]Pedomathemagica: Problem 4

Dominique D. Arrouays, Anne C Richer-De-Forges

ΠΕΔΟ METRON

The Newsletter of the Pedometrics Commission of the IUSS

Issue 33, August 2013

Chair: A-Xing Zhu
Vice Chair: Dick. J. Brus
Coordinator: Murray Lark
Layout: Jing Liu

From the Chair

Dear Fellow Pedometricians and Friends,

Alex reported that over 3000 people attended the 31th Brazilian Congress of Soil Science (actually more than 3500, see the exciting news about this at the IUSS website). Recently (Aug. 18 and 19), I attended the annual conference hosted by the soil remote sensing and soil geography specialty groups of the Chinese Society of Soil Sciences. The number of attendants at this conference reached its record high, from the usual 100 some to over 250. It seems to me that soil science is experiencing a growth which is great and exciting. The reasons for this exciting growth are many but the introduction of digital technology into soil science, I thought to myself, has brought new life into soil science, thus this got to be one of the most important, if not the most important, reason. In this regard, we, pedometricians, should give ourselves a pat on the back, saying "Hmmm, here is good one after all these efforts!"

Inside this Issue

From the Chair…...………………..….……………….. 1

Pedometrics goes to the Tropics ....…2

Spatial analysis of gilgai patterns ……....…………....4

Jenny, PCA and Random Forests ………………......10

Why you don't need to use RPD ……….....………14

Pedometrics 2013 pre-conference workshops …...16

The Working Group on Soil Monitoring ………...17

Activities of the Proximal Soil Sensing Working Group (WG-PSS)……………………….............18

Flyers …………………………………….....…...19

Henglic Wheelersol …………..…………….…… 21

Pedomathemagica …………………………… 22

Great but the question "how do we, the Pedometrics Commission, can benefit from and continue to foster this growth?" I would like to take this opportunity to share some thoughts. Certainly, this is to start the discussion. In a Chinese saying, this is "抛砖引玉" , "Sending the bricks to lure the jades). In other words, using my primitive ideas (bricks) to get your precious advice and polished ideas (the jades). I am writing this for two purposes. The first is to offer an introduction to the new comers about the Pedometrics Commission and to ask our existing pedometricians to help to distribute our welcome to anyone who is interested in pedometrics, and tell them how vibrant this group is and to capture the growth and to foster the growth.

Here is the introduction. Pedometrics, as an organization, is a commission of Division One (Soil in Space and Time) in the International Union of Soil Sciences (IUSS). As a field, based on the definition in Wikipedia Pedometrics is the application of mathematical and statistical methods for the study of the distribution and genesis of soils. It might need to be revised to include modern spatial information gathering and processing techniques (such as global positioning systems, remote sensing, and geographic information systems, now collectively referred to as Geographic Information Science or GIScience) because this (GIScience) becomes increasingly important in pedometrics.

The Commission is one of the most vibrant, if not the most vibrant group in IUSS. In addition to itself, it also houses three working groups directed at emerging areas of pedometrics: Digital Soil Mapping Working Group, Proximal Soil Sensing Working Group, and Soil Monitoring Working Group. The Commission hosts three well attended global scale conferences:

Pedometrics Conference: held by the commission itself every two years, focusing on all aspects of pedometrics;

Global workshop on digital soil mapping: held by the Digital Soil Mapping Working Group every two years (offset the Pedometrics conference by a year), focusing on the techniques and issues for digital soil mapping;

Global Workshop on Proximal Soil Sensing: held by the Proximal Soil Sensing Working Group, focusing on proximal soil sensing.

The Commission grants two awards: the Richard Webster Medal, in honor of Dr. Richard Webster, the founder of pedometrics. The Medal is to honor outstanding researchers in pedometrics. The Award is given every four years (at the World Congress of Soil Sciences). Details of this award can be found at the IUSS website. The other award is the best paper award which is to honor the best paper in pedometrics every year. The award is given at the Pedometrics Conference.

The Commission also offers three communication platforms: the website (www.pedometrics.org), this newsletter, and email list (the pedometrics google group). The website contains a rich set of materials ranging from job ads to archived articles. The newsletters is a place for people to share stories, reports, research thoughts and comments on issues. In this issue we have three of these for you to enjoy.

How to be part of it? Easy, just sign up at the pedometrics googlegroup and come to the conferences, participate in elections and in the award activities, and even help to organize the conferences. For those of you who cannot access googlegroup from where you are, send us (Dick Brus dick.brus@wur.nl and A-Xing Zhu azhu@wisc.edu) an email. We will sign you up.

So, come and be part of this exciting group!

Best wishes,

A-Xing Zhu

Pedometrics goes to the Tropics

By Leigh Winowiecki
International Center for Tropical Agriculture (CIAT)

Pedometrics Conference 2013

As you are all aware, the Pedometrics conference 2013 will be co-hosted by the International Center for Tropical Agriculture (CIAT) and the World Agroforestry Centre (ICRAF) in Nairobi, Kenya (https://sites.google.com/a/cgxchange.org/pedometrics2013/home).

This conference will include a two-day pre-conference workshop on methods for digital soil mapping, a tour of the ICRAF soil and plant spectroscopy laboratory, three full days of conference presentations, a poster session, keynote presentations by the East African Soil Science Society and lead scientists in soil mapping, and a field trip to one of the most spectacular rangeland areas in north-central Kenya.
Both CIAT and ICRAF have been working in Kenya for over 25 years and this conference will bring together international scientists and local Kenyan University students to discuss key advances in pedometrics. The conference organizers are happy to announce that five student scholarships were awarded to African students and a reduced conference fee is available for students.

The Soils of Kenya

Kenya is an incredibly diverse country, both ecologically and culturally, with an area of approximately 582,600 km^2 and a population of about 30 million (from 2000 census data). Just under 70% of its population live in rural areas.

Continental maps produced by ISRIC and JRC (Figure. 1) and country-level maps produced by the Kenya Soil Survey (Figure. 2) indicate that most of the WRB Reference Groups exist in Kenya. This diversity of soils is due to the geographically and climatically diverse regions of Kenya which include the humid regions (> 1000 MAP), sub-humid regions (< 1000 MAP) and semi-arid regions (450-900 MAP), as well as arid to very arid regions (150-500 MAP), combined with a high geologic diversity. The most impressive ecological features of Kenya include the Great Rift Valley, which extends 6,000 km from northern Syria to Mozambique, and Africa's second highest peak, Mt. Kenya (5,199 m). Escarpments to the east and west border Kenya's rift valley, and the floor contains volcanoes, some still active, as well as a series of sodic lakes that are important breeding grounds for great white pelicans and feeding areas for lesser flamingos. The soils of the Rift Valley have volcanic parent material, and most are classified as Andosols, with high P-sorption, high aluminum saturation and also high fluoride content.

The Lake Victoria basin, which covers an area of

~184,200 km² and extends into Tanzania, Kenya, Uganda, Rwanda and Burundi, supports some of the densest rural populations in the world with up to 1,200 persons per km². The dominant soils within the highlands of the Lake Victoria basin include Ferrasols, Nitisols, Cambisols, and Acrisols (FAO-UNESCO, 1988). While soils of the Kano plains within the basin are generally Luvisols, Vertisols, Planosols, Cambisols, and Solonetz (FAO-UNESCO, 1988). The Lake Victoria basin has been plagued by high soil erosion rates, increasing land degradation, which are contributing to both decreased agricultural productivity in the region and diminishing water quality of Lake Victoria (World Agroforestry Centre, 2006).

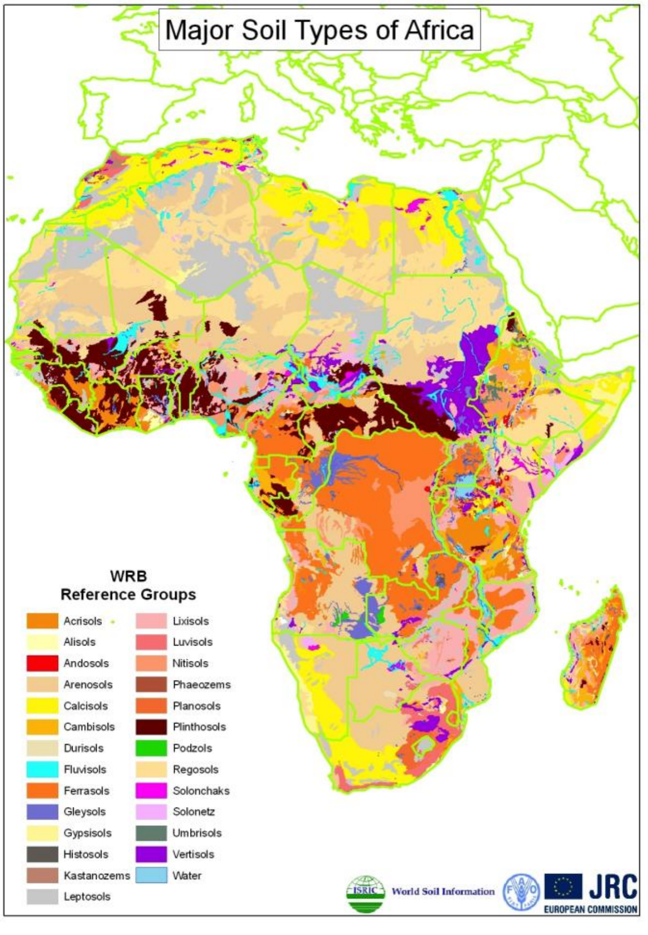


Figure 1. Major Soil Types of Africa

Dominating Kenya's landscape are the dryland ecosystems, which are the arid and semi-arid lands of north-central and eastern Kenya. These areas are inhabited by nomadic pastoralists. The dominant soil types include: Regosols, Plansosols, Lixisols, Solonchaks, Calcisols, Arensosols, among other soil types that are characterized by high sand content. Recent studies have indicated that these soils have low carbon content, high erosion prevalence and are particularly vulnerable to over grazing and compaction (Vågen et al., 2012). Current efforts to curb this degradation with community pastoralists groups have made strides toward rehabilitation and improved productivity. The field trip on the Pedometrics Conference will highlight the soils of this region as well as the conservation efforts (https://sites.google.com/a/cgxchange.org/pedometrics2013/home/field-trip.)

For a set of recently developed maps of soil condition and land degradation for sub-Saharan Africa you can visit the website of the World Agroforestry Centre's GeoScience Lab (http://gsl.worldagroforestry.org/).

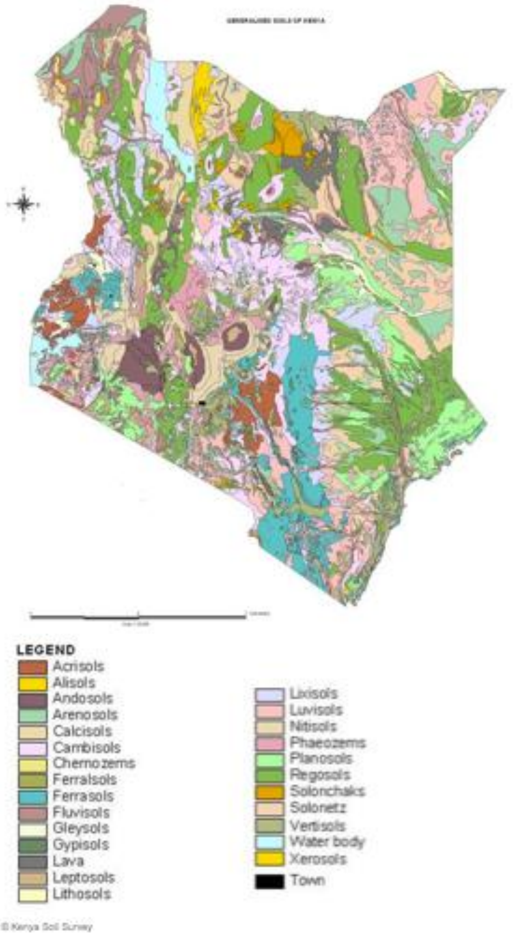


Figure 2. Generalized Soil Map of Kenya
(*from Kenya Soil Survey*)

References:

(*Turn to Page 15*)

Spatial analysis of gilgai patterns

By R. Webster with contributions from
A.E. Milne and R.M. Lark
Rothamsted Research

'Gilgai' is an Australian aboriginal word for a wet hollow. Gilgais are widespread in Australia, especially on the at eastern plains, where they pock the landscape (see Figure. 1). They are thought to have formed as a result of the soil's repeatedly shrinking and swelling as it dries and wets. It seems as though on drying the soil shrinks coherently over some distance until it fails and cracks. Rain or flood carry clay into the cracks where the soil swells and pushes the soil sideways, and in many instances upwards to form 'puffs', too. On every repetition of the cycle the cracks open in the same places, and so the pattern becomes entrenched to what we see today. When viewed from the air the gilgais can appear in characteristic patterns, as in Figure 2, with each roughly circular and separated from its neighbors by similar distances. One can imagine that their distributions represent distortions of hexagonal close packing, and that prompts one to ask whether there is some regularity in the repetition, and if there is then what its characteristic wavelength might be.

Figure 1. Ground photograph of gilgai on the Bland Plain

That idea came to Russell and Moore (1972) more than 40 years ago. Russell and Moore measured the height of the land above a local datum at frequent intervals along transects and estimated the average wavelength and amplitude of the micro-relief by Fourier analysis of their data.

But what of the soil itself? The soil in the gilgais is typically wetter than that on the intervening plain. It can differ in other respects, too; in clay content, depth to carbonate or gypsum, and salinity, for example. It occurred to me that these properties might also repeat in a periodic way. So I pursued the idea by sampling the soil along a transect 1.5 km long across gilgai patterned ground on the Bland Plain of New South Wales.

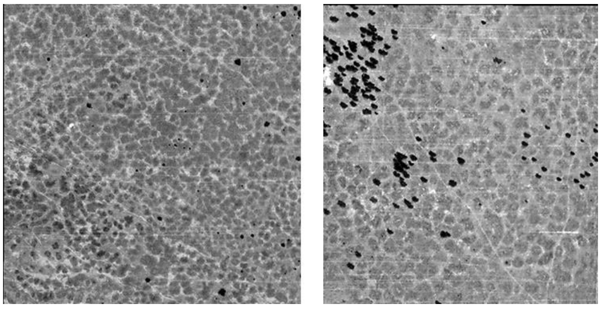


Figure 2. Parts of two digitized aerial photographs of the Bland Plain; at Caragabal (left) and at Back Creek (right).

I recorded the height of the soil surface above a local datum and the morphology, pH and electrical conductivity of the soil down to 1 m on cores at frequent intervals. All of the properties showed strong spatial dependence at that scale. More interesting, however, was that the correlograms of several fluctuated in an apparently periodic manner and that their power spectra, obtained as transforms of the correlograms, had peaks at a frequency corresponding approximately to the distances between the centers of neighboring gilgais, about 34 m - see Webster (1977). It seemed that there was indeed a degree of regularity in the spatial pattern. Margaret Oliver and I (Webster and Oliver, 2007) had already come to that view after modelling the variogram; after all, for a second-order stationary process the variogram and the power spectrum contain the same information.

Electrical conductivity on the Bland Plain

Here are the results of analysis of the electrical conductivity (EC) measured in the 30~40-cm layer of the soil at 4-m intervals and converted to their common logarithms. Figure. 3 displays the variogram and its corresponding correlogram computed by the method of moments; they show distinct waves. A nested model including a periodic component with a wavelength of 34.7 m fits the experimental semivariogram well.

The general equation for the transformation of the experimental correlogram $\hat{\rho}(k)$ to its power spectrum is:

$$\hat{g}(f) = 1 + 2\sum_{k=1}^{K} \hat{\rho}(k)\omega(k)\cos(2\pi f k) \ , \quad (1)$$

for frequency f in the range 0 to 1/2 cycle and lags from 1 to a maximum of K. The quantity $\omega(k)$ is a weight that depends on the limit K and on the shape of the window within which the transform is computed. For present purposes I have set K = 60 and used the Parzen window to minimize the leakage.

Figure. 3 shows the experimental variogram, with a nested model containing a periodic component fitted to it, and the correlogram. The power spectrum derived from the latter appears in Figure 4. Notice the peak at a frequency of ≈ 0.12, which is equivalent to wavelength of 8.4 sampling intervals or 34 m.

In addition to my measuring the height of the land, I recorded at each sampling point the nature of the land surface as 'plain', 'depression', i.e. gilgai, or 'puff'. Lark (2005) analyzed and modelled these records, and he obtained periodic empirical variograms of both plain and depressions with wavelengths of 8 to 9 sampling intervals (32~36 m).

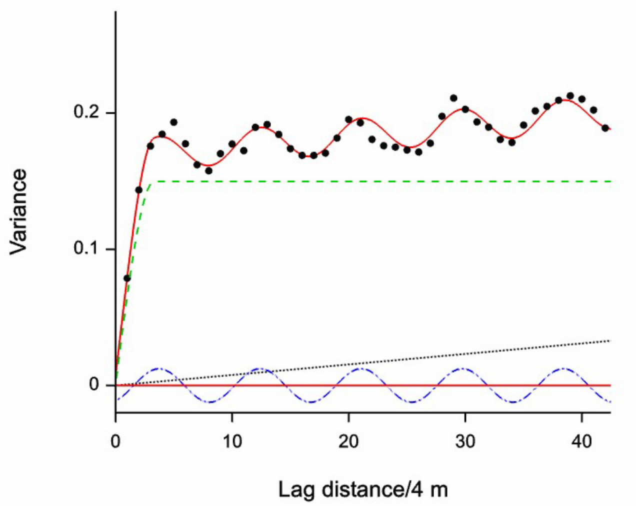


Figure 3. Empirical variogram of $\log_{10}$ electrical conductivity at Caragabal with four components of the fitted model shown separately. Note in particular the periodic component.

The patterns are two-dimensional, of course, and the results of the one-dimensional analysis led to the question: is the two-dimensional arrangement of the gilgais regular?

Measuring properties of the soil, such as its electrical conductivity, at enough sites for a two-dimensional spectral analysis was prohibitively expensive. An alternative was to analyze the photographic images. At the time, the 1970s, however, the available digitizing equipment proved too unreliable to furnish

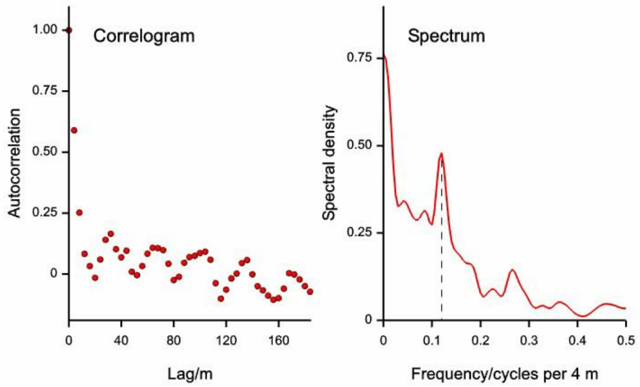


Figure 4. Correlogram of $\log_{10}$ electrical conductivity at Caragabal (left) and its corresponding spectrum computed with the Parzen lag window and bandwidth 60.

sufficient data. An alternative was to process the images optically, a technique then being explored by Preston and Davis (1976) for analyzing photomicrographs of sedimentary rocks.

Optical Fourier transformation

The principle of the method is based on the interference of light waves refracted through a lens and projected on to a screen, which is the physical analogue of the Fourier transformation.

Let us suppose that a gilgai pattern has a characteristic wavelength, which seems reasonable given the photographs reproduced in Figure 2. Let us denote this wavelength as ω. When transformed through the optical system a ring of light should appear in the plane of the Fourier transform at a distance d from the center given by

$$d = \omega\frac{\lambda f}{2\pi} \ , \quad (2)$$

where λ is the wavelength of light and f is the focal length of the lens.

I made transparencies of several aerial photographs, including ones of the scenes in Figures 2. I put them on a light bench, shone a coherent monochromatic beam of light from a laser through them and converted the images to their Fourier transforms in this way. The results were in all instances the same. The transforms were dominated by bright centers away from which the light gradually diminished towards the peripheries; there were no evident subsidiary peaks to indicate periodicity. They were disappointing.

Modern digitizing equipment, hugely increased computing power and new mathematical developments such as wavelet analysis enable us to revisit the question. Alice Milne and Murray Lark and

I took advantage of the opportunity. We described in detail our search for an answer in the *Australian Journal of Soil Research* (Milne et al., 2010), and here I mention the highlights.

Digital analysis of aerial photographs

The photographs showing gilgai patterns that we analyzed were taken of the Bland Plain by the New South Wales Department of Lands and Surveys in the late 1960s. One is of Caragabal station, which I sampled originally, the other is several km to the west at Back Creek. We digitized electronically a rectangular patch corresponding to about 25 ha on the ground of each photograph at a resolution of 157 pixels per cm (≈ 1.3 m diameter of a pixel) and recorded the grey level in the range 0 to 256. The results are as shown in Figure 2.

Correlograms and spectral analysis

We first computed variograms and correlograms from the pixel data for individual rows and columns in the images by the method of moments. Figure 5 shows examples, one from Caragabal and one from Back Creek. Notice how both fluctuate in an apparently periodic way. Their spectra, Figure 6, contain strong peaks corresponding to the lengths of the periods in the correlograms.

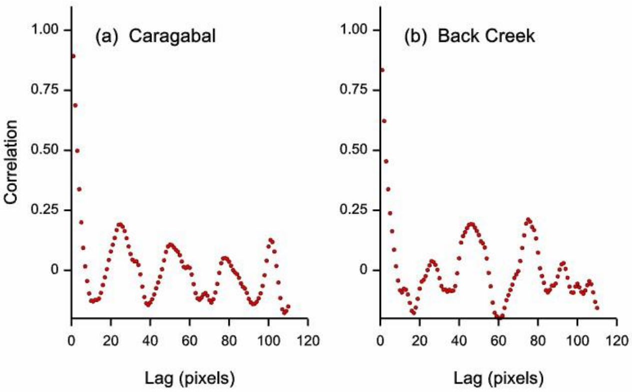


Figure 5. Correlograms of transects across the digitized images of (a) Caragabal and (b) Back Creek.

We also computed the two-dimensional variograms on grids. Again, the autocorrelations are obtained as the complements of the semivariances.

Figures 7 and 8 show the two-dimensional correlograms. Notice how both correlograms decay from their central peaks but with pronounced waves on them.

From these correlograms we computed the two-dimensional spectra. The spectra appear in Figures 9 and 10. Both have peaks at their origins arising from

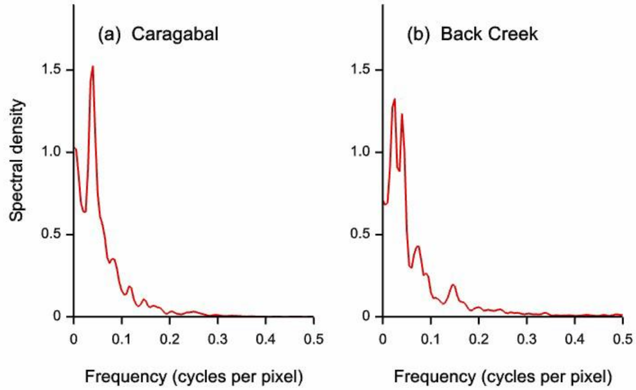


Figure 6. Power spectra computed from the correlograms in Figure 5: (a) Caragabal and (b) Back Creek..

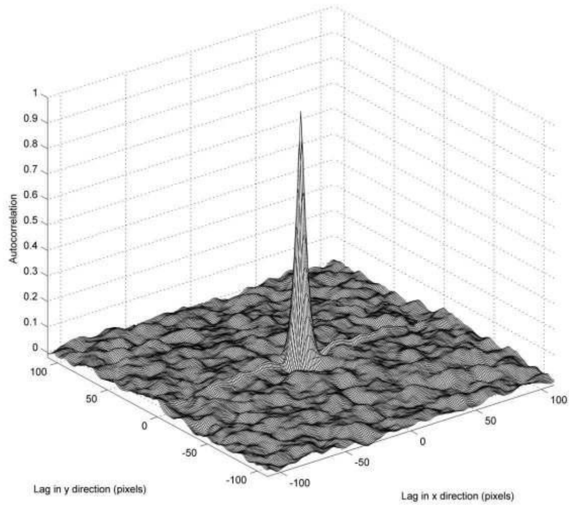


Figure 7. Two-dimensional correlogram of the Caragabal image.

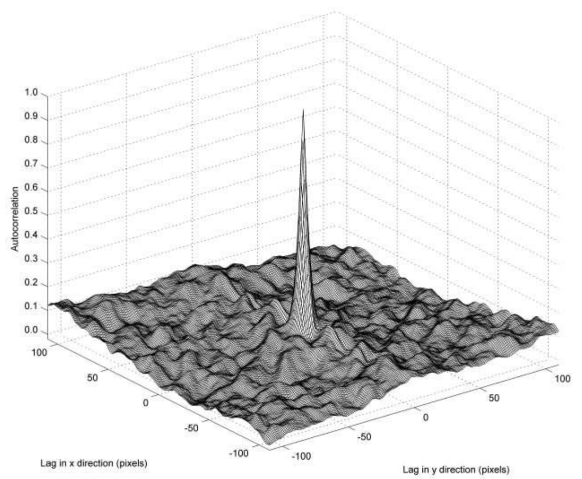


Figure 8. Two-dimensional correlogram of the Back Creek image.

uncorrelated information in the images, just as the optical transforms did. Both, however, have subsidiary rings surrounding the bases of the peaks at distances from 0.025 to 0.04 cycles per pixel. These frequencies correspond to wavelengths on the ground of 52 to 32 m. So, we found periodicity in the two-dimensional spectra confirming our visual impression from the photographs with wavelengths that accord.

Wavelet analysis

My colleagues and I took the analysis a stage further with wavelets. As you may know, spectral analysis loses all information on the positions of the features in images. The correlogram and spectrum average the statistics over the whole field of data. Wavelet analysis, in contrast, retains local information by decomposing data into separate components of both frequency (called 'scale' in the jargon) and location as locally compact wavelets move over the data. In this way it can reveal *where* characteristics of an image or series of data change. For this reason it requires no assumption of stationarity.

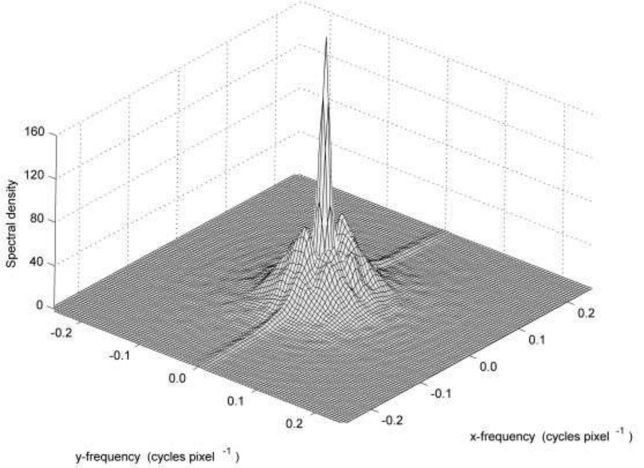


Figure 9. Two-dimensional spectrum of the Caragabal image.

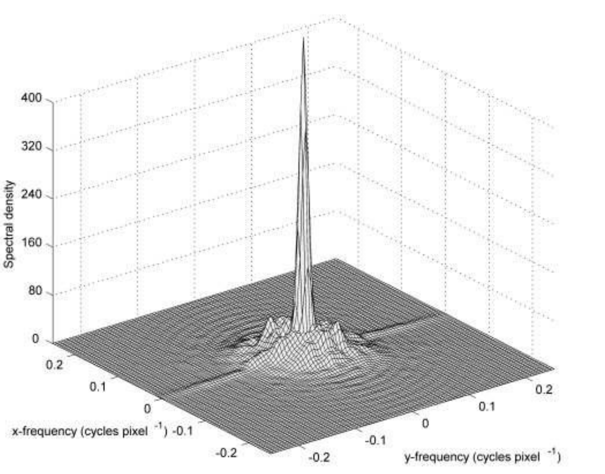


Figure 10. Two-dimensional spectrum of the Back Creek image.

There is a great deal to wavelet analysis, just as there is to geostatistics, and we cannot describe here all aspects of wavelets. We can at best refer you to some of our papers (Lark and Webster, 1999, 2001, 2004; Milne et al., 2010; and the excellent book by Percival and Walden, 2000).

Basically, we have a set of wavelet functions

$$\Psi_{\lambda,x} = \frac{1}{\sqrt{\lambda}}\Psi\left(\frac{u-x}{\lambda}\right), \quad \lambda>0 \tag{3}$$

in which $\Psi(x)$is a 'mother wavelet' centerd at position x, λ is a scale parameter that controls the width over which the wavelet takes non-zero values, and u represents a displacement from x. By convolving the wavelet with the data we obtain wavelet coefficients:

$$W(\lambda, x) = \int_{-\infty}^{\infty} z(u)\frac{1}{\sqrt{\lambda}}\Psi\left(\frac{u-x}{\lambda}\right)\mathrm{d}u\ . \tag{4}$$

Varying x moves the wavelet over the space to provide coefficients at each position. By changing λ we change the scale at which we view the variation. The smaller is the λ the finer is the scale at which we describe the variation about x. Increasing λ dilates the wavelet and coarsens the scale.

In the discrete wavelet transform λ is incremented in a series of powers of 2, thus: $\lambda = 2^j$; j = 1, 2, … to some maximum usually set by the extent of the data. The result of the convolution at each value of j produces a smooth representation and a detailed component of the data, and by changing j we obtain a multi-resolution analysis.

Figures 11 and show our results for Caragabal and Back Creek respectively.

Wavelets also have variances attached to them, and for present purposes we have calculated them by the maximal overlap discrete wavelet transform (MODWT) of Percival and Guttorp (1994). They are given by

$$\sigma_j^2 = \frac{1}{2^j n_j}\sum_{x=1}^{n_j} d_j^2(x)\ , \tag{5}$$

where n_j is the number of data involved in the computation, and $d_j^2(x)$ is the MODWT coefficient at position x for the j^{th} scale.

We calculated the variances in three directions, namely along the rows, down the columns and across the diagonals, and in Figure 13 we plot them against scale. The graphs are similar for the two images. We note first that the maximum variances along the rows and down the columns are similar and larger than those on the diagonals. These show that the variation is isotropic. More importantly, perhaps, both have pronounced peaks at the 16~32 pixel scale, and these accord with the visibly largest variation in the detail components at that scale in Figures 11 and 13. Further, this scale of maximal variation corresponds with the wavelengths estimated in the correlation analysis and peaks in the spectra.

So, we can conclude that the gilgai patterns that appear in plan to be periodic and isotropic are indeed regular and isotropic with wavelengths that we can estimate with confidence.

Reference

Lark, R.M. 2005. Spatial analysis of categorical soil variables with the wavelet transform. *European Journal of Soil Science*, 56, 779~792.

Lark, R.M. and Webster, R. 1999. Analysis and elucidation of soil variation using wavelets. *European Journal of Soil Science*, 50, 185~206.

Lark, R.M. and Webster, R. 2001. Changes in variance and correlation of soil properties with scale and location: analysis using an adapted maximal overlap discrete wavelet transform. *European Journal of Soil Science*, 52, 547~562.

Lark, R.M. and Webster, R. 2004. Analysing soil variation in two dimensions with the discrete wavelet transform. *European Journal of Soil Science*, 55, 777~797.

Milne, A.E., Webster, R. and Lark, R.M. 2010. Spectral and wavelet analysis of gilgai patterns from air photography. Australian Journal of Soil Research, 48, 309~325.

Percival, D.B. and Guttorp, P. 1994. Long-term memory processes, the Allan variance and wavelets. In: *Wavelets in Geophysics* (editors E. Foufoula-Georgiou and P. Kumar), pp. 325~344. Academic Press, New York.

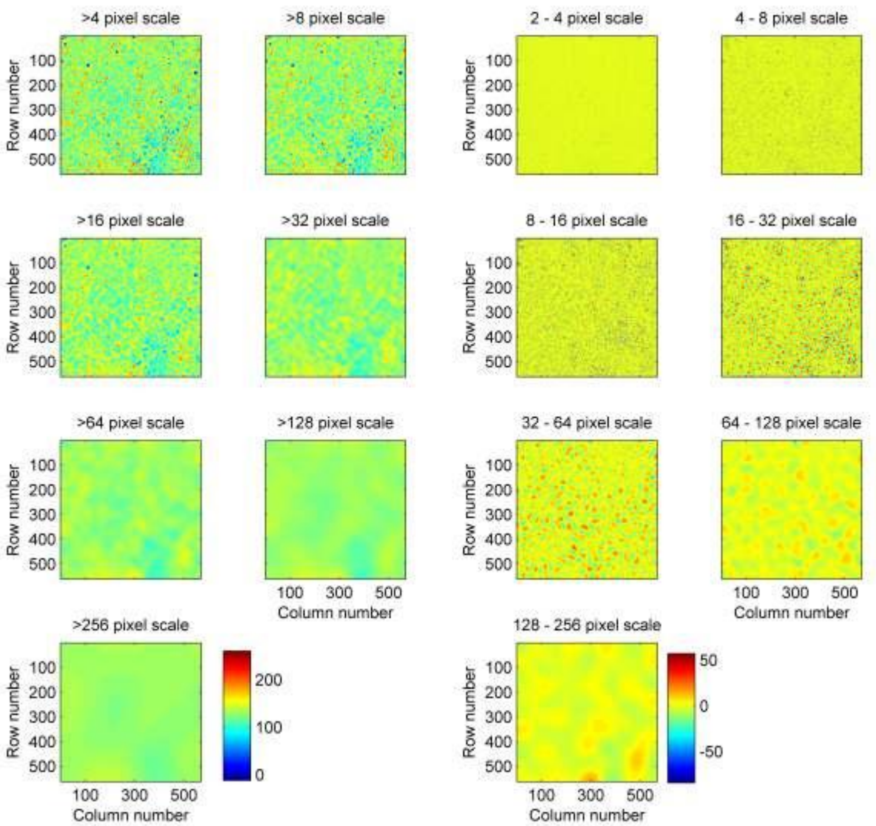
Percival, D.B. and Walden, A.T. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge University Press, Cambridge, UK.

Preston, F.W. and Davis, J.C. 1976. Sedimentary porous materials as a realization of a stochastic process. In: *Random Processes in Geology* (editor D.F. Merriam), pp. 63{86. Springer-Verlag, New York.

Russell, J.S. and Moore A.W. 1972. Some parameters of gilgai microrelief. *Soil Science*, 114, 82~87.

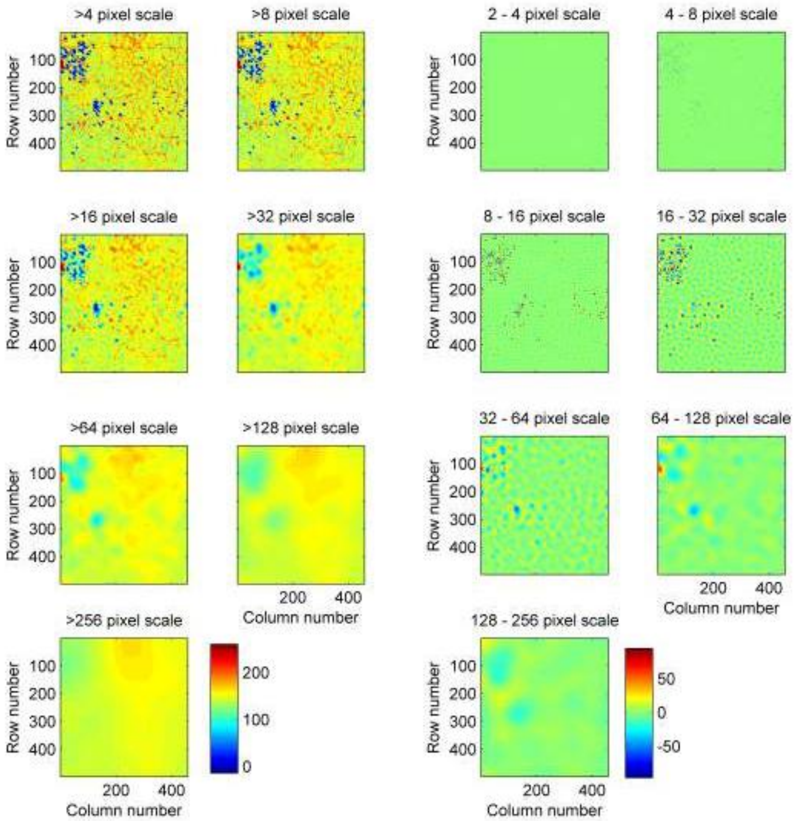
Webster, R. 1977. Spectral analysis of gilgai soil. *Australian Journal of Soil Research*, 15, 191~204.

Webster, R. and Oliver, M.A. 2007. *Geostatistics for Environmental Scientists*, 2nd edition. John Wiley & Sons, Chichester.



(a) **(b)**

Figure 11. Multi-resolution analysis of the Caragabal image; (a) smooth representations, and (b) detail components.



(a) **(b)**

Figure 12. Multi-resolution analysis of the Back Creek image; (a) smooth representations, and (b) detail components.

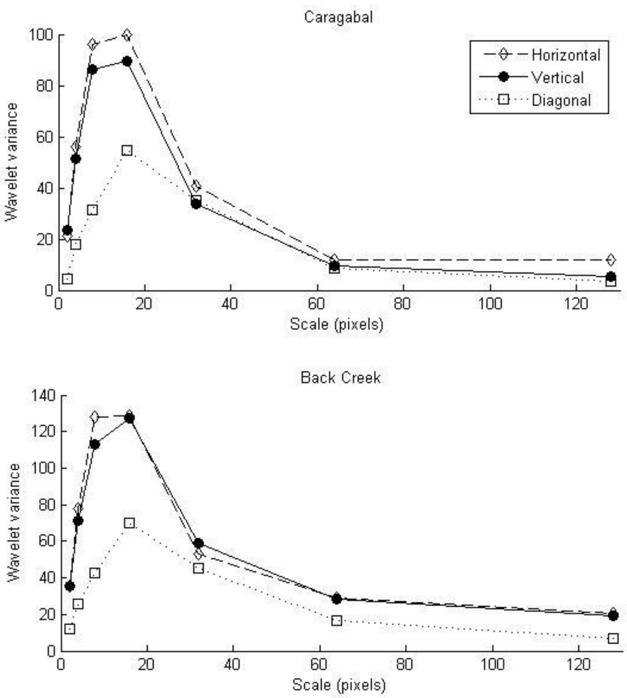


Figure 13. Two-dimensional wavelet variances for Caragabal and for Back Creek. The points are plotted at the lower bounds of the scale range on the abscissa, and the lines that join them are for visual clarity only.

Jenny, PCA and Random Forests

By Budiman Minasny & Alex. McBratney
University of Sydney

The 2011 Hans Jenny Memorial Lecture in Soil Science was delivered by Prof. Garrison Sposito from UC Berkeley. He called his talk - The Genius of Soil. The video is available at http://youtu.be/y3q0mg54Li4. In the last part of the lecture, Gary drew attention to one of the little known paper by Hans Jenny in 1968 which was presented at the *Study Week on Organic Matter and Soil Fertility*, April 22-27, 1968, organised by the Pontificia Academia Scientiarum a scientific academy of the Vatican. This is probably the early Global Soil Carbon workshop.

The paper by Jenny et al. (1968) was the first chapter in this book, available at http://tinyurl.com/begpa2h (Jenny's appendix paper on "The image of soil in landscape art, Old and New" in the same book is better-known than this paper). In the study, Jenny collected 97 soil samples across a moisture transect in the Sierra Nevada, California where the variation in the factors of soil formation were to some degree controlled. The mean annual precipitation (MAP) is between 80 and 2000 mm, and mean annual temperature (MAT) between 10 and 16°C. The flora was restricted to pine tree and grass. The aspect is always southeast, with slopes varying from 0 to 30%. The parent materials are acidic and basic igneous rocks. Jenny used this data to quantitatively fit "an integrated clorpt model" where all factors were simultaneously modelled in the form of a multivariate linear regression:

$$s = a + k_1\ \text{MAP} + k_2\ \text{MAT} + k_3\ \text{Parent Material} + k_4\ \text{slope} + k_5\ \text{Flora} + k_6\ \text{Latitude}$$

In addition, Jenny also realised there would be correlation among the independent factors:

"When the independent variables X_1, X_2 , X_3, … are highly self-correlated (collinear) the slope coefficients *b* become unstable, even meaningless as to sign. Regressing for example N against Precipitation (*P*) (in.) and Temperature (*T*) (°F) gives

$$N = 0.350 + 0.0012\,P - 0.0055\,T$$

with R^2= 0.324. Introducing Leaching value (*Li*, in.), which is highly collinear with *P*, results in

$$N = 0.375 + 0.0037\,P - 0.0062\,T - 0.0029\,Li$$

The slope coefficient of *P* has tripled and that of *Li* is negative, which is absurd from the viewpoint of soil leaching. R^2 remains essentially unchanged as 0.327.

The handicap of self-correlation can be overcome by computing "*principal components*" (Some of the content of this paper is later used in the last chapter of Jenny's 1980 *The Soil Resource Book*, pp. 361-363).

Gary in his talk indicated that this was the first paper that used PCA in soil studies. Intrigued by his comment, we tried to find out whether there are predecessors. Jim Wallis one of the co-authors of Jenny's paper (who then worked at I.B.M. Watson Research Center, Yorktown Heights, N.Y.) wrote to Gary (email message from Jim Wallis to Gary Sposito, May 15, 2011):

"It is not possible for me to say definitely that my work/paper was the first use of principal-components regression in soil science, but the probability that it was is extremely high. What is certain is that very few people at the time seemed able to understand the methodology or provide references to similar work. It seemed that it would help me with my dissertation on accelerated soil erosion, and I used it in my dissertation - it was highly controversial at the time.

A sidelight on how I arrived at the methodology follows. There was a Professor Meredith in the Psychology Department at the time, he taught a graduate course in Factor Analysis which I unofficially audited, and it seemed to me that if one used principal-component regression to determine the number of factors at work in soil formation (eigenvalues >1) and rotated the matrix into the variable space by Varimax that you would have a quantitative measure of Jenny's CLORPT equation. I wrote a 120-variable computer program to do just that. Jenny was not on campus that year, but he came back in the spring of 1966, got excited by its possibilities for pedology, although I had little to do with the writing of our joint paper, beyond a few conversations and notes that did not get preserved. He demanded that I give a seminar to the Soils Department on the subject, so I did."

Jim Wallis introduced PCA to hydrology in a 1965 paper[1] (Wallis 1965), and also wrote a FORTRAN program called WALLY1. Jim Wallis is a well-known hydrologist who wrote the first paper on fractal in hydrology with Benoit Mandelbrot (1968), and he was the president of the Hydrology section at the AGU.

[1]The hydologists always seem to be a couple of years ahead of the soil scientists

Jenny, PCA and Random Forests

The earliest references to techniques in Principal Component Analysis (PCA) were Karl Pearson in 1901 and Hotelling (1933). However it was not until the 1960s with the availability of computers that the analysis became practical. Earlier papers on soil can be found on mostly on factor analysis (rotated principal components or principal factors that are not necessarily orthogonal). The thrust for factor analysis was largely from the social sciences (Psychometrics) rather than the physical ones. Rayner's 1966 paper was on principal coordinate analysis which involves finding the principal components corresponding to similarity matrices. This analysis was invented by John Gower at Rothamsted principally to help James Rayner with the soil similarity problem, however as it turns out it is virtually the same as multidimensional scaling which was invented in the 1950's by the psychologists. The first use of multivariate statistical methods for soil (data) that we know of is Cox and Martin (1937). See the Ordination article by Dick Webster in Pedometron 29.

Searching through the Web of Knowledge, we found several earlier papers that used PCA in soil studies. A paper by Gyllenberg (1964) from Finland and another one by Skyring and Quapling (1968) from Canada used PCA as a way to describe soil microbial diversity. Yamamoto and Anderson (1967) used PCA (instead of multiple linear regression) to find the degree of association between soil physical properties (soil aggregate stability, erodibility) and the soil-forming actors for wildland soils of Oahu, Hawaii. This bears the closest resemblance to Jenny's 1968 paper. Their study was also inspired by Jim Wallis' paper in hydrology (Wallis, 1965). There was also a PhD dissertation by John Berglund in 1969 from State University College of Forestry at Syracuse University, where PCA was used to develop and interpret prediction equations to estimate forest productivity from its soil properties. Dick Webster and his student Ignatius Wong (1969) used PCA to analyse soil data collected along a transect. The main use here was for ordination - many soil properties were combined into a first principal component so that soil property variation could be plotted as a graph along a transect.

While Jenny may not be the first to use PCA in soil studies, the 1968 paper lays the fundamentals of what is now called digital soil mapping. It should be a good reminder for us on how to mindfully choose the best covariates and model. We need to remember that Jenny's linear model is used to explain the factors that control the distribution of soil properties, not specifically as a spatial prediction function.

Jenny (1980) wrote:

"The computer's verdict of tangible linkages of soil properties to the state factors pertains to today's environment. Either the pedologically effective climate has been stable for a long time, or past climates are highly correlated with modern ones, or the chosen soil properties have readjusted themselves to today's precipitation."

Nowadays (notwithstanding its simplicity) PCA is still extensively used in soil science and pedometrics, for drastically reducing the number of variables in soil spectral data, finding patterns (clusters) in the data, reducing dimensions of microbial diversity data, or satellite images, etc. According to Scopus, since 2010, there has been an average of 450 papers per year on the application of PCA to soil data.

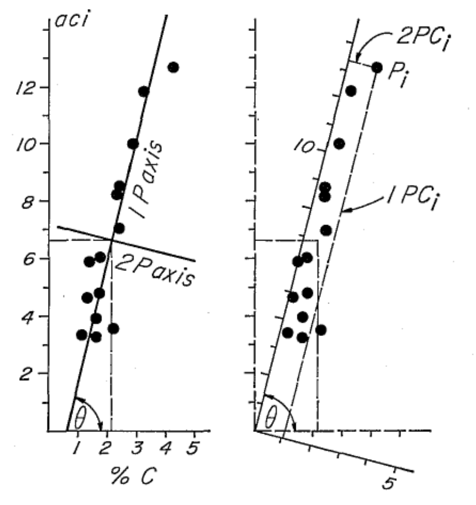
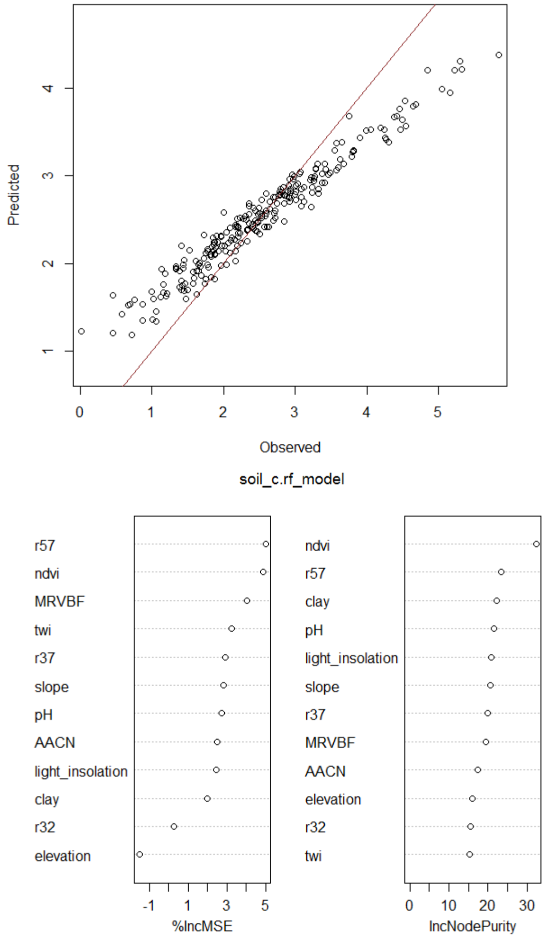


Figure 1. Illustration of converting original variables (aci, C) to first and second principle components (from Jenny et al., 1968).

Pygmy Forest to Random Forests

Research in digital soil mapping now has moved from carefully controlled environmental factors to "real-world" soil data, either collected from stratified random sampling or using legacy soil data. Models of the Jenny et al. (1968) type are still being developed (Gray et al., 2012), while others prefer to use data-mining techniques. Data-mining models are usually treated as a black-box as they are complex and cannot be easily or explicitly written out. However, the results can be expressed as significant predictors or variables of importance and usually interpreted as 'knowledge discovery' from databases which are then sometimes justified a *posteriori* by principles of soil genesis. As opposed to a process-based model, where the process needs to be specified first, the data-mining approach is said to "learn" the process through the data. As an example, the Random Forests technique has been used a lot in digital soil mapping as it is freely available and it has been claimed that "Random forests does not overfit. You can run as many trees as you want" (From the Random Forests Manual by Breiman and Cutler). In addition, the author also made claims that it is: "The most accurate current prediction", "a complex predictor can yield a wealth of 'interpretable' scientific information about the prediction mechanism and the data."

An example of the use of Random Forests is given in Figure 2, which shows the prediction for some surface soil carbon data in the Hunter Valley, NSW, Australia, where the fit is excellent on the training data (using 100 trees), R^2 = 0.94. The variable of importance indicated that in addition to indices calculated from Landsat bands, terrain attributes of MrVBF (Multi-resolution Valley Bottom Flatness) and TWI (topographic wetness index) play important roles. The map confirmed this and it is in accordance with our pedological knowledge, where carbon concentration is expected to be larger in areas with higher moisture and areas of deposition (knowledge discovery).



But wait a minute, Figure 3, shows the fit on an internal validation (out of bag estimates) and an independent validation data, where there is no fit at all. The soil carbon data has very little correlation with any of the terrain attributes and is very weakly correlated with some Landsat imagery. It is obvious that Random Forests can easily overfit the data. Overfitting implies the model describes the noise in the data (perfect fit on the training data), while has poor predictive capability in the validation data. (The data and R code are available to download from here, and you can experiment yourself with the notion that RF can fit anything). It is quite interesting that scientists take the statement "Random Forests does not overfit" as the truth, and repeatedly quote this in many papers without any question.

A recent news article mentioned the latest breakthrough in technology: "With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart." Perhaps we should tell you that we need to explore Deep Learning tools for pedometrics. But we think now we need to remind ourselves that explicit linear models should be at least considered as a starting point for exploratory data analysis before trying the fancy tools. There is no magic algorithm that can fit everything — yet not overfit.

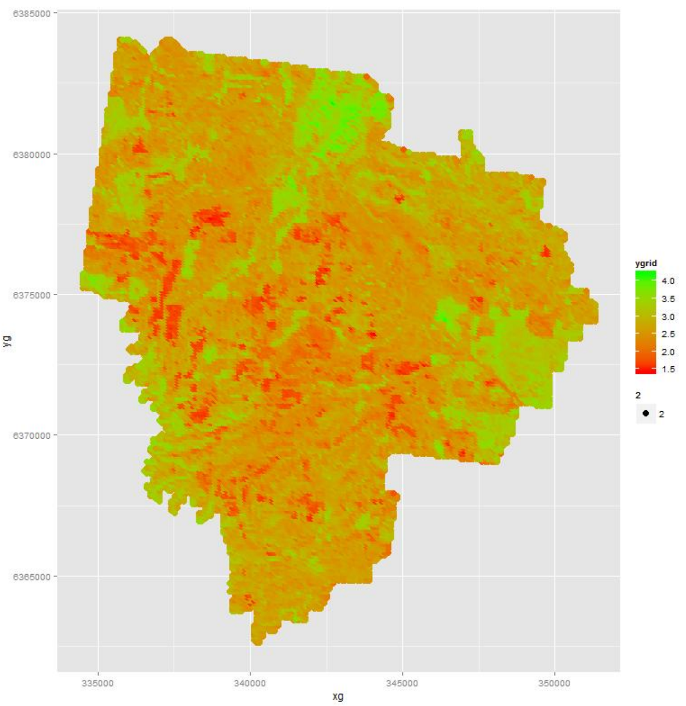
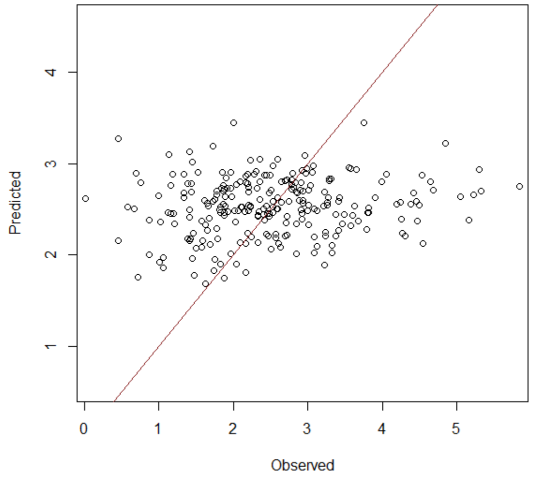


Figure 2. The prediction of soil carbon content in the Hunter Valley using random forests, its predicted map, and variable of importance (for prediction).

Jenny, PCA and Random Forests

Summary

We've come a long way from Jenny's pedological study in the Pygmy Forest to using Random Forests for making soil predictions. Technology has advanced, powerful computers that can handle complex algorithms and there is now widespread availability of high-resolution covariates. We still stick to the same principle that while we need to make use of all of the new technologies, common sense and parsimony must prevail over fancy tools.



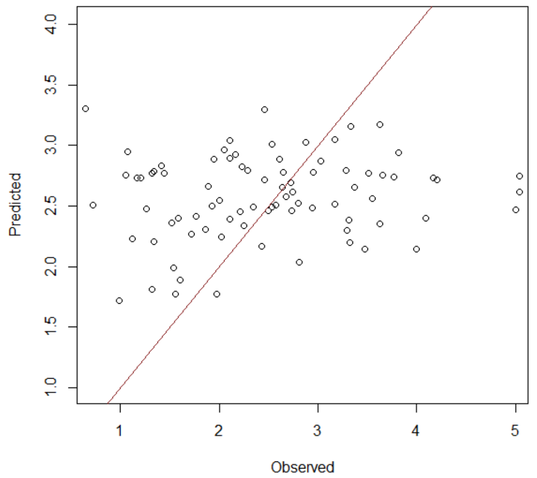


Figure 3. The out of bag fit vs. observed carbon content fitted using random forests (up) and the fit for an independent validation dataset (down).

References

Berglund, J.V., 1969. The Use of Modal Soil Taxonomic Units for the Prediction of Sugar Maple Site Productivity In *Southern New York. Dissertation Abstracts* 29B, 2696-2696.

Cox, G.M., Martin, W.M., 1937. Use of a discriminant function for differentiating soils with different azotobacter populations. *Journal paper No. J451 of the Iowa Experimental Station*, pp. 323–332.

Gray, J., Bishop, T., Smith, P., Robinson, N., Brough, D., 2012. A pragmatic quantitative model for soil organic carbon distribution in eastern Australia. In *Digital Soil Assessments*, CRC Press, pp. 115-120.

Gyllenberg, H.G., 1964. An approach to numerical description of microbial isolates of soil bacterial populations. *Ann Acad Sci Fennice*, Ser. A Biol 81, 1-23.

Jenny, H., 1980. *The Soil Resource*. Springer-Verlag.

Jenny, H., Salem, A.E., Wallis, J.R., 1968. Interplay of soil organic matter and soil fertility with state factors and soil properties. *Study Week on Organic Matter and Soil Fertility*, Pontif. Acad. Sci. Scripta varia, 32, 5-36.

Mandelbrot, B.B., Wallis, J.R., 1968. Noah, Joseph, and operational hydrology. *Water Resources Research* 4 , 909-918.

Rayner, J.H. 1966. Classification of soils by numerical methods. *Journal of Soil Science*, 17, 79-92.

Skyring, G.W., Quapling, C., 1968. Soil bacteria: principal component analysis of descriptions of named cultures. *Canadian Journal of Microbiology* 15, 141-158.

Wallis, J.R., 1965. Multivariate statistical methods in hydrology - A comparison using data of known functional relationship. *Water Resources Research* 1, 447-461.

Webster, R., 2010. An early history of ordination in soil science Ordination. *Pedometron* 29, 20-24.

Postscript by Alex.

The availability of principal components and more general multivariate methods for looking at soil took off fairly quickly after the sixties. When I did my first serious pedometrics work, which was in the long hot summer of 1976, with Dick Webster at Yarnton, software for doing PCA, discriminant analysis, principal coordinates etc. was readily available in programs such as Genstat, BMDP, SPSS and SAS. They were the powerful forerunners of S and then R. In my alma mater at Aberdeen another mentor the soil physical chemist Michael Court very much favoured principal factor analysis over principal components analysis. Largely with Dick Webster's help I learned the mechanics of the multivariate methods – and they continue to serve well. They should be in any pedometrician's toolbox.

Why you don't need to use RPD

By Budiman Minasny & Alex. McBratney
University of Sydney

Another great myth in pedometrics[1] is the use of RPD as a measure of the goodness of fit. RPD, Ratio of Performance to Deviation, is the ratio of the standard error in prediction to the standard deviation of the samples, which is frequently used in (soil) NIR literature, and now in digital soil mapping for assessing the usefulness or goodness-of-fit of calibration models. It attempts to scale the error in prediction with the standard deviation of the property. If the error in estimation is large compared with the standard deviation, then the model is not performing well.

RPD was initially used by Williams (1987) for assessing the goodness of fit for NIR calibration (in agricultural and food products). Batten (1998) wrote "Williams (pers. comm.) suggested that RPD values greater than 3 are useful for screening, values greater than 5 can be used for quality control, and values greater than 8 for any application."

Then a paper by Chang et al. (2001) used RPD for assessing the ability of NIR spectra to predict soil properties. In that paper the authors made 3 arbitrary categories: Category A: RDP > 2.0, Category B: RDP 1.4–2.0 and Category C: RDP< 1.4. This somehow was interpreted by other authors as the 'standard' classification and referred excellent models when RPD >2; fair models when 1.4 < RPD < 2; and non-reliable models when RPD <1.4. Since then some other authors also have slightly modified this to make a new criterion for general quality of soil prediction.

There is no statistical or utilitarian basis as to how these thresholds were determined. And people have tend to use these RPD classes to designate their models as 'excellent' (RPD > 2 becomes the golden standard) with no further questions asked. Veronique Bellon Maurel et al. (2010) questioned the use of RPD and pointed that the normalization in RPD only works assuming normally distributed values. For different soil properties, due to the difference in their statistical distributions, the std. deviation will not have the same interpretation in terms of the range of values. It does not correctly represent the spread of the population. They recommended the use of RPIQ (ratio of performance to interquartile range/ IQ = Q75 - Q25). It is a better index than RPD, based on quartiles, which better represents the spread of the population.

Looking again to the definition of RPD, we can see that it is in fact just inversely related to R^2.

$$\text{RPD} = \text{Sd}/\text{SEP}, \text{ and } R^2 = 1 - SS_{res}/SS_{tot}$$

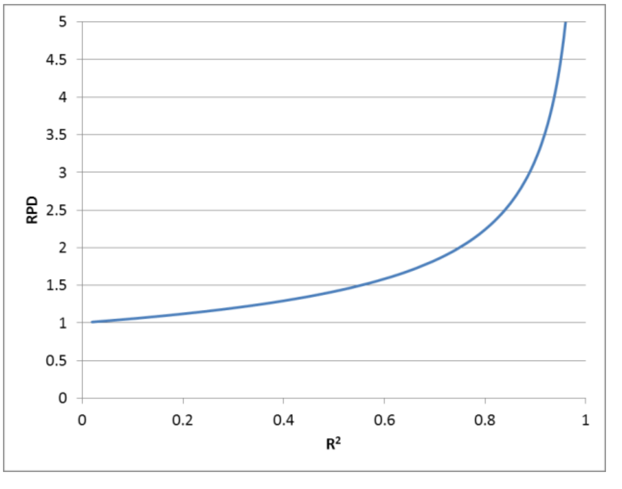
where SEP = standard error of prediction, which is calculated as root mean squared error $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2}$

Sd = Standard deviation of the sample $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$

SS_{res} = Sum of squared error $\sum_{i=1}^{n}(y_i - f_i)^2$

and SS_{tot} = total sum of squares (which is proportional to sample variance) $\sum_{i=1}^{n}(y_i - \bar{y})^2$

Plotting R^2 and RPD, we can see the exact relationship. For a normally distributed variable, and large sample size, $\text{RPD} = (1-R^2)^{-0.5}$.



So we might as well say that if $R^2>0.75$ (equal to RPD >2) the model fits (predicts) quite well and if $R^2< 0.5$ (equal to RPD < 1.4) it doesn't fit so well, rather than using the RPD classification.

Pedometricians, please

1) do not quote both RPD and R^2, they are the same measure,
2) do not use the classification of RPD to justify that your models are excellent or poor, it is no different than using R^2 and there is no basis for this classification. It is all relative!
3) The important measure is how uncertain is the prediction, or what is the prediction interval. This is rarely calculated.

[1] One other myth is that random forests never overfit.

References

Batten, G.D. 1998. Plant analysis using near infrared reflectance spectroscopy: The potential and the limitations. *Australian Journal of Experimental Agriculture* 38, 697-706.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, P., Roger, J-M., McBratney, A.B., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends in Analytical Chemistry* 29, 1073-1081.

Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh C.R., 2001. Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. *Soil Science Society of America Journal* 65, 480-490.

Williams, P.C. 1987. Variables affecting near-infrared reflectance spectroscopic analysis. In: *Near-Infrared Technology in the Agricultural and Food Industries* (eds P. Williams & K. Norris), pp. 143–167. American Association of Cereal Chemists Inc., Saint Paul, MN.

(*Continued from Page 3*)

References

FAO-UNESCO. 1988. Soil map of the World (revised legend). World Soil Resources Reports. FAO. Rome.

World Agroforestry Centre, 2006. Improved Land Management in the Lake Victoria Basin: Final Report on the TransVic project. ICRAF Occasional Paper No. 7. Nairobi. World Agroforestry Centre.

Vågen, T-G., Davey, F.A., Shepherd, K.D. 2012. Land Health Surveillance: Mapping Soil Carbon in Kenyan Rangelands. In P.K.R. Nair and D. Garrity (eds.), Agroforestry - The Future of Global 455 Land Use, Advances in Agroforestry 9, DOI 10.1007/978-94-007-4676-3_22.

More Readings:

(*Leigh has kindly provided some articles about the latest research progress on Kenya soil. Because of the limited space, only the titles are listed here. You may access those articles through the links below the titles.*)

- Landsat-based approaches for mapping of land degradation prevalence and soil functional properties in Ethiopia

http://www.pedometrics.org/papers/Vagen et al 2013 RS Envrionment.pdf

- Mapping of soil organic carbon stocks for spatially explicit assessments of climate change mitigation potential

http://www.pedometrics.org/papers/Vagen and Winowiecki 2013 Mapping SOC stocks.pdf

- Land health surveillance: mapping soil carbon in Kenyan rangelands

http://www.pedometrics.org/papers/Vagen et al 2012_Land Health Surveillance.pdf

Pedometrics 2013 Pre-conference Workshops

Workshop1 (Aug. 26):

Title: Basic Geostatistics with R **Instructor:** Gerard Heuvelink (ISRIC World Soil Information)

Brief Description:

This workshop reviews basic geostatistical methods and shows how these methods are implemented in the R language for statistical computing. It requires no prior knowledge about geostatistics but assumes that participants are familiar with basic statistical concepts such as probability distribution, mean and variance, correlation and linear regression. The main topics addressed are: relationship between spatial variation and semivariogram, semivariogram estimation from point observations, ordinary kriging and regression kriging. The first part of the workshop is a lecture that explains the theory and illustrates it with real-world examples. The second part is a computer practical in which participants execute all steps of a basic geostatistical analysis in R using a soil pollution dataset from the Netherlands.

Contents:

- Lecture: basic statistics and geostatistics, including variogram estimation and ordinary and regression kriging
- Computer practical: introduction to R, basic statistics and geostatistics with R

Workshop2 (Aug. 27):

Title: Spatial prediction of soil variables using 3D regression-kriging - GSIF and plotKML packages for R

Instructor: Tom Hengl (ISRIC World Soil Information)

Brief Description:

This workshop continues with more advanced geostatistical methods that can be implemented in the R environment for statistical and geographical computing, namely 3D regression-kriging (soil properties) and multinomial logistic regression (soil classes). The focus of this workshop is on using R packages for daily work, i.e. operational soil mapping. The lecturer will use two packages for R (GSIF and plotKML for Google Earth) that have been developed for the purpose of automating soil mapping and that have been used for 3D mapping of key soil properties in Africa at 1 km resolution. The first part of the workshop is a lecture that explains the design and main functionality of the GSIF package. In the second part participants will use GSIF and plotKML for 3D regression-kriging of soil properties in the Ebergötzen area, Germany. This workshop will also include a demonstration of how the soil grids at 1 km were derived for the African continent using the WorldGrids repository of covariates and the AfSP database that contains over 15,000 African legacy soil profiles.

Contents:

- Demo: spatial prediction of soil properties and classes for Africa

Workshop3 (Aug. 27):

Title: DSM Using SoLIM **Instructor:** A-Xing Zhu (University of Wisconsin-Madison)

Brief Discription:

SoLIM, Soil-Land Inference Model, is a new technology for digital soil mapping (DSM). It makes use of the state-of-the-art techniques in geographic information processing techniques and artificial intelligence techniques for predictively mapping under fuzzy logic. The focus of this workshop is the operation of SoLIM Solutions 2013 for DSM. SoLIM Solutions 2013 encompasses most of the recent developments in DSM techniques under the SoLIM framework. The workshop will also provide a brief introduction of the SoLIM framework and an operational perspective of the new system in deploying the SoLIM technology, the CyberSoLIM, an effort for bridging the digital divide in DSM. The workshop will be given by A-Xing Zhu and assisted by three other key members of the SoLIM group. The workshop will provide attendees with the full version of SoLIM Solutions 2013.

Contents:

- DSM Using SoLIM: Framework
- DSM Using SoLIM: Operation
- DSM Using SoLIM: New Advances – the CyberSoLIM system

The Working Group on Soil Monitoring

The Working Group on Soil Monitoring was formally established during the last World Congress of Soil Science in 2010. The aims of the group are to encourage inter and intra disciplinary collaborations into the design, implementation and interpretation of soil monitoring networks. An article describing the key issues to be addressed by the group has been published in Pedosphere (Arrouays et al., 2012).

The first activity arranged by the group was a special session at the 2011 Pedometrics meeting in Trest, Czech Republic. The session concentrated upon mathematical and statistical issues of soil monitoring and consisted of seven research talks and a keynote reviewing the outstanding research challenges. The research talks illustrated how diverse threats to soil quality such as compaction and contamination could be monitored.

A symposium addressing more general issues in soil monitoring was arranged at the Eurosoils 2012 meeting in Bari. This tacked fundamental problems such as the design of soil monitoring networks and the requirements and challenges of monitoring key soil parameters such as soil carbon and bulk density. Contributions from this symposium and the accompanying poster session will be included in a special issue of the European Journal of Soil Science which is currently in preparation.

Further sessions have been arranged for the 2013 IUSS Division 1 Congress in Ulm Germany and the 2014 World Congress in Jeju, South Korea. The Ulm symposium is concerned with the interdisciplinary challenges of soil monitoring whereas the Jeju meeting will explore how soil monitoring can benefit mankind and the environment.

In March 2014 an international workshop entitled 'Soil Change Matters' will be organized by the IUSS WGs on Soil Monitoring and Global Soil Change, Soil Science Australia and the Victorian Department of Primary Industries (VDPI). The meeting will be hosted by VDPI in Bendigo, Australia. It will explore the extent to which soil change can be quantified and explained through monitoring and modelling. The meeting will bring together policy makers and scientists to discuss policy needs, the limitations of our current understanding and the implications for future research.

Further details on all of these forthcoming events can be obtained from the WG secretary.

Contacts

Chairman: Dominique Arrouays,
Dominique.Arrouays@orleans.inra.fr

Secretary: Ben Marchant, benmarch@nerc.ac.uk

Reference

Arrouays, D. et al. (2012) Generic Issues on Broad-Scale Soil Monitoring Schemes: A Review. *Pedosphere*, 22, 456-469.

Activities of the Proximal Soil Sensing Working Group (WG-PSS)

The working group has been actively involved in the organising of workshops and sessions at different conferences:

- IUSS WG-PSS session at 19th WCSS, Brisbane 1–6 August 2010
- EGU Session on soil spectroscopy, Vienna 3–8 April, 2011
- 2nd Global Workshop on PSS, Montreal 15–18 May 2011
- WG-PSS session at EUROSOIL, Bari 2–6 July 2012
- 3rd Global Workshop on PSS, Potsdam Germany 26–29 May 2013

We have also produced some publications:

- Book from papers of the 1st global workshop on high resolution soil sensing and mapping, Sydney 2008
- Special issue from papers in EGU session 'Diffuse reflectance spectroscopy for soil and land resource assessment, Vienna 2009
- Special issue from papers of 2nd global workshop on proximal soil sensing, Montreal 2011 - January 2013
- Special Issue based on topic 4 of Eurosoil 2012 'Advanced Techniques and Modelling' Papers being processed now
- A special issue is being considered to report the science presented 3rd Global Workshop on PSS, Potsdam Germany 26–29 May 2013

We have a website with general information on the WG and on PSS as well as our meetings: www.proximalsoilsensing.org

The Third Global Workshop on Proximal Soil Sensing 26-29 May, 2013, Potsdam, Germany.

The workshop was organised to bring together global community devoted to advancements in technologies related to measurements by sensors placed in proximity to the soil being tested. As with the previous workshops, it was held under auspices of international union of soil science (IUSS) working group on proximal soil sensing (WG-PSS). Locally, it was organised by Leibniz-Institute of Agricultural Engineering (ATB), Leibniz-Institute of Vegetable and Ornamental Crops (IGZ Großbeeren), and the University of Potsdam (Potsdam, Germany). Over 80 researchers from various disciplines and 23 countries were present. The workshop included two full days of presentations, a field demonstration event and a number of networking and discussion sessions.

European Geosciences Union, Soil System Sciences Division

At its Annual General Meeting at the 2013 European Geosciences Union (EGU) meeting in Vienna the Soil System Sciences Division of EGU voted to establish a new subdivision called Soils: Informatics and Statistics. This new subdivision will be organizing sessions for the 2014 EGU convention in Vienna (27th April–2nd May 2014). While the programme has yet to be approved and finalized, the following proposed sessions might be of particular interest to pedometricians.

- *Communication of uncertainty about information in earth sciences*
- *Sampling in space and time*
- *Modelling and visualization: new informatics tools for soil science*
- *Digital soil sensing, assessment and mapping: novel approaches for spatial prediction of key soil properties and for spatial assessment of soil functions*
- *Soil system modelling: strategies and software*
- *Complexity and nonlinearity in soils*
- *Scaling Connectivity*
- *Dynamic Soil Landscape Modelling*
- *Soil mapping perspectives: soil spatial information for land management and decision making*
- *Carbon sequestration in agricultural soils: the need for a landscape scale approach*
- *Soil Apps.*

Sessions for the 2014 congress will be announced later in 2013. For more information about these, and about the new subdivision, visit http://gsoil.wordpress.com/ or contact the Subdivision Chair mlark(at)bgs.ac.uk

Flyers

Lightning image © David Marland

Soils change in response to land use, land management and climate. Understanding the mechanisms and rates of change in fundamental soil properties, and their extent across the landscape, is critical for management of soil and land to ensure enduring productivity and provision of ecosystem services.

The workshop will bring together scientists who can explain the critical changes in soils, particularly during the past century of increasingly intense land use. The workshop will include dialog between policy makers and scientists to clarify policy needs as well as the current capability of soil knowledge systems and soil monitoring approaches to fulfill them.

This international workshop is organised by the Victorian Government's Department of Environment and Primary Industries and supported by Soil Science Australia and the International Union of Soil Sciences.

Key Dates

Call for Papers — *Open now*

Papers Due — *30 Sep 2013*

Early bird closes — *16 Dec 2013*

Soil Science Australia

City of Greater Bendigo

International Union of Soil Sciences

Provisional Program

Monday 24th March

	Optional field trip (limited numbers)

Tuesday 25th March

All day	Soil change workshop technical presentations
	Soil Matters Symposium * (general audience)
Evening	"Soil Matters" (free event open to the public)

Wednesday 26th March

All day	Policy and Science workshop
	Soil change workshop technical presentations
Evening	Workshop dinner

Thursday 27th March

Morning	Soil change technical presentations & Plenary
Afternoon	Working group meetings and workshop

*The Soil Matters Symposium is a one day event comprising a series of presentations from national and international authorities on soil science as well as local and regional examples of soil issues and management. A 1 day registration is required for this event ($150).

Workshop Purpose

The workshop is designed to bring together practitioners in policy, science and management to address the questions –

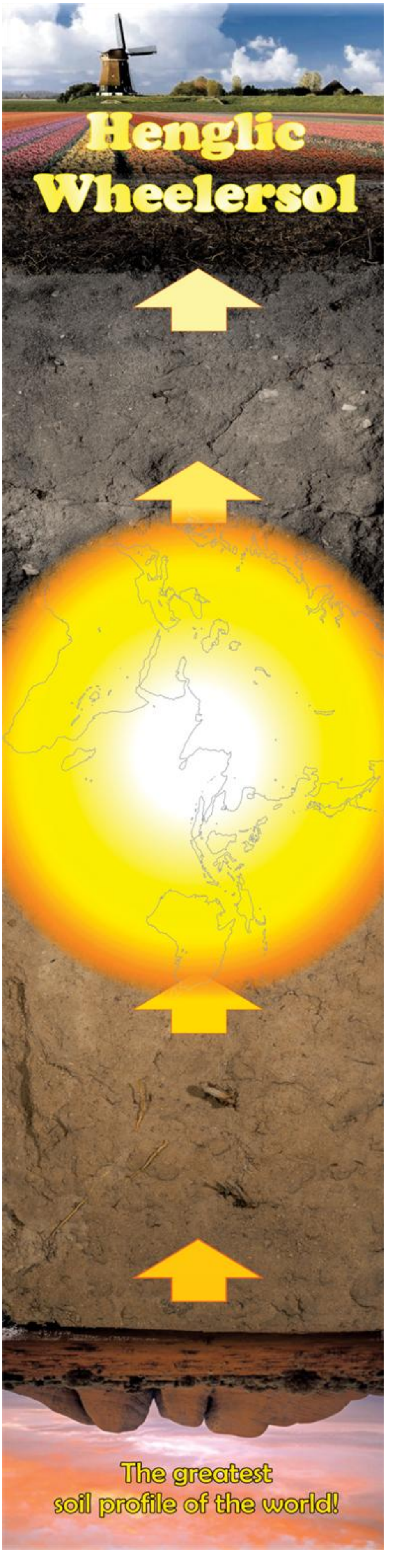
- What is changing in soil?
- Why does it matter (impact, extent, critical thresholds, system collapse) or does it? Who cares and who pays?
- How do we measure change and are we monitoring important changes?
- How fast are changes occurring and expected to occur?
- Should anything be done about these changes - science understanding,? Technical ability to measure and monitor change? Policies for intervention or prevention?

Further information and registration: www.soilmatters.org

Department of Environment and Primary Industries

Victoria

The first pedometric marriage, fusing European and Australian soil in the greatest soil profile in the world!!!



Pedomathemagica

Problem 1 (not difficult)

During a field campaign collecting validation data, pedometrician Ganlin runs into a farmer. "You are a clever and educated guy", says the farmer, "so you should be able to solve my problem". The farmer explains that he has four sons who are quite jealous about each other and who constantly quarrel about the management of their father's land (see shape below). He decides to divide the land give each of his sons an equal share, but this must be done in a way that not only the area of each of the four resulting parcels is the same, but also their geometry. How to do this?

(from Gerard Heuvelink)

Problem 2 (difficult)

One of the problems with legacy soil profile data is that their geographic coordinates may not always have been recorded properly. Recently, legacy data officer Johan came across a very odd recording. The metadata stated: "Go to the drinking well and walk from there to the old elm tree, measure the distance, take a right turn (90 degrees), walk the same distance and mark the point that you get to. Go back to the drinking well, walk to the pine tree, measure the distance, take a left turn (90 degrees), walk the same distance and mark this point too. The profile is located exactly half way the straight line connecting the two marked points." One day Johan happens to be in the neighbourhood and decides to trace the position of the soil profile. Arriving at the spot, he finds that the elm tree and pine tree are still there, but that the drinking well did not survive the passage of time: it had completely disappeared. Johan is still wondering about and looking for the location of the profile, can you help him?

(from Gerard Heuvelink)

Problem 3

Alf and Bert the soil surveyors are about to go on a trip to map a remote corner of Ruritania. Alf does all the planning, but the weekend before they are due to travel he wins a big mathematical bet off Flossie the barmaid and wakes up on Monday morning with a severe hangover. Bert has to pack the Landrover, using Alf's notes. All is well until he comes to the equipment list which you can see below.

2: Augers
3: Spades
5: Buckets
7: Plane Tables
11: Laptops
13: GPS
FTA-Code for Ruritania survey: 1486485000

Pedomathemagica

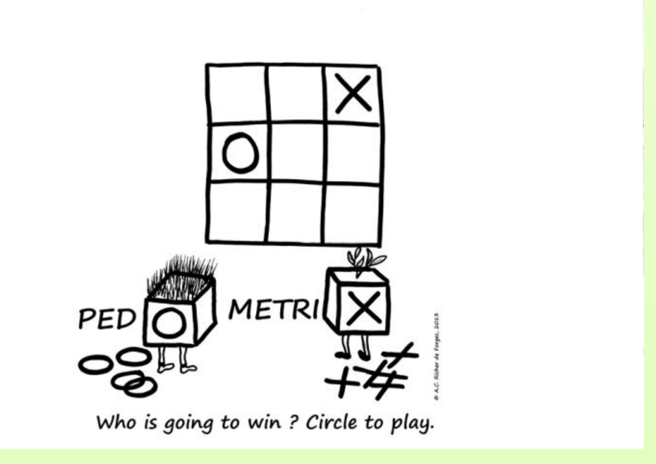
Bert collects 2 augers, 3 spades etc. but realizes that all is not as it seems when he gets to 7 plane tables. Alf only ever takes one plane table when soil surveying, which he likes to have in case the GPS should break down or he requires shelter from particularly heavy rain. Furthermore, why would Alf want 13 GPS? The institute doesn't have that many and, anyway, regulations only allow one GPS to be taken on any one expedition. The numbers cannot mean what they seem to mean, and what does "FTA-Code for Ruritania survey" mean? It's not the project number, it's not a set of co-ordinates, it's not even the budget in Ruritanian Reuros (too small a number for that). Bert turns the list over and reads:

> *Bert,*
> *Get it right and I'll buy all the Ruritanian lager. Get it wrong and ....*
>
> *Alf*

Can you work out how many of each item on the list Alf intended to pack?

(from Murray Lark)

Problem 4



1) If circle is clever, whatever cross is (only one answer)

2) If circle and cross are stupid and choose the location randomly (give a probability)

3) If circle is stupid and choose the location randomly and cross is clever (give a probability)

The first winner will receive a unique tee-shirt with the drawing.

(from Dominique Arrouays and Anne Richer de Forges)

Pedomathemagica

Solutions for Pedometron 32

Problem 1 (easy)

Answer: There is as much red wine in the white barrel as there is white wine in the red barrel. This is simply because at the start there was as much red wine as there was white wine, and so it just has to be that all the volume taken up by the red wine in the white barrel must be equal to the volume taken up by the white wine in the red barrel, because no wine got lost. We can also show it mathematically: The amount of litres of red wine in the white barrel is $1 - \frac{1}{21} = \frac{20}{21}$. The amount of white wine in the red barrel is $1 \cdot \frac{20}{21} = \frac{20}{21}$.

To answer the second question let $R(k)$ be the amount of red wine in the red barrel after k mixings. We then have that: $R(k+1) = \frac{19}{20}R(k) + \frac{1}{21}\left(20 - R(k) + \frac{1}{20}R(k)\right) = \left(\frac{19}{20} - \frac{1}{21} + \frac{1}{420}\right)R(k) + \frac{20}{21} = \frac{1}{21}(19R(k) + 20)$. Since $R(0) = 20$ we can calculate how large k should be until $R(k) = 0.51 \cdot 20 = 10.2$ or smaller. This happens at $k = 40$. Philippe has a lot of work!

(by Gerard Heuvelink)

Problem 2 (medium)

Answer: Philippe can be certain that Marc will have to pay the bill by playing it cleverly. This problem was in fact quite difficult. It is a variant of the Chinese Nim game, see https://en.wikipedia.org/wiki/Nim for an explanation of the game and a winning strategy. The winning strategy is best explained by writing the number of beer glasses of each row in binary format:

001
010
011
100
101
110

Adding the number of ones in each column gives 3 – 3 – 3. What Philippe should do is remove that many beers such that the sum for each column becomes even. He can do this by drinking five beers from the bottom row, which gives:

001
010
011
100
101
001

This gives a sequence of sums 2 – 2 – 4, which are all even. Whatever is Marc's next move, each time Philippe must make sure that after his move the sum of the number of ones in each of the columns of the binary representation is again even. This is always possible. With this strategy, Philippe is guaranteed to win. Try it out yourself with your friends or colleagues and impress them! More details and explanation of the Sprague-Grundy theorem (which states that "every impartial game under the normal play convention is equivalent to a nimber"!) at https://en.wikipedia.org/wiki/Nim.

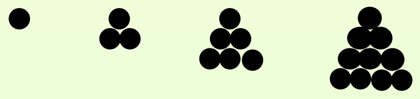
(by Gerard Heuvelink)

Pedomathemagica

Problem 3:

Answer: The number of presents received on day n is the nth triangular number Tn. A triangular number is the number of points in one of the series of triangular arrays created by adding, at the nth step, a row of n points to the base of the previous triangle in the series, so that the first few terms of the series are:

n:	0	1	2	3	4
T_n:	0	1	3	6	10



I always find it easier to think of the nth triangular number as the number of elements on or above the diagonal of an $n \times n$ matrix, i.e. $\frac{n^2-n}{2} + n = n(n+1)/2$. Easy, and who needs R?

Before thinking about the second question, note that if we double the nth triangular number T_n we get $n(n+1)$ which is called the nth *pronic* number.

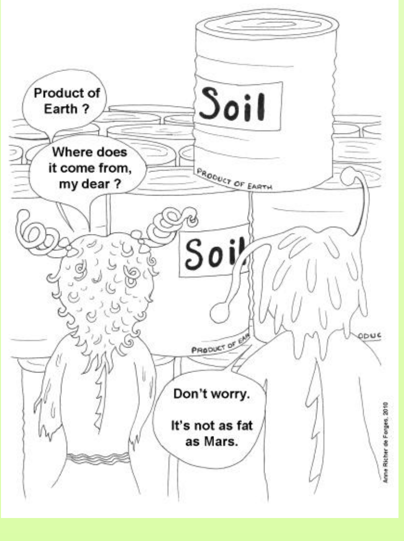
Now, since the two integers provided by Flossie are odd, and not equal to each other, we can write the larger and the smaller as $2i+1$ and $2j+1$ respectively where i and j are integers and $i>j$. The number of pints that Guinevere orders is therefore:

$$(2i+1)^2 - (2j+1)^2$$
$$= \{4i^2 + 4i + 1\} - \{4j^2 + 4j + 1\}$$
$$= 4\{i(i+1) - j(j+1)\}.$$

Now, note from above, that the two terms in the braces are the ith and jth pronic numbers, and so the number of pints is

$$= 4\{2T_i - 2T_j\}$$
$$= 8\{T_i - T_j\}$$

Since the triangular numbers are obviously all integer, it follows that, what ever odd numbers Flossie provides, the number of pints is exactly divisible by eight, so Alf is certain to win the bet. I found this question (without the answer, I should point out) in a maths examination paper set in Oxford in 1854 and printed in facsimile in *Mathematics in Victorian Britain* by Flood, Rice and Wilson (OUP, 2011).



(by Murray Lark)