

## From the Chair



Dear Pedometricians,

Welcome to another edition of *Pedometron*. Once again we have a diverse and informative set of articles for your perusal, and I hope that they will inspire you in your work, and also encourage you to contribute something to these pages.

A recent paper in the *European Journal of Soil Science* has the intriguing title *Visions of a more precise soil biology* (Andr en *et al.*, 2008). In this paper the authors set out a critical account of what they see as the major problems that soil biologists face in developing their science, a science that is challenging because, in the words of the paper

"certain inherent properties of soils and biological systems (large heterogeneity, many levels of physical structures including micro-sites, the difficulty of making direct observations especially at depth, many different species that are difficult to extract and/or do not grow on ordinary laboratory media and the wide divergence of landscapes) make it difficult to perform scientific investigations with great precision."

The paper focuses less on particular technical issues than on aspects of the general culture of contemporary science which the authors think are obstructive. For example, they observe that the inertia by which old and outworn paradigms retain their hold on the scientific community is due, at least in part, to "the increased workload and productivity demands on scientists .... you are too busy to take in new information and therefore stick to what you know."

Of particular interest to pedometricians are the comments that are made on statistics. The paper gives an example of the difficulty of demonstrating the effect at field scale of processes that can be easily demonstrated in laboratory measurements, the particular example is the effect of earthworms on the rate of decomposition of organic material in soil. They admit that a satisfactory test of the hypothesis that earthworms increase the rate of decomposition has

not been achieved at field scale. They propose that one approach to this problem would be through meta-analysis of published studies, but admit that the principal problem with this is that field studies in which no effect of earthworms on decomposition is found are unlikely to be published because the result is both unexciting and counter to most scientists' expectations. They go on to argue for more and better statistical evaluation and input into research projects and papers. That is all very well, but it seems to me that their central problem requires something more than sound statistics, rather it requires new thinking on how processes, which operate at different spatial scales, are studied and quantified *in situ*. There may well be real effects of earthworms in the field experiments (it would be odd if there were not), but other factors, operating at different spatial scales, obscure these. To disentangle the effect of the worms in the field requires sampling and experimental designs which allow scale-specific effects to be detected and estimated. As pedometricians we have some ideas on how this might be done. What is needed is a dialogue with the soil biologists, in which pedometrics/statistics is not regarded as a service discipline, along with nucleic acid sequencing and soil chemical analy-

### Inside This Issue

From the Chair	1
Call for Best paper 2007	2
Report from HRDSSM Sydney	3
Report from EGU Vienna	6
Another Journey on the Road to Pedometrics	7
Soil Bibliometrics	10
Headbanging	12
Statistical Distribution of Humus	17
Upcoming Events	21
Alex's Preferred Pedometrics Paper	22
Book Review	24
Profiles	25
Pedomathemagica	26

sis, but rather as a collaborative field of science with relevant experience to share.

Part of the difficulty is illustrated by an example of statistical mispractice which Andrén and colleagues call the "thesis syndrome". They state that a thesis, based on work in which 200 hypotheses are tested, with  $P=0.05$  as the criterion of significance, would be expected to report 10 cases of significant findings simply through random variation (a healthy number of positive results on which to build a thesis). Over a scientist's career, they observe, a number of false positive results (so-called type 1 errors) will be reported, along with false negatives "even at the 1% or 0.1% significance levels."

Well, yes and no. It is certainly the case that a set of 200 tests on random numbers is expected to return 10 results with  $P<0.05$  by definition. This is part of the reason for the concerns about data mining that have been expressed by statisticians, and that I presented at the Pedometrics meeting in Tübingen last year. It is also true that a set of multiple tests (tests on separate, not necessarily independent hypotheses) cannot be presented as significant with  $P$  at some threshold used on all tests independently. This is the motivation for multiple testing procedures such as Bonferroni's approach, or control of the False Discovery Rate. However, I would want strongly to resist the notion that the  $P$ -value at which null hypotheses are rejected determines some expectation of the proportion of false hypotheses that are incorrectly reported as true in the course of a thesis, scientific programme or scientific career (*Mistakes, I've made a few, but then again, on average only five percent of the time.*) Behind this interpretation of significance testing lies the dead hand of the approach to statistical inference advanced by Neyman and Pearson, all too often confused with the approach of Fisher who first developed the use of  $P$ -values for inference.

Neyman and Pearson proposed that no probabilistic inference can be made about a particular experiment but rather "Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them, in following which we ensure that, in the long run of experience, we shall not often be wrong." (Neyman and Pearson, 1933). This approach can be regarded as part of a broader project to restrict scientific activity to purely deductive tests of hypotheses, a project in which the (in my view much over-rated) work of Karl Popper was a leading part (Goodman, 1998).

Fisher, however, would have none of it. He wrote: "The concept that the scientific worker can regard himself as an inert item in a vast co-operative concern working according to accepted rules, is encouraged by directing attention away from his duty to form correct scientific conclusions, to summarize them and to communicate them to his scientific colleagues, and by stressing his supposed duty mechanically to make a succession of automatic

'decisions'...The idea that this responsibility can be delegated to a giant computer programmed with Decision Functions belongs to a phantasy of circles rather remote from scientific research" (Fisher, 1973).

For Fisher science is not an automated method for processing and selecting among random hypotheses, but rather relies on the responsibility of the scientist to form appropriate, plausible and fruitful hypotheses, the evidence for which is then weighed by the  $P$ -value. This is one reason why statistics can never function satisfactorily as a pure service facility for science, but should be fully integrated into the development of science, particularly of branches such as soil science in which heterogeneity in time and space is all-pervasive.

Andrén, O., Kirchmann, H., Kätterer, T., Magid, J., Paul, E.A. and Coleman, D.C. 2008. Visions of a more precise soil biology *European Journal of Soil Science*, 59, 380-390 doi: 10.1111/j.1365-2389.2008.01018.x

Fisher, R.A. 1973. *Statistical methods and scientific inference*. 3rd ed. Macmillan, New York.

Goodman, S.N. 1998. Multiple comparisons explained. *American Journal of Epidemiology* 147, 807-812.

Neyman J, and Pearson E. 1933. On the problem of the most efficient tests of statistical hypotheses. *Philos Trans R Soc A*. 231, 289-337.

Murray



# Report from the 1<sup>st</sup> Global Workshop on High Resolution Digital Soil Sensing and Mapping



by Liesbet Cockx



The first months of 2008 burdened me with rather conflicting feelings. On the one hand, the first workshop exactly matching my PhD interest was going to be organized in Sydney, causing quite some excitement in my scientific mind. On the other hand however, I had to be separated from a lovely six-months-old daughter. But regardless the strong maternal instinct, I can now conclude that this first Global Workshop on High Resolution Digital Soil Sensing and Mapping was not to be missed! The workshop had a very interesting program with a wide range of soil sensor topics, brought by an international group of scientists (photograph by Nathan Odgers).

After a warm (but rainy) welcome, the workshop was kicked off by an interesting keynote of J.J. De Gruijter on Sampling for High Resolution Soil Mapping. He brought a clear message on the distinction between sensor sampling and calibration sampling, each with a different aim and a different sampling pattern. Sensor sampling needs sufficient coverage in the geographic space which was optimally obtained using a square grid, while calibration sampling should focus on choosing a pattern in the attribute space.

During the rest of the first day, all types of soil sensors were covered, going from remote field spectroscopy over physical-mechanical soil sensors to EMI and EC sensors. The seven-minute talks worked as a tornado, touching a lot in a short time and supplying us with loads of material to discuss.

The second day we were all immersed in a refreshing and complete overview on the Development of On-the-Go Soil Sensor Systems by Viacheslav Adamchuk. Starting with relating the human senses to the types of soil sensors, passing over the principles behind the sensors, he ended with conclusions that were going to be the fads and fancies of this workshop. Multi-sensor fusion will provide the ability to separate the effects of individual soil properties while at the same time site-specific calibration and validation remain essential steps in the soil mapping procedure. Alex McBratney gave us a seven minutes history class to acknowledge the first on-the-go soil sensor built by Haines and Keen, already in





1925. So despite our massive sensor - enthusiasm, in fact there is “nothing” new under the sun.

Bo Stenberg discussed in his keynote both the potential of spectroscopy for soil applications as well as the calibration-issue and future research suggestions. The last keynote by Noel Cressie dealt with kriging very large datasets; he solved the problem of inverting an  $n \times n$  matrix by fixed rank kriging (and inverting an  $r \times r$  matrix).

Besides three intensive scientific days, we also went on a field excursion where we were exposed to the Australian sun and to a wide range of proximal soil sensors. At the experimental farm, field demonstrations and commercial presentations were held. The obtained sensor measurements were displayed and used as a basis for a nice discussion.

The informal side of the workshop is worthwhile mentioning as well: every evening local pubs were explored and I was even introduced to a

non-Belgian best beer of the world called “Little Creatures”. The workshop dinner was held in the Waterfront Restaurant, with splendid view on the Opera House and the Harbour Bridge: an entertaining combination of socializing and sight-seeing. To conclude I would like to thank Raphael, Alex and the complete team for the interesting program of this workshop and its excellent organization!

For more photos, see:

<http://picasaweb.google.co.uk/pedometrics/Hrdsms>  
and <http://picasaweb.google.com/nathanodgers/HRDSSMWorkshop>.





- 3) The speaker's energy is also intensified, the delivery more passionate and engaging.
- 4) The blowhard in us all is truncated. Speakers are forced to distill their message to the most valuable and salient points, no more than one or two.

by Brent Myers

What an excellent meeting my friends.... The 1st Global Workshop on HRDSSM was a great start for a new workshop tradition. There have been and continue to be other venues for bringing high-res soil sensor and mapping scientists together, but a pure focus on these topics in an intensive workshop setting is needed.

This need was obvious in the discussion and in the questions that were raised at the meeting. These centered on the fundamental problems, yet unsolved, such as global versus local and direct versus indirect sensor calibrations, reference method accuracy, sensor fusion, the ubiquitous and troublesome covariance of water content with almost any soil property, and on from there. The focus of a global scientific community assailed these problems in an exceptional array of presentations and deliberations.

The venue was very nice and service was excellent, especially at tea. The lunches were delightful - light and crisp fare. They brought the pleasures of sitting with friends and colleagues in the grass or in the shade of a eucalyptus tree. Evenings at the pub were every bit as important as the formal program - so many interesting people. And no one who attended will forget the magnificent view of Sydney Harbor, downtown Sydney, and the Sydney Opera House as seen from the conference dinner at The Rocks. These non-scientific intangibles produced very memorable impressions. I can almost be there as I write this.

A most remarkable feature of the workshop was its format. Except for keynote presentations, talks were limited to seven minutes. Seven minutes! I must say I was skeptical...and worried! But here's why it works:

1) Lengthy introductions are obviated. Workshops and small conferences are topically focused. Those attending know the theory behind the methods and technologies. Keynotes can provide the topical introductions.

2) With compressed introductions we get straight to the meat. The attention of the audience is focused and intensified by the pace. There is no time to drift away.

5) If you need to say more, or need to see more to understand, the paper is there for added detail.

6) The seven minute format is time efficient. More talks can be included and still expand the time for discussion. The value of the meeting is effectively doubled.

7) Finally, discussion is emphasized. The talks become fuel for the really important part - the community dialogue about the promising methods, difficult problems, and research needs.

I'll conclude my portrayal of the HRDSSM workshop by advertising Budiman's shopping list of soil physical sensing. It came in a discussion after a session on physical and mechanical sensors.

#### Budiman's Shopping List

- Particle size
- Bulk density
- Available water capacity
- Hydraulic conductivity

We all have a similar list. I call my list the Killer Applications of Soil and Landscape Modeling. Geeks are always talking about 'killer apps'. These are software and hardware solutions that everyone adopts: e-mail, spreadsheets, word processing, cellular (mobile) phones, iPods, the internet, etc. I'd suggest that the items in Budiman's list are the killer apps that soil sensor technologies need to provide in order to be adopted. There are more things to be added to this list, it focuses mainly on indirect calibrations, but it provides an organizing principal for the future of high resolution digital soil sensing and mapping.

To the organizers and all in attendance... thanks for a great workshop.





Thomas Mayr on Digital Soil Assessment

## *Budi & Murray*

The European Geophysical Union (EGU) held its annual General Assembly in Vienna 13-18 April 2008. The Soil System Sciences programme sponsored the Digital Soil Mapping (incorporating the geophysical measurement) session. The session took the whole of Friday with 17 oral presentations and 19 posters. Thomas Mayr started with a keynote on Digital Soil Assessment. We were reminded on the final output of DSM is not merely a map, but has to be linked with applications such as soil threats. Clearly an advancement beyond static map is required for evaluating the changes in soil and was echoed in most presentation and discussion.

We are also introduced to a couple of new EU projects on using geophysical measurement for soil: DIGISOIL and iSOIL (in competition with Alfred's iPodzol). We are certainly eager to see the developments and results in the coming years. The oral presentation featured a range of Digital Soil Mapping examples with a variety of approaches from various countries: Northern Ireland, South Korea, India, Brazil, the Netherlands, Belgium, Russia, USA, Austria, Australia, Romania, Spain, Czech, Switzerland, Italy, Hungary, Morocco, Egypt, Germany, France, and the world. This is certainly a good representation. The session went really well with a good turn-out of attendees and some interesting discussion. There were great interactions with geophysicists and soil scientists, modellers, and remote sensing experts. Here we see the merging of disciplines such as soil physical measurement, geophysical measurement, hyperspectral remote sensing, and digital terrain modelling.

With more than 8200 attendants to this meeting, it is not easy to meet someone. It seems that the EGU has embraced many soil scientists in their program, the

chair of the Soil System Sciences (SSS) noted that this is the largest SSS meeting ever in EGU with more than 580 abstracts in 25 sessions (228 orals and 358 posters). Perhaps in the future it will keep growing with more soil scientists from Europe especially Eastern Europe joining the meeting.

The Soil Science System offered a range of topics from fauna effect on soil properties to Iodine and Selenium in soils. There were a couple sessions of interest to pedometricians: Scale, Scaling, and Nonlinearity in the Earth's Surface, Soil and Solid Earth, and Diffuse reflectance spectroscopy in soil science and land resource assessment. The advantage of such meeting is that we can listen to presentations from geosciences on uncertainty, modelling, Mars, etc. and also lectures such as the Philippe Duchaufour Medal by Alain Ruellan, and the John Dalton Medal lecture by Mike Kirkby.

It has been a successful meeting for DSM, and the people from EGU-SSS were clearly impressed and were talking about doing something similar next year! Finally thanks to Florence Carré for all of her very hard work in putting the session together. And also to Nunzio Romano for initiating the session.



# Another journey on the road to pedometrics

Margaret Oliver

An interest in soil that began with the Chernozem in my sixth form at school has never left me. The University of Bristol seemed just the right place to pursue this interest with a degree course in Geography and Geology starting in 1962. I was especially lucky to be taught by Len Curtis, an inspiring lecturer. He took us out into the field most weeks, and through this I became increasingly fascinated by the seemingly infinite variation in soil. I am not alone in having continued with a career in soil science from Bristol; others you may know include Frank Courtney, Stephen Northcliff, Steve Reynolds, Steve Trudgill and Steve Staines.

Towards the end of my time at Bristol Professor Peel invited me to study for a PhD in there. I was to be married soon after graduating in 1965 and would be living in the West Midlands, and so he asked Professor Roland Moss to supervise me at the University of Birmingham. In the event, and with some regret, I decided not to pursue the offer. Nevertheless, it was the right decision for me and it led to unexpected fortuitous outcomes in the long run. The vacancy at Birmingham was taken instead by John Norris whom I met a couple of times at lectures at the University. John's research was on the application of multivariate analysis, and on completion of his PhD he went to work with CSIRO in Australia where he sadly met an untimely death in a road accident. At this stage I could have had no idea that John's death would involve someone who would have a large impact on my life, or that in following John's work later would lead me to this person.

After producing two children and part-time teaching and lecturing, I decided in 1977 to pursue my original aim of doing research. I went to Birmingham University where to my surprise Roland Moss became my supervisor. He had a lively group of PhD students, and the weekly tutorials were stimulating in those early days of my return to academia. The Nature Conservancy Council, which managed the Wyre Forest near to Bewdley in the West Midlands of

England, wanted information on the soil there because it had never been mapped in detail. There was an old geology map of the area, but no one was sure how accurate it was. Roland suggested I should map the soil by applying the USDA's Seventh Approximation, as it was known then. After reading the documentation I realised that the task would be impossible, and so I decided to follow John Norris in applying a multivariate statistical approach instead. My rebellious streak left me having to make many subsequent decisions by myself. Even at this early stage I was aware of how important a sound approach to sampling was for later statistical analysis. I was fortunate in having been taught statistics at Bristol by David Harvey whose lectures to both students and staff were part of the quantitative revolution at the time. After my long absence from the study of soil, I thought it prudent to consult the Soil Survey of England and Wales which had a regional office in Wolverhampton. Mike Hodgson was in charge, and who better could I have had to advise me on survey technique? We talked endlessly about sampling and multivariate analysis. He was quite sceptical of what I was planning to do, but this only spurred me on even more, of course! Nevertheless, his scepticism made me think hard about what I was going to do, which is such an important part of planning any survey, piece of research work or a PhD. Mike dispatched me into the field with John Hollis, another great soil surveyor with a deep interest in soil classification. I learned a lot from John while revising my field skills, but at the same time I was thinking about the weaknesses of the conventional approach to classification.

In my first year in the Wyre Forest I, with the help of a technician, dug and sampled 201 pits. Even though this was hard work, the Forest was a wonderful place in which to work with a lot of wild life and impor-





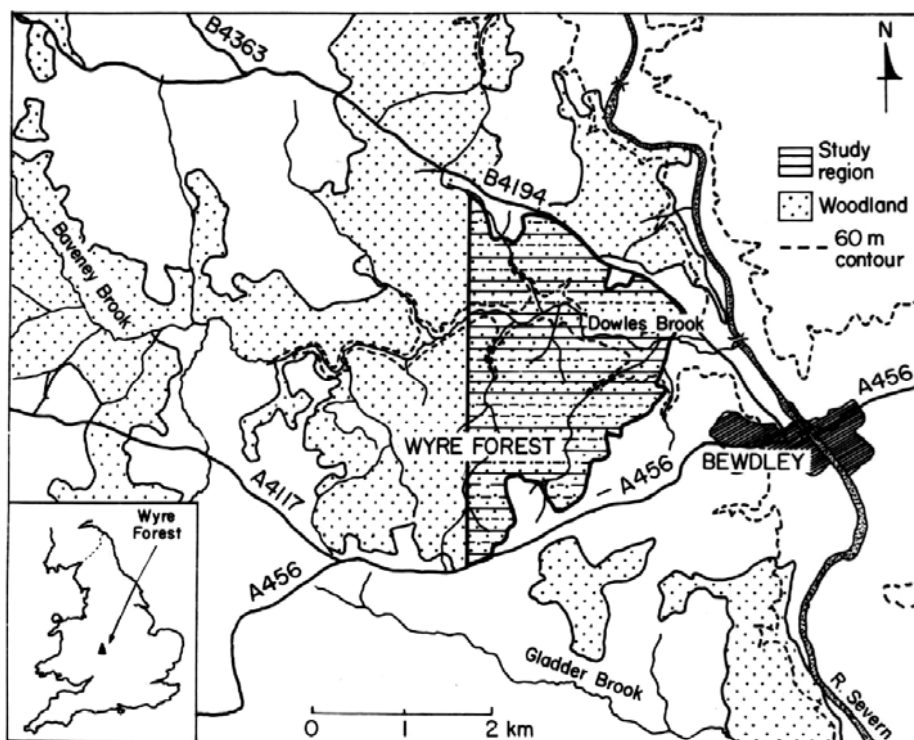
analyses, whereas others could be used for all analyses. The advisory staff at Birmingham's computer centre guided me in using SPSS and GenStat, but not on the suitability of certain variables

tantly very interesting soil. Every soil pit seemed as exciting as opening a new hand of cards in bridge (bridge players will recognise that rush of adrenalin as a challenge is presented by each new pattern in a hand). The variation in the soil there amazed me, and there seemed to be no relation between adjacent pits. I was unaware that the Forest would provide exactly the right place to test methods that I had not even heard about in 1978. With data on 603 soil samples, my aim was to explore them by several multivariate methods.

While still doing fieldwork I read Richard Webster's first book on Quantitative and Numerical Methods in Soil Classification and Soil Survey, published in 1977. At the time I had no idea what the long term effect of this book would be! Even with it at hand, however, I could not see clearly how to deal with some of the many variables I had (78 for each profile) for my analysis. Some properties could be used for some

for a particular analysis. After much head-scratching about how to analyse my data Mike Hodgson encouraged me to contact Richard Webster for guidance. For some inexplicable reason I was reluctant to do this - I must have had some premonition that this would lead me on a difficult path. By now I had also realised some potential deficiencies in my sampling scheme because there seemed to be no relation between adjacent soil pits in terms of their properties. I knew nothing of spatial dependence at that time, and I had not set out to do a spatial study. However, I had recently read Stephen Nortcliff's paper in the Journal of Soil Science (1978) in which he used a nested sampling design to determine the approximate scale of spatial variation. I discovered that Richard had been involved in this work, and so I now had two reasons for contacting him. I wrote him a long letter asking his advice about my analyses. In his usual efficient style, he replied by return of post suggesting a meeting with him, much to the chagrin of my supervisor. I

was full of trepidation as I drove to the Weed Research Organization at Yarnton on the outskirts of Oxford. I was directed to an office that from memory seemed to be down a dismal corridor to be greeted by someone much younger than I had expected with a mass of blond hair and wearing sandals on a cold March day in 1979. At this meeting, which resembled a viva voce examination, we discussed ways in which I might analyse my data to advantage. They would not be straightforward. Later in the afternoon I met Alex McBratney for the first time, and I was quite innocent of the how





my world was about change for the next 30 years! I didn't know at the time that Richard, Trevor Burgess and Alex were working on the first publications of geostatistics for the Journal of Soil Science (1980 - 1981). I left Yarnton and drove home with my head reeling with thoughts and mixed feelings about the day - overwhelmed with the possibilities and daunted at having to find my way through this particular forest alone.

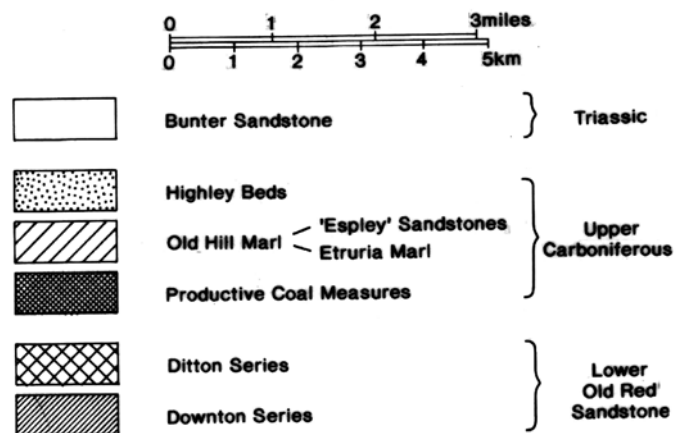
I settled down to tackle principal components analysis, numerical hierarchical and k-means clustering, which were so difficult at that time because from Birmingham all multivariate jobs had to go to the Manchester computer centre. The whole process was slow; one mistake in a line of code meant a delay of several hours before it could be submitted again.

I met Richard again at a Rothamsted open day in the summer of 1980. Over lunch we chatted about life after my impending qualifying MSc viva for which he had been appointed external examiner. My supervisor had left for another University by now, and, after a penetrating examination, Richard volunteered to supervise my research to the second stage. The next three years were exciting and hard work - I was working part time because I also wanted to give my two young children as much of my attention as possible.

I was now keen to understand why the soil in adjacent pits was so different. The discovery that the first variograms I computed from my first survey of the Wyre Forest were pure nugget was disappointing, but not entirely surprising. I had realised intuitively before my first meeting with Richard that the soil profiles I had exposed appeared to lack any spatial relationship with one another. I needed to explore the spatial scale by a nested approach, and my second survey was the first application of an unbalanced nested design to soil survey. This was an illuminating component of my research, which was followed by transect surveys. These two surveys generated more than 1500 samples. A daunting set of data resulted for the nested and geostatistical analyses that were to follow. The analyses and model-fitting were done on the slow computers of the time where there were long delays between submitting jobs and obtaining the results. It is quite amazing that several models can be fitted in the blink of an eye now on a PC.

I continued to work with Richard after completing my PhD, and we have had an exciting and productive working relationship over this time with many diverse problems to tackle. With the

help of Christian Lajaunie at the Centre de Géostatistique we tackled the analysis of the risk of children's developing cancer. This was a difficult subject, but fascinating and worthwhile. In 1994 I moved from Birmingham to Reading where I had 10 happy and fruitful years and built my own research group which included Benjamin Warr, Zöe Frogbrook, Samantha Baxter, Ruth Kerry and Sarah Rodgers. Ben spent time at the Centre de Géostatistique and has remained in Fontainebleau. With Zöe, Samantha and Ruth I entered the field of precision agriculture for which geostatistics is proving so useful. It is gratifying to think that as I slowly retreat from the mainstream of research there are some young and able researchers to take the subject forward. Zöe and Ruth continue to explore new analyses in soil science, and Samantha is taking a break to look after baby Stephanie. Sarah, after working on malaria in Mali, is established in the field of spatial epidemiology, which I could have also found so interesting. I have thoroughly enjoyed my academic life, even though I came to it late. There has been a strong structural component to my research. The apparently random component of meeting Mike Hodgson and Richard, however, has probably had the greater effect, and that seems entirely appropriate for a geostatistician!



In this article, we would like to introduce a bibliometric measure called the *Degree of Separation*. We can define the Degree of separation (dS) between two articles as the number of papers that need to be sought for the two articles to “meet”.

For example: We searched for the linkage between published work on on-the-go soil strength measurement. As we showed in the previous issue of *Pedometron* (No. 23), the first work is from Haines and Keen (1925). After this paper there seemed to be no paper on this subject for a while until Stafford and Hendrick (1985, 1988) and with the advent of site-specific soil and crop management. We searched for some key papers on the on-the-go soil strength measurement and

tabulated them in Table 1. We calculated the degree of separation dS between the papers. For example, the dS between articles Alihamsjah et al. (1990) and Chung et al. (2001) is 1, as Alihamsjah et al. (1990) was cited by Chung et al. (2001). This is shown as row H, column M of Table 1 below. Another example is for Adamchuk & Christenson (2007) and Stafford and Hendrick (1988) is 2. In Adamchuk & Christenson (2007), no reference is given to Stafford and Hendrick (1988), but it gives reference to Adamchuk et al. (2001). In Adamchuk et al. (2001) there is a reference to Stafford and Hendrick (1988). Therefore there are 2 degrees of separation between the two articles (shown as row E, column O of Table 1 below).

The degrees of separation of the main papers on this topic are shown in a matrix in Table 1. The

**Table 1 Degrees of separation between papers on on-the-go soil strength measurement**

	D	F	G	H	I	J	K	L	M	N	O
A	∞*	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
B	1	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
C	1	∞	∞	∞	∞	∞	∞	∞	∞	1	∞
D	0	∞	∞	∞	∞	∞	∞	∞	∞	∞	∞
E	∞	1	1	∞	∞	2	3	2	3	∞	2
F		0	∞	∞	∞	1	2	2	2	∞	2
G			0	∞	∞	∞	∞	1	2	∞	1
H				0	∞	1	1	2	1	∞	2
I					0	∞	∞	2	1	∞	2
J						0	1	1	1	∞	1
K							0	∞	1	∞	2
L								0	∞	∞	2
M									0	∞	1
N										0	∞
O											0

\*∞ as we possibly have not found all intervening papers this might be better expressed as >5

A Keen & Haines (1925)

B Haines & Keen (1925a)

C Haines & Keen (1925b)

D Haines & Keen (1928)

E Stafford & Hendrick (1985/88)

F Owen et al. (1987)

G Glancey et al. (1989)

I Van Bergeijk et al. (1996/2001)

J Adamchuk et al. (2001)

K Hall & Raper (2005)

L Mouazen & Ramon (2006)

M Chung et al. (2006)

N Watts et al. (2006)

O Adamchuk & Christenson (2007)

rows and columns are in chronological order and the lower half of the matrix is blank because we assume earlier papers cannot refer to later ones. There appears to be a literature developing but there is also a relatively high degree of non citation - many of the cells show infinite separation\*. The most remarkable finding is that none of the early work of Haines and Keen appears to have been referenced in the scientific literature (see rows A, B, C, D) until the Broadbalk measurements were repeated recently (Watts et al., 2006), and Watt et al.'s work seems to be separated in the citation sense from the main body of contemporary soil strength sensing work (see column N)

Haines and Keen's work was lost or seen as irrelevant until the new impetus given by precision agriculture in the 1990's when similar, but independent, work evolved again.

The 'degree of separation' analysis can be used to identify and help towards the development of ideas in a discipline.

## References

- Adamchuk, V.I., Morgan, M.T., Sumali, H., 2001. Application of a strain gauge array to estimate soil mechanical impedance on-the-go. *Transactions of the ASAE* 44, 1377-1383.
- Adamchuk VI, Christenson PT., Development of an instrumented blade system for mapping soil mechanical resistance represented as a second-order polynomial. *Soil and Tillage Research* 95, 76-83.
- Alihamsyah, T., Humphries, E.G., Bowers Jr., C.G., 1990. A technique for horizontal measurement of soil mechanical impedance. *Transactions of the ASAE* 33, 73-77.
- Chung, S.O., Sudduth, K.A., Hummel, J.W. 2006 Design and validation of an on-the-go soil strength profile sensor *Transactions of the ASABE* 49, 5-14
- Glancey, J.L., Upadahyaya, S.K., Chancellor, W.J., Rumsey, J.W., 1989. An instrumented chisel for the study of soil-tillage dynamics. *Soil and Tillage Research* 14, 1-24.
- Haines, W.B. and Keen, B.A. 1925a. Studies in soil cultivation. II. Test of soil uniformity by means of dynamometer and plough. *Journal of Agricultural Science (Cambridge)*, 15: 387-394.
- Haines, W.B. and Keen, B.A. (1925b) Studies in soil cultivation. III. Measurements on the Rothamsted classical plots by means of dynamometer and plough. *Journal of Agricultural Science (Cambridge)*, 15: 395-406.
- Haines, W.B. and Keen, B.A. (1928) Studies in soil cultivation. IV. A new form of traction dynamometer. *Journal of Agricultural Science (Cambridge)*, 18: 724-732.
- Hall, H.E., Raper, R.L., 2005. Development and concept evaluation of an on-the-go soil strength measurement system. *Transactions of the ASAE* 48, 469-477.
- Keen, B.A. and Haines, W.B. (1925) Studies in soil cultivation. I. The evolution of a reliable dynamometer technique for use in soil cultivation experiments. *Journal of Agricultural Science (Cambridge)*, 15: 375-386.
- Mouazen AM, Ramon H., 2006. Development of on-line measurement system of bulk density based on on-line measured draught, depth and soil moisture content. *Soil and Tillage Research* 86, 218-229.
- Owen, G.T., Drummond, H., Cobb, L. and Godwin, R.J., 1987. An instrumentation system for deep tillage research. *Transactions of the ASAE* 30, 1578-1582.
- Stafford, J.V., Hendrick, J.G. 1985. Dynamic control of pan rupturing tines. *ASAE Paper No. 85-1547*.
- Stafford, J.V., Hendrick, J.G., 1988. Dynamic sensing of soil pans. *Transactions of the ASAE* 31, 9-13.
- Van Bergeijk, J., Goense, D., 1996. Soil tillage resistance as tool to map soil type differences. In: Robert, P.C., Rust, R.H., Larson, W.E. (Eds.), *Proceedings of the Third International Conference on Precision Agriculture, ASA-CSSA-SSSA, Madison, Wisconsin*, pp. 605-616.
- Van Bergeijk, J., Goense, D., Speelman, L., 2001. Soil tillage resistance as a tool to map soil type differences. *Journal of Agricultural Engineering Research* 79, 371-387.
- Watts, C.W., Clark, L.J., Poulton, P.R., Powlson, D.S. and Whitmore, A.P., 2006. The role of clay, organic carbon and long-term management on mouldboard plough draught measured on the Broadbalk wheat experiment at Rothamsted. *Soil Use & Management*, 22: 334-341.

### Soil Science on D-Day

Murray has written an article on "Science on the Normandy Beaches: J.D. Bernal and the Prediction of Soil Trafficability for Operation Overlord" for *Soil Survey Horizons* 49:12-15 (Spring, 2008). This is following his "From the Chair" editorial in Issue No. 21 of *Pedometron*. You can access the article for a limited time at:

<https://www.soils.org/press/releases/2008/0414/002/>



## Introduction

Scientists are known for their penchant to come up with new words, or to use old words in a new sense. The word 'pedometrics' is a well known example of such a neologism. It was coined by McBratney (1986) as the soil science counterpart of research areas like econometrics, biometrics, psychometrics, and geometrics.

A scientist who definitely earned a place in the neologistic hall of fame is John Wilder Tukey. He has enriched the language with numerous new words and gave new meanings to existing ones. The word 'bit' for instance, a contraction of 'binary digit', was proposed by him in 1946. And 'software', as opposed to 'hardware', was first mentioned in printed form in Tukey (1958). He also introduced for pedometricians familiar terms like 'box-and-whisker plot', 'jackknifing', 'prewhitening', and 'hat matrix'. A less well-known neologism contributed by him is 'headbanging'.

It was coined by Tukey and Tukey (1981, p.247) to refer to a specific kind of nonlinear smoother over the plane. Although occasionally mentioned by pedometricians (McBratney, 2004; Goovaerts, 2005), applications of headbanging seem to be mainly restricted to smoothing mortality rates (Kafadar, 1996; Kafadar, 1999), health risk factors (Williams Pickle and Su, 2002) and temperature trends (Kafadar, 1994; Mungiole *et al.*, 1999; Mungiole and Pickle, 1999; Kafadar and Morris, 2002; Lu *et al.*, 2005). However, headbanging might also be of interest to the Pedometrics community. In particular, to reveal hidden spatial structures in noisy soil data.

## Smoothing

Before going into details about the headbanging algorithm itself, let's start with a brief introduction to smoothing. An effective smoother should retain genuine structures in the data and remove small scale spatial variation or noise. Stated another way, a contiguous area of high values (like a ridge) is more likely to represent a real phenomenon and has to be preserved whereas a high isolated value surrounded by relatively low values is more likely to be an outlier and has to be removed. In this way, smoothing may potentially shed some light on the process(es) that generated the data.

In mathematical terms, smoothing boils down to the decomposition of spatial data  $y_m(s)$  at location  $s$  into a component representing the signal  $y(s)$  and a component representing the noise  $\varepsilon$ :

$$y_m(s) = y(s) + \varepsilon$$

The signal  $y(s)$  will be estimated by the 'smooth'  $\tilde{y}(s)$ .

## Headbanging algorithm

Headbanging is a two dimensional extension of Tukey's (1977) highly robust running median smoother, also known as 3R, *i.e.*, running medians of length three repeated until convergence. The headbanging algorithm consists of the following steps:

1. Construct so-called triples, *i.e.* ordered sets of three points where the central point in the order is roughly at the midpoint of the other two (Tukey and Tukey, 1981). The selection of triples will be illustrated by means of Figure 1 in the next paragraph.
2. Compute the 'low screen' and 'high screen' for each point to be smoothed. For each triple, the end point with the lower value will be assigned to the set  $L$  and the end point with the higher value to the set  $H$ . The low screen is given by the median of  $L$ , and the high screen by the me-

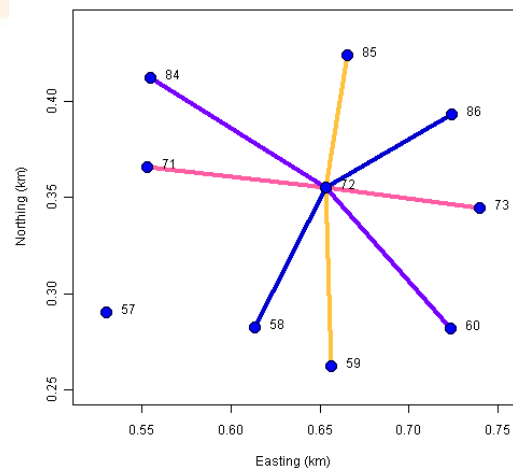


Figure 1: Four sets of triples corresponding to centre point 72. Triples are connected by coloured line segments.

dian of  $H$ .

3. Compute the new (smoothed) value at each point as the median of its low screen, its high screen, and its current value.
4. Return to step 2 and repeat the procedure until some convergence criterion is met.

Tukey and Tukey (1981) referred to this algorithm as ‘headbanging’ because it bears some resemblance to repairing a wall in which the heads of the nails that have popped out needs to be banged down.

The selection of triples (see step 1) will be illustrated by means of Figure 1. Suppose triples have to be found for point 72. Then the first step is to find its nearest neighbours, i.e., all points within a specific distance from point 72. In Figure 1, all points except for point 57 are neighbours of point 72. Point 57 is too far out to be a neighbour. Next, all (nearly) collinear triples will be selected, i.e. triples where the angle formed by the end points and centre point is approximately  $\rho$  radians. In Figure 1, triple (71, 72, 73) has an angle of nearly  $\rho$  radians and will be retained, whereas triple (58, 72, 59) will be rejected because its angle is too small. If the total number of triples exceeds an a priori maximum, then only the ‘thinnest’ triples will be retained, i.e., those triples for which the distance from the centre point to the line segment connecting the end points is smallest. This criterion tends to favour collinear pairs as well as pairs close to the point to be smoothed (Kafadar and Morris, 2002). In Figure 1, for example, triple (71, 72, 73) is thinner than triple (58, 72, 86).

To reduce bias, Tukey and Tukey (1981) recommend to twice the headbanging procedure. This means that the

final smooth is the sum of the ‘smooth’ and the ‘smooth of the rough’. In other words, headbanging is first applied to the raw data, and then to the residuals (or the ‘rough’). Both smooths are summed together to obtain the final smooth. ‘Twicing’ is also common practice for one dimensional nonlinear smoothers like Tukey’s 3R. In addition, it is also another neologism coined by Tukey.

## Headbanging in action

The performance of headbanging will be illustrated by means of a limited simulation study.

### Simulation of spatial data sets

Spatial data sets have been simulated that comply with the following model:

$$y(s) = m(s) + r(s)$$

where the ‘real’ value  $y(s)$  consists of a trend component  $m(s)$  and a spatially dependent noise component  $r(s)$ . In this simulation study,  $m(s)$  is either zero (representing a plane) or a step function (representing a cliff). Component  $r(s)$  is a Gaussian random field with zero expectation, unit variance, and an exponential autocorrelation structure.

Surface  $y(s)$  has been sampled by means of stratified simple random sampling. A total of  $n = 250$  locations  $s_i, i \in \{1 \dots n\}$  have been selected. At these locations, measurements  $y_m(s_i)$  have been simulated by adding spatially independent Gaussian noise  $\varepsilon$  (i.e., measurement error) to  $y(s_i)$ :

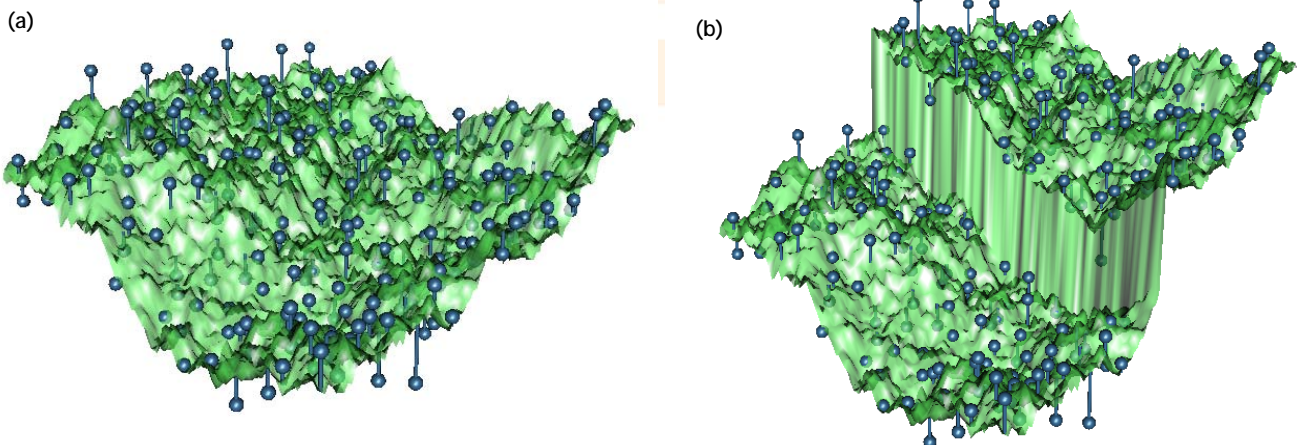


Figure 2: Realizations of  $y_m(s)$  (steel blue spheres, or ‘nails’) for the plane (a) and the cliff (b). The semi-transparent green surface represents the real world  $y(s)$ .

$$y_m(s_i) = y(s_i) + \varepsilon$$

The green surfaces in Figure 2 represent realizations of  $y(s)$  where  $m(s)$  is either a plane (Figure 2a) or a step function (Figure 2b). The steel blue spheres (the ‘nails’) represent observations  $y_m(s_i)$ . These are just two realizations of  $y_m(s_i)$ . To estimate the variation in smoothing efficiency, 100 equally probable samples of size 250 have been selected by stratified simple random sampling. In addition, 100 realizations of measurement noise  $\varepsilon$  have been added to each sample. This has been done for three noise levels, *i.e.* 10%, 20%, and 50% of the variance of  $r(s)$ . In other words,  $100 \times 100 \times 3 = 30\,000$  realizations of  $y_m(s_i)$  have been simulated for both the plane and the cliff structure.

## Headbanging versus LOESS

Observations  $y_m(s_i)$  will be smoothed by headbanging and the more familiar locally weighted polynomial regression smoother (Cleveland, 1979; Cleveland and Devlin, 1988; Cleveland *et al.*, 1988). The latter is also known as LOESS and will be used as a reference. The mean squared error, defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n [y(s_i) - \tilde{y}(s_i)]^2$$

where  $\tilde{y}(s_i)$  is the smoothed value at location  $s_i$ , will be used as a measure of smoothing efficiency (lower MSEs correspond to better smoothers).

We used the LOESS implementation in R (R Development Core Team, 2008). The degree of smoothing is controlled by parameter ‘span’. Apart from the default value for span, we also applied LOESS with optimized span settings, *i.e.*, the span that minimized the MSE. In this way, LOESS has been given a head start. Note that this optimization can’t be done in practice because  $y(s)$  is not known. Therefore, LOESS with optimized span should be interpreted as a ‘best case’ situation for the LOESS smoother.

## Results

Figure 3 gives bubblecharts of absolute errors

$|y(s_i) - \tilde{y}(s_i)|$  for LOESS with default span (L), for LOESS with optimized span ( $L^*$ ) and for headbanging (H). Bubble charts for the unsmoothed data are also given as a reference (R). The top row gives the results for the plane and the bottom row for the cliff. The cliff face runs North-South at Easting = 0.5.

For the plane (top row), headbanging and optimized LOESS give similar results. They both outperform LOESS with default settings. For the cliff (bottom row), both LOESS smoothers tend to oversmooth the data near the cliff face. Headbanging on the other

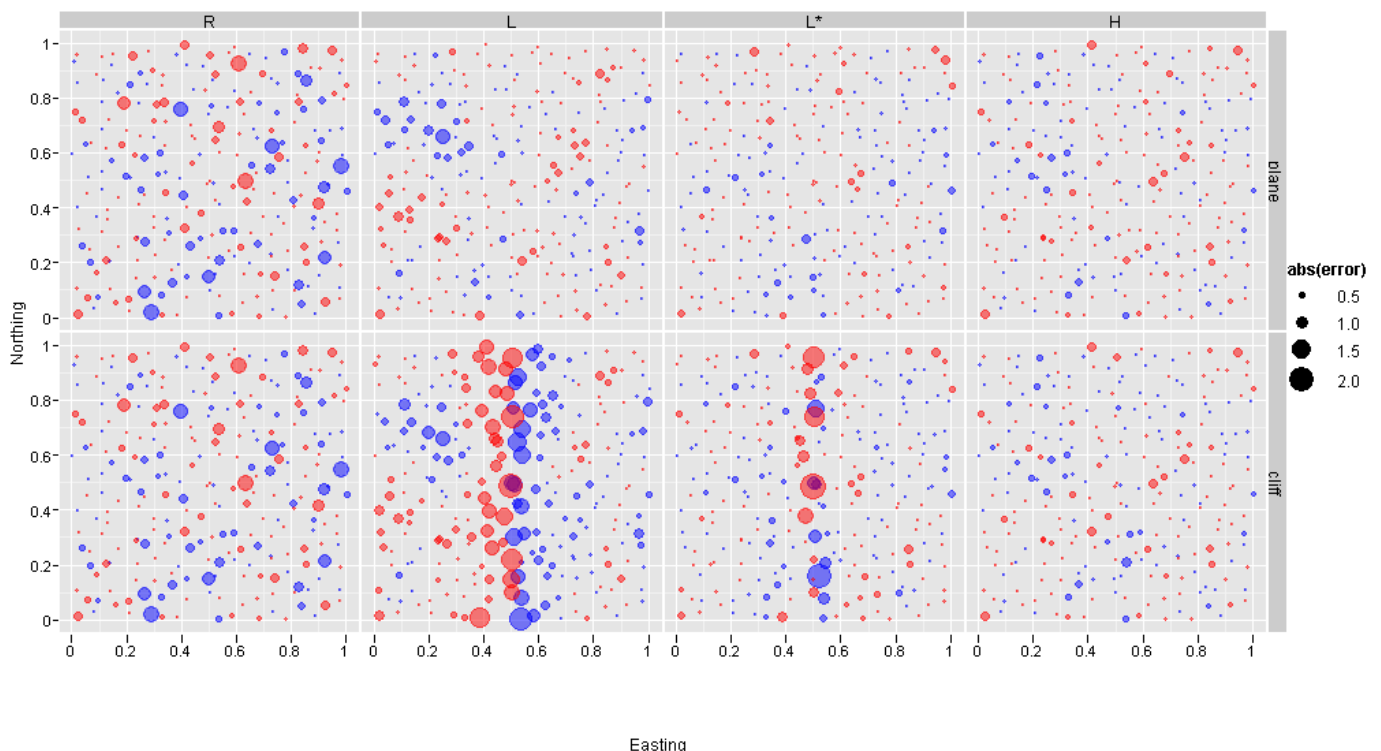


Figure 3: Bubble charts of absolute errors for one realization of the plane (top row) and the cliff (bottom row). The columns represent successively, the unsmoothed situation (R), LOESS with default span (L), LOESS with optimized span ( $L^*$ ) and headbanging (H). The cliff face runs North-South at Easting = 0.5 (bottom row only). The size of each bubble gives the absolute error. Blue bubbles denote underestimation, red bubbles overestimation.

hand far better preserves this structure. Although not shown in Figure 3, the differences are even more pronounced when isolated outliers are present. LOESS greatly affects the neighbourhood of these outliers by introducing local bias. Headbanging on the other hand, primarily bangs down the outliers, leaving the neighbourhood mostly unaffected.

Figure 3 pertains only to one realization of  $y_m(s_i)$  for the plane and one for the cliff. It is far more interesting to compare smoothing efficiencies for an ensemble of realizations. In Figure 4, the MSEs of the 30000 realizations of  $y_m(s_i)$  for the plane and the cliff are represented by box-whisker plots (another invention of Tukey). For the cliff, headbanging outperforms both LOESS smoothers. It better preserves the underlying spatial structure. For lower noise levels, LOESS has even higher MSEs than the unsmoothed data. For the plane, however, the optimized LOESS smoother is clearly the winner, and headbanging comes in second place for the lower noise levels. However, it should be recalled that the optimized LOESS is a best case situation and that this optimization can't be done in practice.

Kafadar and Morris (2002) came to similar conclusions. In addition, they found that headbanging is more effective on smaller sample sizes than LOESS, in particular when the underlying spatial pattern has some

structure.

## Different ways of headbanging

The headbanging algorithm described in this paper closely resembles that of Tukey and Tukey (1981). The main difference lies in the objective function to be minimized in order to find appropriate triples. However, other variants of the headbanging algorithm exist.

Mungiole et al. (1999) introduced a weighted headbanging algorithm, in which weighted medians were taken in stead of ordinary medians. They used the inverse of the standard error of mortality rates as weights. Other weighting schemes are also possible. Another aspect in which the algorithms differ is the way edge effects are taken into account. In this paper, we followed the procedure proposed by Tukey and Tukey (1981).

## Headbanging: give it a go yourself

Unlike its one dimensional counterparts, headbanging is rarely applied during the preliminary stages of exploratory data analysis. One reasons might be the limited availability of headbanging software.

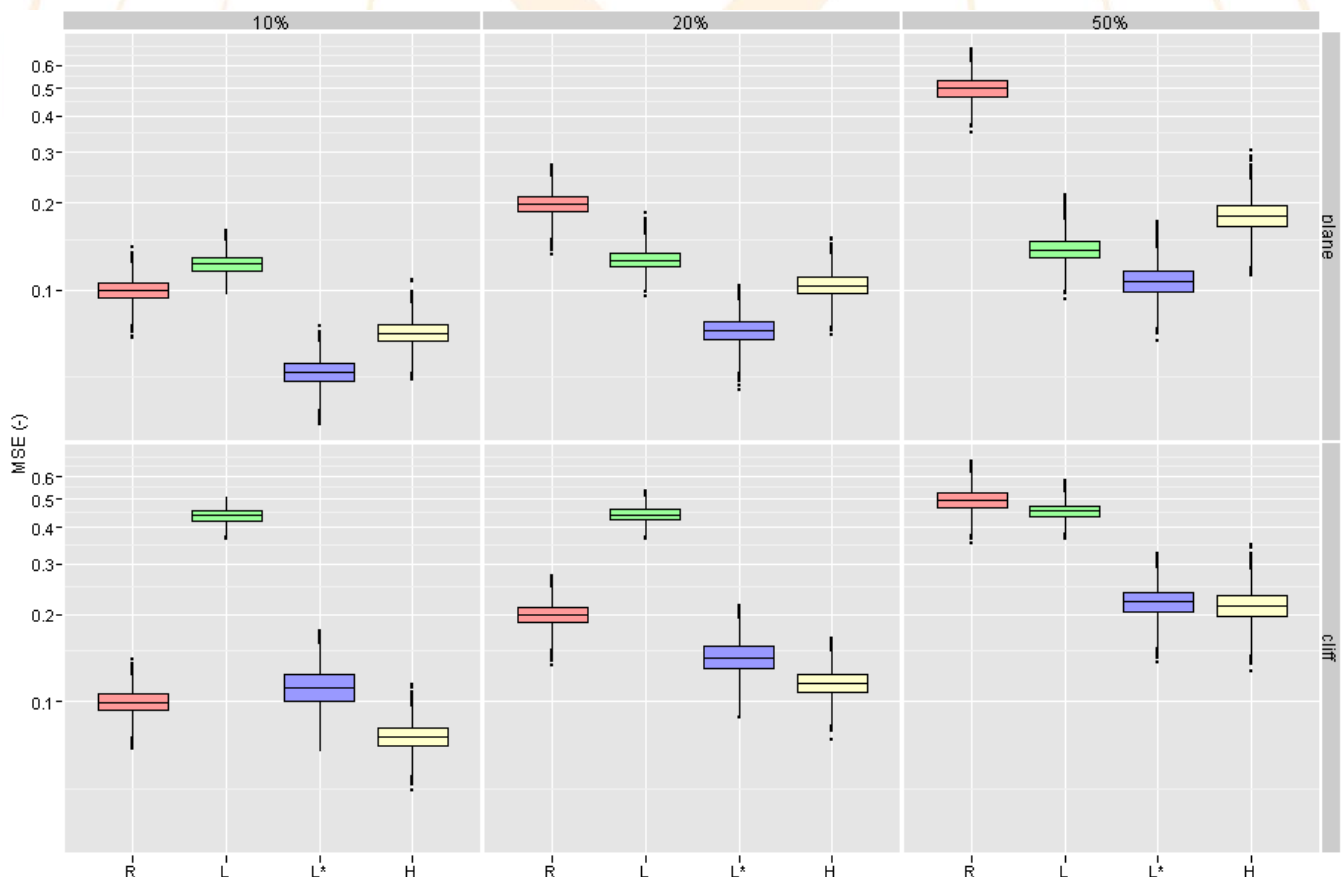


Figure 4: Box-whisker plots of the MSEs for LOESS with default span (L), for LOESS with optimized span ( $L^*$ ) and for headbanging (H). Box-whisker plots for the unsmoothed data are given as a reference (R). The top row gives the results for the plane and the bottom row for the cliff. The columns correspond to the three levels of measurement error, i.e., 10%, 20%, and 50% of the variance of  $r(s)$ .

In order to write this paper, I've implemented a version of the headbanging algorithm in R (R Development Core Team, 2008). If you're eager to give it a go yourself, just send me an email ([dennis.walvoort@wur.nl](mailto:dennis.walvoort@wur.nl)), and I'll send you the R-package.

## References

Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* 74(368): 829-836.

Cleveland, W. and Devlin, S. (1988). Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association* 83(403): 596-610.

Cleveland, W., Devlin, S. and E., G. (1988). Regression by local fitting. methods, properties, and computational algorithms, *Journal of Econometrics* 37: 87-114.

Goovaerts, P. (2005). Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using poisson kriging, *International Journal of Health Geographics* 4(31).

Kafadar, K. (1994). Choosing among two-dimensional smoothers in practice., *Computational Statistics & Data Analysis* 18: 419-439.

Kafadar, K. (1996). Smoothing geographical data, particularly rates of disease, *Statistics In Medicine* 15: 2539-2560.

Kafadar, K. (1999). Simultaneous smoothing and adjusting mortality rates in U.S. counties: Melanoma in white females and white males, *Statistics in Medicine* 18: 3167-3188.

Kafadar, K. and Morris, M. (2002). Nonlinear smoothers in two dimensions for environmental data, *Chemometrics and Intelligent Laboratory Systems* 60: 113-125.

Lu, Q., Lund, R. and Seymour, L. (2005). An update of U.S. temperature trends, *Journal of Climate* 18: 4906-4914.

M<sup>c</sup>Bratney, A. (1986). Introduction to Pedometrics: a course of lectures, CSIRO Australia, Division of Soils Technical Memorandum 53/1986, Private Bag 2, Glen Osmond, S.A. 5064, Australia.

M<sup>c</sup>Bratney, A. (2004). Book review: Everitt, B.S., 2002. *The Cambridge Dictionary of Statistics.*, *Geoderma* 121(1{2}): 155-156.

Mungiole, M. and Pickle, L. (1999). Determining the optimal degree of smoothing using the weighted head-banging algorithm on mapped mortality data, *ASC '99- Leading Survey & Statistical Computing into the New Millennium, Proceedings of the ASC International Conference, September 1999.*

Mungiole, M., Pickle, L. and Hansen Simonson, K. (1999). Application of a weighted headbanging algorithm to mortality data maps, *Statistics in Medicine* 18: 3201-3209.

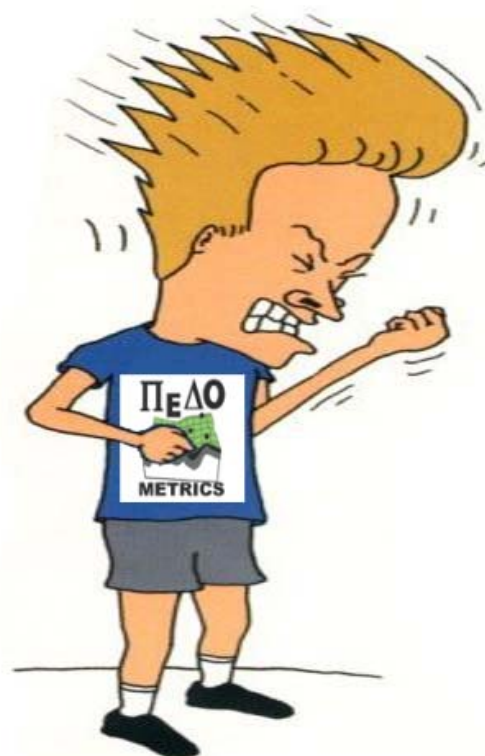
R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>

Tukey, J. (1958). The teaching of concrete mathematics, *The American Mathematical Monthly* 65: 1-9.

Tukey, J. (1977). *Exploratory Data Analysis*, Addison-Wesley, Reading Massachusetts.

Tukey, P. and Tukey, J. (1981). Summarization; smoothing; supplemented views, in V. Barnett (ed.), *Interpreting Multivariate Data*, John Wiley & Sons, Chichester, pp. 245-275.

Williams Pickle, L. and Su, Y. (2002). Within-state geographic patterns of health insurance coverage and health risk factors in the United States, *American journal of preventive medicine* 22(2): 75-83.





I am convinced as a pedometrician that the state of the art in theoretical and applied soil science, e.g., in soil monitoring, requires more specific mathematical description of soil properties and their relationships with soil-forming factors, as well as the reliable evaluation and prediction of changes induced by current anthropogenic and natural processes. Many advances have been made in the development of statistical, informational, and other mathematical methods for the quantitative description of analytical results and data processing. I believe that the further development of soil science calls for the expansion of probabilistic thought and the mathematical specification of basic theoretical concepts of pedology, including the concept of "soil properties." This specification may include their statistical probability distributions, as the simplest mathematical models of soil properties accounting for their variability.

It is known that soil properties exhibit quantitative diversity that is related to soil-forming factors by "soft" statistical rather than "hard" functional relationships. Multiple factors and the cumulative synergetic action of processes with different degrees of organization determine not only the average levels but also the changes in property values; therefore, variability should be considered as an intrinsic systemic property of the soil. On the other hand, variability determines the quantitative measure of properties and, hence, affects the practical and economic value of the soil. Therefore, the statistical probability distributions of properties, which represent a mathematical model of variability, have a scientific value for the systemic characterization of soil and an applied value for the specification of the soil quality and status.

You may ask me what I mean by "statistical probability" distribution? Everybody knows statistical or probability distribution, but statistical and probability, hmm? Indeed theoretically, statistical and probability distributions are different concepts. We mean statistical distributions as histograms from observation data, and probability distributions as mathematical functions. The latter are theoretical ones, because they are theoretically introduced from some basic physical con-

cepts. However there is no theory, which allows us to know this function with confidence. We have only very indistinct preposition about normality and sometimes log-normality of distribution.

There are procedures of hypothesis testing to support inferences about statistical distributions. Numerous authors showed that empirical statistical distributions of soil properties frequently differ from the normal (Gaussian) or log-normal distribution. However, any available information on their mathematical functions is still scanty and sometimes contradictory. This is explained by methodological problems and the ambiguity of hypothesis testing. In spite of this we propose the model of changes in the statistical distribution of features upon the nature and anthropogenic processes indicates that the statistical distributions can be of various shapes (Mikheeva, 1997; 2001; 2005a, b). Some ideas of this were in Hans Jenny's famous book (Jenny, 1948).

So we suggested that there are specific functions, which describe specific probability distributions of soil properties that were a result of evolution of soils as complex system. Normality is only one case among plenty of cases. However we realized that these functions could only be obtained by statistical analysis of actual soil data. Moreover we need large reliable data sets analysed by standard methods. Only such functions we proposed be statistical probability distributions of soil properties. New developments have been made in this field of mathematical statistics, the application of which allowed us to analyze and identify the mathematical functions of the statistical probability distributions (Lemeshko, 1995).

We analyzed data of large-scale soil surveys, which were done at an area part of the Kulunda steppe on the south of Western Siberia (approximately 1.6 million ha) in 1965, 1975, 1982 and 1989 by standard methods (*All-Union Guidelines*, 1973). The region is characterized by a dry continental climate, and its relief is a gently undulating plain. The soil cover consists of chestnut soils (70%), meadow-chestnut soils, meadow soils, solonetz, and solonchaks with different degrees of hydromorphism (in Russian classification).

The chestnut soils significantly vary in texture, from loose sands to loams, which is a result of the ancient Quaternary alluvial genesis of the territory. The texture of a soil largely determines its physical properties and water regime, which in turn determine the chemical properties. The content of humus in the soils studied is 1 - 3% depending on the soil type. Loamy sandy and sandy loamy soils are predominant; they have an evenly colored humus-accumulative horizon and are uniform in density and display deep effervescence (Mikheeva and Kuz'mina, 2000).

At the first stage we grouped these data according to soil-genetic principles. Every group refers to a soil type and a soil class, namely, this data belongs to a soil type and a texture class. The number of these samples was  $n = 50 - 650$ , depending on the abundance of the soil groups. The exactness of this grouping was shown by the unimodality of the distributions and high level of significance.

"The identification" of a probability distribution implied the selection of the best one among the 27 known functions to describe the empirical distribution using a special program. For this purpose, the parameters of each of the 27 functions were estimated using the method of maximum likelihood from the actual soil data and the

hypothesis about the agreement between the empirical data and the tested distribution was checked.

Most researchers select the best distribution using a single fitting criterion for a specified significance level (usually 0.01, 0.05, or 0.1) and accept the hypothesis of agreement when the corresponding statistic does not exceed the critical value. It is not taken into account that the fitting criteria differ in terms of the agreement measure and that the agreement with a criterion does not suggest an agreement with other criteria. Our experience shows that several criteria with different measures of agreement should be used, and the decision should be based on their integrity. Our program system used six criteria for testing the hypotheses; the Pearson  $\chi^2$  criterion (two modifications), the likelihood ratio test (two modifications), the Kolmogorov's criterion, the Smirnov criterion, and two von Mises criteria ( $\omega^2$  and  $\Omega^2$ ).

The conventional procedure of hypothesis testing involves the non-rejection of the null hypothesis. However, frequently there is no reason to reject several hypotheses; i.e., a number of alternatives remain. At the same time, the distribution with the best fit should

be selected. Therefore, the probability is calculated for each  $i$ -th criterion and each  $j$ -th theoretical distribution (where  $i$  is the index of one of the  $K$  criteria used, and  $j$  is the index of one of the  $R$  tested theoretical distributions):

$$P\{S_i > S_{ij}^*\} = \int_{S_{ij}^*}^{\infty} g_i(s) ds = \alpha_{ij},$$

(1)

where  $S^*$  is the corresponding statistics of the criterion used for the  $i$ -th distribution, and  $g_i(s)$  the known distribution density function of statistic  $S_i$  provided that the  $H_0$  hypothesis is true.

Then, in testing the hypothesis about the agree-

Table 2. Some statistical probability density functions for soil properties

Function name	Density function
Su-Johnson	$f(x) = \frac{\theta_1}{\sqrt{2\pi} \sqrt{(x-\theta_3)^2 + \theta_2^2}} \exp\left\{-\frac{1}{2}\left[\theta_0 + \theta_1 \ln\left[\frac{x-\theta_3}{\theta_2} + \sqrt{\left(\frac{x-\theta_3}{\theta_2}\right)^2 + 1}\right]\right]^2\right\}$
Exponential functions	$f(x) = \frac{\theta_2}{2\theta_1 \Gamma(1/\theta_2)} \exp\left\{-\left(\frac{ x-\theta_0 }{\theta_1}\right)^{\theta_2}\right\}$
Double power	$f(x) = \theta_0 \theta_1 \exp\{-x\theta_0 - \theta_1 \exp(-x\theta_0)\}$
Maximum value	$f(x) = \frac{1}{\theta_1} \exp\left\{-\frac{x-\theta_0}{\theta_1} - \exp\left(-\frac{x-\theta_0}{\theta_1}\right)\right\}, x > \theta_0$
Minimum value	$f(x) = \frac{1}{\theta_1} \exp\left\{\frac{x-\theta_0}{\theta_1} - \exp\left(\frac{x-\theta_0}{\theta_1}\right)\right\}, x < \theta_0$
Normal	$f(x) = \frac{1}{\theta_1 \sqrt{2\pi}} e^{-\frac{(\ln x - \theta_0)^2}{2\theta_1^2}}$
Log-normal	$f(x) = -\frac{1}{x \theta_1 \lg 10 \sqrt{2\pi}} e^{-(\lg x - \theta_0)^2 / 2\theta_1^2}$
Weibull	$f(x) = \frac{\theta_0 (x - \theta_2)^{\theta_0 - 1}}{\theta_1^{\theta_0}} \exp\left\{-\left(\frac{x - \theta_2}{\theta_1}\right)^{\theta_0}\right\}$
Cauchy	$f(x) = \frac{\theta_0}{\pi [\theta_0^2 + (x - \theta_1)^2]}$
Nakagami	$f(x) = \frac{2}{\Gamma(\theta_1)} \left(\frac{\theta_1}{\theta_0^2}\right)^{\theta_1} (x - \theta_2)^{2\theta_1 - 1} \exp\left\{-\frac{\theta_1 (x - \theta_2)^2}{\theta_0^2}\right\}$

ment with the  $j$ -th distribution by the  $i$ -th criterion, if  $\alpha_{ij} > \alpha$  (where  $\alpha$  is the preset significance level), there is no reason to reject the hypothesis of agreement with the  $j$ -th distribution according to the  $i$ -th criterion. Let there be no reason to reject the hypothesis about the agreement with many laws noted by indices  $R1$  (out of  $R$  indices). Then, the distribution law is selected for which

An unambiguous conclusion may usually be drawn. However, if an uncertainty remains (e.g., in the case of similar distribution laws), the multicriterion problem of decision-making is solved. In this case, a simple compromise criterion is composed in the form

$$\max_{j \in R1} \sum_i \omega_i \alpha_{ij},$$

where  $\omega_i$  is the weighting coefficient of the  $i$ th criterion,  $\sum_i \omega_i = 1$ .

For example, the statistical probability distribution of humus in the fine-loamy sandy chestnut soil in 1975 was identified by the program as a Nakagami distribution with the parameters  $\theta_0 = 0.63$ ,  $\theta_1 = 0.89$ , and  $\theta_2 = 0.98$ . The program calculates the criterion statistics and probabilities  $\alpha_{ij}$ , using Eq.(1) (Table 2). It is notable that the calculated probabilities significantly exceed the standard significance levels. This implies that, among the 27 distributions examined, the Nakagami distribution is characterized by the highest probability of agreement between the empirical data and the theoretical distribution according to the totality of the criteria used. Similar calculations were conducted for more than 100 samples.

The seeming awkward equation poses no problem, because a computer performs all the calculations. Thus, this paper includes a bank of statistical probability distributions for humus content in chestnut soils derived from the analysis of large statistical samples from vast territories (Fig.1) (for more information look through Mikheeva, 2005 b).

In Figure 1 are shown statistical probability distributions of the contents of humus in a lithosequence of chestnut soils: coherent sands, coarse-loamy sands, fine-loamy sands, sandy loams and loams in horizon A (a) and horizon B1 (b).

Table 2. Statistical criteria and probabilities  $\alpha_{ij}$

Criterion	Statistics	Probability
Likelihood relations (2)	2.5159	0.28
Likelihood relations (4)	2.5159	0.64
Chi-square (2)	2.4962	0.29
Chi-square (4)	2.4962	0.65
Kolmogorov	1.1425	0.62
Smirnov	0.4883	0.78
w-square	0.0467	0.90
W-square	0.3908	0.86

These functions are distinguished accordingly to the texture by shift along abscissa, diffusion (dispersion) and form. They differ among the genetic horizons too. The distributions are characterized by a narrow central part and skewness (double power and Johnson's Su distributions) except for the fine-loamy sands variety (double exponential distribution), which has no skewness. In sands, loamy sands, and sandy loams, the distributions of the humus overlap significantly, especially in the B1 horizon where the differences in the humus content are determined by a small shift and an increase in the dispersion for the same distribution type. In loamy soil, the humus distribution is more shifted with respect to other soils, and it has a very broad central part for the B1 horizon (Fig. 1b).

The skewness and kurtosis of the humus distributions in the soils of the steppe zone are explained by the low content of humus in these soils, which is close to its lower threshold. When the texture becomes finer and the degree of hydromorphism increases, the humus content increases. This results not only in a shift of the distribution but also in an increase in the variance and in changes in the skewness and kurtosis; i.e., a wave of the statistical probability distribution is observed.

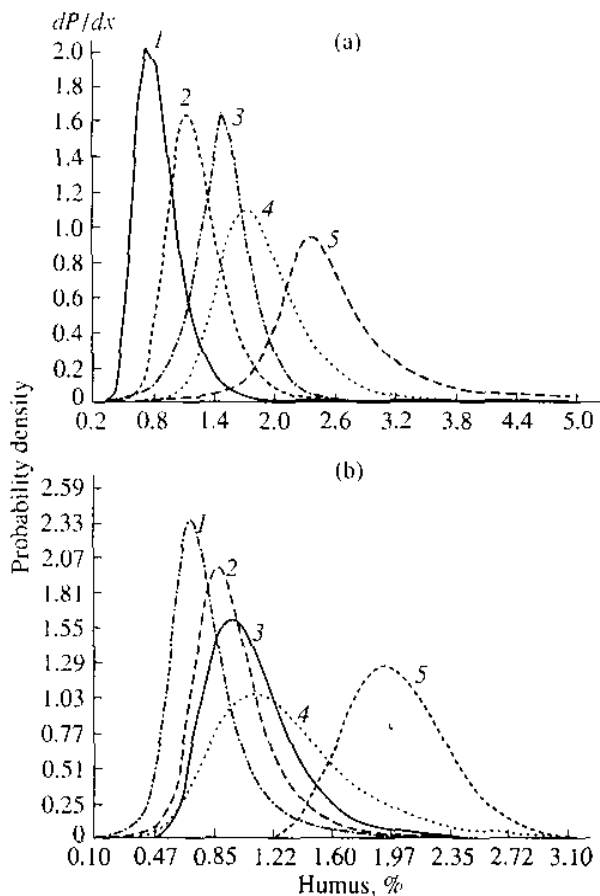


Fig. 1. Statistical probability distributions of humus in horizons (a) Ap and (b) B1 of chestnut soils.

(a):

- (1) Coherent sands, maximum value distribution,  $\theta_0 = 0.18, \theta_1 = 0.77$ ;
- (2) Coarse-loamy sands, Johnson's Su distribution  $\theta_0 = -1.34, \theta_1 = 1.98, \theta_2 = 0.4, \theta_3 = 0.91$ ;
- (3) Fine-loamy sands, double exponential distribution,  $\theta_0 = 1.5, \theta_1 = 0.32, \theta_2 = 1.32$ ;
- (4) Sandy loams, Johnson's Su distribution,  $\theta_0 = -1.45, \theta_1 = 1.91, \theta_2 = 0.56, \theta_3 = 1.36$ ;
- (5) Loams, Johnson's Su distribution,  $\theta_0 = -0.61, \theta_1 = 0.99, \theta_2 = 0.38, \theta_3 = 2.25$ .

(b):

- (1) Coherent sands, Johnson's Su distribution,  $\theta_0 = -0.87, \theta_1 = 1.28, \theta_2 = 0.19, \theta_3 = 0.63$ ;
- (2) Coarse-loamy sands, Johnson's Su distribution,  $\theta_0 = -0.99, \theta_1 = 1.46, \theta_2 = 0.25, \theta_3 = 0.76$ ;
- (3) Fine-loamy sands, maximum value distribution,  $\theta_0 = 0.19, \theta_1 = 0.94$ ;
- (4) Sandy loams, maximum value distribution,  $\theta_0 = 1.11, \theta_1 = 0.35$ ;
- (5) Loams, Nakagami distribution,  $\theta_0 = 0.93, \theta_1 = 0.41, \theta_2 = 1.14$ .

In this way extensive analysis of the statistical probability distributions of soil properties, expressed as mathematical functions best describing the occurrence probabilities of their values, showed that the distributions regularly and adequately describe the changes in the soil proper-

ties even under similar conditions and represent mathematical models of variability. This fact reflects the continuous nature of the genetic dependence of soil properties on the soil-forming factors and processes. The distribution functions of soil properties change their shape under the long-term effect of natural and anthropogenic processes, which allow them to be used as a scientific tool for studying the tendencies in the soil cover changes. This result has also a practical value, because the assessment of the hazard and responsibility of natural systems to changes is always based on the statistical distributions of their properties.

Please, forgive me for my not very good English

## REFERENCES

- I. V. Mikheeva, "Changes in the Spatial Variability of Soil Properties under Anthropogenic Impact," *Pochvovedenie*. No. 1, 102-109 (1997) [*Eur. Soil Sci.* 30 (1), 87-93 (1997)].
- I.V. Mikheeva, *Statistical Probability Models of Soil Properties (with Chestnut Soils of the Kulunda Steppe as an Example)* (Ross. Akad. Nauk, Novosibirsk, 2001) [in Russian].
- I.V. Mikheeva, *Monitoring and Statistical Probability Assessment of Stability and Changeability of Nature Objects under Modern Processes (with Chestnut Soils of the Kulunda Steppe as an Example)* (Ross. Akad. Nauk, Novosibirsk, 2005) [in Russian]-.a
- I.V. Mikheeva, "Spatial Fluctuations and Statistical Probability Distributions of Chestnut Soil Properties in the Kulunda Steppe", *Pochvovedenie*. No. 3, 316-327 (2005) [*Eur. Soil Sci.* 38 (3), 278-288 (2005)]. -b
- Jenny, H., *Factors of Soil Formation: A System of Quantitative Pedology* (McGraw-Hill, New York, 1941; Inostrannaya Literatura, Moscow, 1948).
- B. Yu. Lemeshko, *Statistical Analysis of Univariate Observations of a Random Variable: a Program System* (Novosibirsk, 1995) [in Russian].
- All-Union Guidelines for Soil Survey and Compilation of Large-Scale Maps of Land Tenure* (Kolos, Moscow, 1973), p. 95 [in Russian].
- I.V. Mikheeva and E. D. Kuz'mina, "Statistical Characteristics of Soil Particle-Size Composition Formula." *Pochvovedenie*, No. 7. 818-828 (2000) [*Eur. Soil Sci.* 33 (7), 713-722 (2000)].



Dr. Irina V. Mikheeva. is a lead researcher in the Laboratory of geography and genesis of soils of the Institute of Soil Science and Agrochemistry of Siberian Branch of Russian Academy of Sciences (Novosibirsk). She graduated from the Novosibirsk

Technical University as applied mathematician. Then she specialized as ecologist in Soil Department of Moscow State University, and then re-

ceived her PhD in 1993, and doctor degree in 2002 in biology (soil science). The areas of her interests and experience include mathematical methods, mostly probabilistic methods and approaches, in soil science and ecology, quantitative evaluation of soil processes caused by natural and anthropogenic impact.

[mikheeva@issa.nsc.ru](mailto:mikheeva@issa.nsc.ru) ; [pulya80@ngs.ru](mailto:pulya80@ngs.ru)

### Workshop on Statistical Aspects of National-Scale Soil Monitoring

11 – 12<sup>th</sup> December 2008  
Rothamsted Conference Centre  
Rothamsted Research  
Harpenden, AL5 2JQ  
England

We invite contributions to a workshop on statistical aspects of national-scale soil monitoring. The workshop will consist of invited keynote presentations, contributed talks and discussions within break-out groups. The statistical challenges of implementing monitoring networks will be discussed under three broad themes: (i) Design of monitoring networks, (ii) Analyses of Data, (iii) Interactions with policy-makers. Further details are available at:

[http://www.pedometrics.org/articles\\_view.asp?id=55](http://www.pedometrics.org/articles_view.asp?id=55)

Ben Marchant & Murray Lark

Email: [ben.marchant@bbsrc.ac.uk](mailto:ben.marchant@bbsrc.ac.uk)

Abstract is  
Due 31 May



### 3rd Global Workshop on Digital Soil Mapping

30<sup>th</sup> Sept. – 3<sup>rd</sup> October 2008  
Utah State University  
Logan, Utah, USA

Digital Soil Mapping: Bridging Research, Production, and Environmental Applications

Digital soil mapping is advancing on different fronts at different paces throughout the world. The goal of the third workshop is to review and discuss the state-of-the art in digital soil mapping, and explore strategies for bridging research, production, and environmental application. For more information on program, abstract submission, registration, travel and housing, visit:

<http://dsmusa.org/>

## Upcoming Events

Accuracy 2008, The eighth symposium on spatial accuracy assessment in natural resources and environmental sciences. 25-27 June 2008. Shanghai, China  
<http://2008.spatial-accuracy.org/>

The First International Conference on HYDROPEDOLOGY. 28-31 July 2008. Penn State University, USA.  
[hydropedology.psu.edu](http://hydropedology.psu.edu)

EUROSOIL 2008, 25 - 29 August 2008, Vienna, Austria.  
[www.ecsss.net](http://www.ecsss.net)

The 3rd Global Workshop on Digital Soil Mapping. Utah State University, Logan, USA, October 2008.  
<http://www.digitalsoilmapping.org>

International Geostatistics Congress Santiago, Chile 1-5 Dec 2008 <http://www.geostats2008.com/>

# Alex's Preferred Pedometrics Papers III

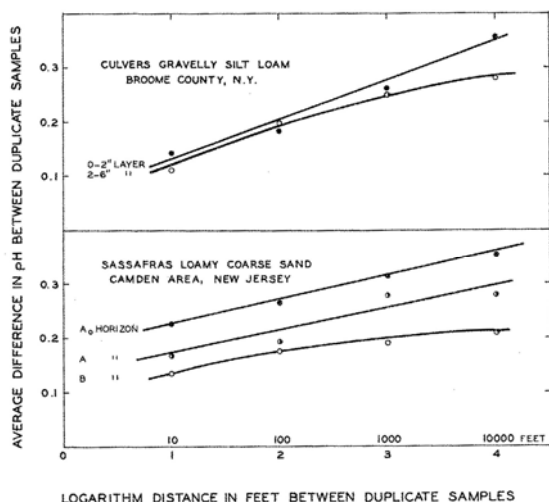
My top six most preferred pre-geostatistical pedometrical papers (3)

Youden, W.J., Mehlich, A., 1937. Selection of efficient methods for soil sampling. Contributions of the Boyce Thompson Institute for Plant Research 9, 59-70.

This is a true classic - it's pedometrics epitomised precociously - I can find no other way of summarising this paper. What do I mean? A potent mixture of sound soil science and statistics - Youden was a first-rate statistician and Mehlich a world-leading soil scientist. So with this admixture of science, statistics and focus we had an early model for good Pedometrics.

**What did they do?** They told us, perhaps originally, how soil varies with distance quantitatively. They discovered that there was a lot of variance at short distances, but also that it increased with distance. Seventy years later, can we answer why? They devised the classical spatial nested design and used methods-of-moments analysis to estimate the variance components. Youden and Mehlich suggested near the end of their paper that considerations of soil variation and designs should be given the same consideration as the design and analysis of yield experiments which were being much studied at that time. I think the proper understanding of the relationship between soil and yield and a good statistical model for that relationship, still eludes us.

**Where did it take us?** Over the years people followed the design, e.g., Webster & Butler (1976), Nortcliff (1978), and others later. Eventually we understood that the information was essentially that of the variogram - this was pointed out most



clearly by Miesch (1975). The true value of this information only came into its own with the soil geostatistical work of Giltrap, and Burgess & Webster in the UK, and Nielsen and others, in the USA. Some of us found Youden and Mehlich's hierarchical nested design rather costly for the shorter distances and preferred unbalanced designs but we needed REML to estimate the components properly (Pettitt & McBratney, 1993). The idea of distances on some sort of logarithmic spacing is still a powerful one when one has no idea of the spatial scale of variation. The combination of this idea with a grid has been suggested by several people (or just imagined by me?) as a 'one-stop-shop' design for combined variogram estimation and kriging.

Of course, one of the marvellous things they did was to publish all their data. They are there for us all to analyse. Great pedometric papers often contain the data as well as their analysis.



Fig. 1. A cartoon from the July 30, 1937 edition of the *Richmond (Virginia) Times Dispatch*, entitled "What Came Out of It", the same day the Sassafras soil was sampled. Perhaps one of our American readers can explain the context of this cartoon. Something to do with the 1936 US presidential election?

Verbal imagery is often missing from the modern paper - but not in 1937. In the Culvers gravelly silt loam ‘wild strawberries were invariably present at the sampling points’ and for the Sassafras loamy coarse sand ‘blueberries were always found in the areas sampled.’ We are told the exact dates of the sampling work - the Culvers soil was sampled on June 22<sup>nd</sup> 1937 - the day Joe Louis KOs James J. Braddock in 18 for the world heavyweight boxing title in Comiskey Park, Chicago, also the day Camille Chautemps (remember him?) becomes Prime Minister of France. The Sassafras soil was sampled on July 30<sup>th</sup>, a Friday. On the same day the League of Nations Commission rejected Great Britain’s request for immediate discussion of the Palestine question..... la plus ça change .....

Bibliometrically, this paper is a quiet achiever showing a late flourish. From 1950 onwards the paper has had a citation rate of around 1.5 per year which is slowly increasing (Fig. 2). Most of the papers before 1975 were from statisticians pure and applied - the general soil science recognition came after that date, peaking around 1995 (Fig. 3) - and it’s still being cited.

Finally, I found another thread that had eluded me in earlier study; there was a practical environmental motive for their work. After all, they were working in the heavily-populated and industrialised states of New York and New Jersey. “The problem of distinguishing whether differences in properties found between widely separated areas are such as occur naturally in the [soil] type or are indicative of the superimposition of unusual conditions, also requires knowledge of the variation found within the [soil] type. In some cases soils have been exposed to

industrial waste gases, such as sulphur dioxide, and it becomes apparent to ascertain whether the soil has been damage [acidified], and if so, the limits of the area affected. It should be clear that supposed differences in properties must be considered in view of the distance involved.” I don’t recall an application that has approached the problem quite from this point of view!

There was so much new in this paper that it has taken 70 years for its significance to be fully uncovered.

References

Miesch, A.T., 1975. Variograms and variance components in geochemistry and ore evaluation. In *Quantitative Studies in the Geological Sciences* (ed. E. H. T. Whitten). Geological Society of America Memoir 142,333-340.

Nortcliff, S., 1978. Soil variability and reconnaissance soil mapping: a statistical study in Norfolk. *Journal of Soil Science* 29, 403-418.

Pettitt, A.N., McBratney, A.B., 1993. Sampling designs for estimating spatial variance components. *Applied Statistics - Journal of the Royal Statistical Society Series C* 42, 185-209.

Webster, R., Butler, B.E., 1976. Soil classification and survey studies at Ginninderra. *Australian Journal of Soil Research* 14, 1-24.

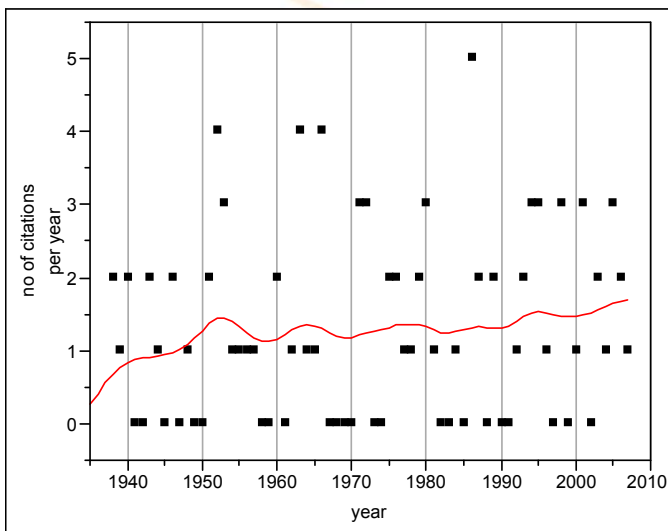


Fig. 2. Citations per year of Youden and Mehlich (1937) by all scientists.

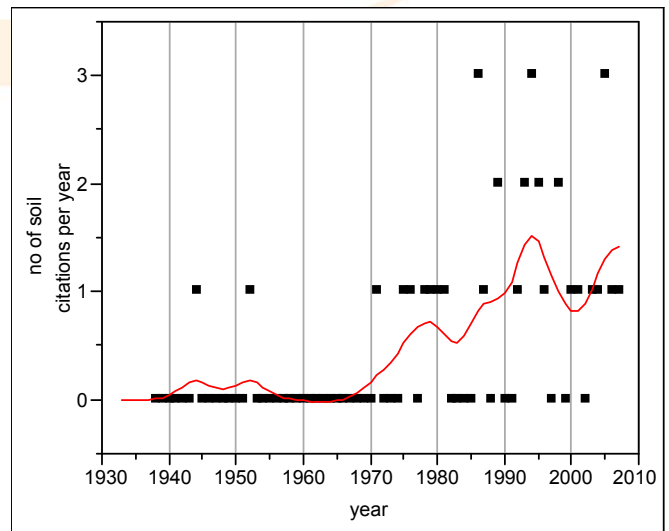
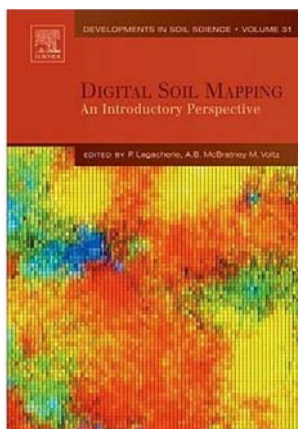


Fig. 3. Citations per year of Youden and Mehlich (1937) by soil scientists.

# Book Review



**Digital soil mapping—An introductory perspective**, *Developments in Soil Science*, vol. 31, edited by P. Lagacherie, A.B. McBratney and M. Voltz. Elsevier Science, Amsterdam.

Soil mapping has come a long way. During the late 1970s and early 1980s the first digital soil databases were made, followed by soil information systems, computer aided cartography and GIS.

These developments facilitated fast recombination and reclassification of soil data for producing maps of derived properties, performing land evaluation and constructing parameter fields for dynamic models. However, the content of these soil information systems was derived from digitized soil maps and descriptions obtained by field surveys. During the 1980s, the application of geostatistics to soil science and the development of Pedometrics made it possible to partly automate the mapping process: first by interpolation of soil attributes, followed by the interpolation of soil classes and the concurrent mapping of soil properties and classes using auxiliary information such as digital terrain models and remotely sensed imagery. The combination of automated mapping of soil classes or attributes and the tools for inferring derived mapping products and parameters is called “Digital Soil Mapping”. It has gained popularity because it facilitates soil mapping in large and remote areas, cuts labour costs and it is easier to quantify the uncertainty.

Development in digital mapping and inference methods warranted a conference in 2004 in Montpellier and the book edited by Lagacherie, McBratney and Voltz are its proceedings. The book consists of 41 contributions representing the state-of-the-art in digital soil mapping, with methodological papers as well as papers providing applications. The book offers three things:

- (i) an excellent vision paper by Lagacherie and McBratney providing a blueprint for “spatial soil inference systems”, and identifying a number of challenges that digital soil mapping will be facing in the near future;
- (ii) a number of theoretical and methodological “goodies” of interest to researchers; and
- (iii) inspiration for practitioners that search for digital mapping and inference methods.

The papers are grouped into 6 different sections. Section 1 shows examples (e.g. from Australia and Brazil)

This is an altered version of a review that will appear in a forthcoming issue of the *Geoderma*

in which digital soil mapping has moved from the research phase to effective production of soil maps for large areas. Section 2 focuses on the architecture and creation of soil databases, where attention is paid to semantic matching of descriptions from multiple sources and a framework is given for including information about the quality of soil data for subsequent error propagation. The majority of papers in Section 3 deal with sampling theory. Although soil sampling remains necessary to ground truth digital mapping projects, the chapters are somewhat at odds with the remaining parts of the book. They could have been published in a soil science or geostatistical journals and the scientific quality of these contributions is very high.

Instrumental to digital soil mapping, especially in data-poor environments, is the use of environmental covariates, either in geostatistical, classification or regression type of methods. In Section 4, various papers deal with new types of covariates, mostly from remote sensing and digital elevation models. Section 5 starts with three papers on the construction of spatial models of soil variability: one presenting a way of including a priori pedological knowledge, one including pedological models and hydrological information and the last paper using wavelets for a multi-scale interpretation of soilscapes. After these three theoretical papers, a large number of contributions are listed with example application of mapping of soil classes and soil attributes. The final section deals with soil map representation, with contributions about quality assessment of soil attributes, representing uncertain boundaries and map visualization.

The lay-out of the book is of high quality, with a nice uniform chapter format, high quality figures, thick glossy paper and a hardback cover. I would highly recommend it to researchers that would like to gauge the state-of-art of the subject, as well as to practitioners searching for examples applicable to their own mapping problem. The subtitle of the book is “An introductory perspective”, by which the editors suggest that the book is just at the beginning of things. Indeed, certain areas are not covered such as spatio-temporal mapping (soil wetness, carbon pools) and mechanistic pedogenetic modelling. Hopefully, future Workshops on Digital Soil Mapping will have contributions on these topics. Moreover, global change research and modelling is in desperate need of a high-resolution (90 m × 90 m) global map of soil physical and soil chemical properties, or better “a global spatial soil inference system” Would this be a splendid activity for the digital soil mapping community?

Marc Bierkens  
Department of Physical Geography,  
Utrecht University, The Netherlands  
E-mail : [m.bierkens@geo.uu.nl](mailto:m.bierkens@geo.uu.nl)



## Pedometrician profile

Beate Zimmerman

Smithsonian Tropical Research Institute,  
Panama



### How did you first become interested in soil science?

During practical experience after high school I assisted scientists with their fieldwork, which included soil sampling. This experience prompted me to study geocology instead of biology. My real start with soil science was during my master thesis, and since then I continued working on soil hydrology in the humid tropics.

### How were you introduced to pedometrics?

This was during my Ph.D. research. I attended a workshop about pedometrics held by Murray Lark and Andreas Papritz, which helped considerably to upgrade my own data analysis. Last summer I visited the Pedometrics conference for the first time.

### What recent paper in pedometrics has caught your attention and why?

Lark, R.M., B.R. Cullis, and S.J. Welham (2006). On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science*, 57, 787-799.

I like this paper because it presents a complex problem in a comprehensible way. The approach is transferable to soil hydraulic data, which I have focussed on so far.

McKenzie, N., and D. Jacquier (1997). Improving the field estimation of saturated hydraulic conductivity in soil survey. *Australian Journal of Soil Research*, 35, 803-825.

This is a very interesting idea, which I like particularly because of its applicability to remote areas, where field determination has to substitute for laboratory analysis.

### What problem in pedometrics are you thinking about at the moment?

I'd like to gain further insight in how to use mixed modelling to describe the spatial variation of environmental variables over a variety of scales. This would entail separating variation into its random and deterministic component at every scale of interest. In addition, the predictive power of ancillary variables also changes with scale. To combine those issues into a reliable model, for example a geostatistical model or a pedotransfer function, appears to be a challenge at any rate. In many tropical countries, difficult field conditions and data scarcity further complicate matters.

### What big problem would you like pedometricians to tackle over the next 10 years?

Developing up-scaling and prediction tools, which are flexible enough for application in many parts of the world. Improving the link between pedometrics and scaling issues in hydrology.

Beate Zimmermann recently obtained her Ph.D. from the University of Potsdam for a thesis on spatial and temporal variability of saturated hydraulic conductivity of soils in Ecuador which had undergone differing degrees of disturbance. She now works for the Smithsonian Tropical Research Institute in Panama, studying soil hydrological properties on catchments with contrasting vegetation and management.

## Non-Pedometrician profile

Alfred Hartemink

ISRIC—World Soil Information



### How did you first become interested in soil science?

As a toddler the sandpit was my domain and I enjoyed both the feel and taste (to the extent that my mother took me to the GP who comforted her by saying that it would not last forever). I remember seeing my first soil pit during a biology class - a sort of podzol - and I enjoyed the afternoon in the field and pit. I was fascinated by the combination of physics, chemistry etc into something that visibly affected the real world: plant growth. As I saw the inequality in the world I decided to focus my study on tropical regions, and with time I specialised in soil science but still find the broad picture as interesting as the small detail.

### What are the most pressing questions at the moment in your area of soil science?

I have been a soil surveyor for some years and later on moved into soil fertility. Broadly speaking soil survey has to incorporate soil processes and temporal aspects whereas soil fertility has to come out of the plot and test-tube and become spatial. One of the most pressing questions is how we make a new soil properties map of the world, but there are several others.

### What statistical and mathematical methods are used in your area of soil science?

Ordinary statistics, nothing fancy.

### Are you aware of any work by pedometricians that might be relevant to your science?

Yes. I try to read widely - also the papers that I find hard to understand (or don't have my interest *per se*). Much is happening in soil science - there is an awful lot of repetition but there are also new developments in for example modelling soil formation. It is sometimes hard to separate my interest from what is relevant.

### What big problem would you like pedometricians to tackle over the next 10 years?

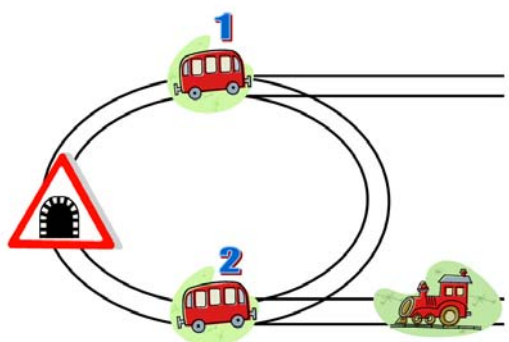
There are methodological issues and issues regarding the status of pedometrics in the soil scientific community. I think all of us would like to see more accurate predictions of a range of soil properties at ever-finer resolutions. That is a clear goal and the integration of soil depth and temporal matters with models of other disciplines needs tackling. The problem might be to come up with something really new - that should also be a goal in itself. Furthermore: pedometrics should become mainstream. One way of doing that - and that is maybe not a popular thing to say in the pedometrics newsletter - is to drop its name and abolish the artificial boundary between the pedometrics community and other soil scientists. Overall, the soil science community needs defragmenting to become more effective and visible - that will be good for all pedometricians.

# Pedomathemagica

With Gerard Heuvelink

## Problem 1 (EASY?)

Two train wagons 1 and 2 with pedometricians enjoying an excursion must change places. The locomotive can pull and push, can drive backwards and forwards and can push and/or pull two wagons at the time. It can also pass through the tunnel, but the wagons cannot (because the pedometricians are afraid of the dark). Once the wagons have changed places, the locomotive must return to the original location. What is the minimum number of movements needed to accomplish this task (each stop of the locomotive means one movement)?



## Problem 2 (MEDIUM-HARD)

Three pedometricians Rosina, Achim and Pierre argue about who is to be first author of their joint paper. They decide to solve it in an original way and fight a duel. They stand in a triangle, each with a pistol armed with (paint?) bullets, and let chance decide who may fire first, second and third. Once a turn of three shots is completed, firing is repeated in the same order as before. A person that is hit is out and the others continue until only one person is left. Rosina is an excellent marksman and never misses, Achim is successful in 80 per cent of his shots and Pierre has only a 50 per cent chance of hitting the target. Each individual maximizes their chances. Who has the biggest chance of becoming first author? How big is this chance?

## The Answers??

Next issue of Pedometron (Sometime at the end of the year 2008)



## ☆ Pedometrics 09 ☆

Will be held in Beijing, October 2009, hosted by China Agricultural University. Mark your diary and watch out for the date!

# Pedomathemagica

## Answers to last issue's problems

### Problem 1

We might solve the problem using trial and error. First note that David must have two more bags than Tom because each has the same number of bags when David gives one of his bags to Tom. Now suppose that Tom has two bags and David four. If Tom gives one to David then Tom has only one left and David has five, this is not twice as many. When Tom starts with three and David with five, then Tom handing over one bag to David leaves him with two bags whereas David has now six, still not twice as many. It also does not work when Tom has four and David six bags, but when Tom starts with five and David has seven, then Tom will have four left after giving one to David, and David will have eight, which is twice as many as four. The solution thus is that Tom has five and David seven bags.

We can also solve the problem in a more mathematical way. Let  $T$  be the number of bags of Tom and let  $D$  be the number of bags that David has. These variables must satisfy:

$$D-1 = T+1 \rightarrow D = T+2$$

$$D+1 = 2 \times (T-1)$$

Combining yields:

$$T+2+1 = 2T-2 \rightarrow T = 5, D = 7$$

### Problem 2 (HARD)

First some definitions. Five hundred kilometres will be called a 'unit', the amount of petrol sufficient to take the jeep 500 kilometres will be called a 'load' and a 'trip' is a journey of the jeep in either direction from one stopping point to the next.

Two loads will carry the jeep a maximum distance of 1 and  $1/3$  units. This is done in four trips by first setting up a depot at a spot  $1/3$  unit from the start. The jeep begins with a full load, goes to the depot, leaves  $1/3$  load, returns (and has an empty tank upon return), picks up another full load, arrives at the depot and picks up the depot's  $1/3$  load. It now has a full load again, sufficient to take it to the end point 1 and  $1/3$  units from the start.

Three loads will carry the jeep 1 and  $1/3$  plus  $1/5$  units in a total of nine trips. The first depot

is  $1/5$  unit from the start. Three trips put  $6/5$  loads in the depot. The fourth trip takes the jeep back to the start point, it picks up the remaining full load and arrives at the first depot, with  $4/5$  load in its tank. This, together with the fuel in the depot, makes two full loads, sufficient to carry the jeep the remaining 1 and  $1/3$  units, as explained in the preceding paragraph.

We are asked for the minimum amount of petrol required to take the jeep 800 kilometres. Three loads will take Maggie 766.7 kilometres (1 and  $1/3$  and  $1/5$  units), so we need a third depot at a distance of 33.3 kilometres from the start. In five trips the jeep can build up this depot so that when the jeep reaches the depot at the end of the seventh trip, the combined fuel of jeep and depot will be three loads. As we have seen, this is sufficient to take the jeep the remaining distance of 766.7 kilometres. Seven trips are made between starting point and first depot, using  $7/15$  load of petrol. The three loads of fuel that remain are just sufficient for the rest of the way, so the total amount of petrol consumed will be 3 and  $7/15$ , which corresponds with 1733.3 kilometre.

Clearly the above is not a rigorous proof that 1733.3 km is the minimum. Perhaps such a proof may be obtained using theory from Linear Programming and Operations Research, but I was not able to find or construct it. One aspect that suggests that we may not yet have reached the optimal solution is that five trips are used to build the third depot 33.3 km from the start. The third depot will then have stored 3 -  $6/15$  loads. In order not to have to leave petrol behind at the third depot, Maggie must start the seventh trip with a load of only  $7/15$ , which seems suboptimal. Does anyone have a better solution?

There is no maximum to the desert width that Maggie can cross. Using the approach described above, the maximum width that she can cross is  $(1 + 1/3 + 1/5 + 1/7 + 1/9 + 1/11 + 1/13 + \dots) \times 500$ . This sum is infinitely large because:

$$\begin{aligned} & 1 + 1/3 + 1/5 + 1/7 + 1/9 + 1/11 + 1/13 + 1/15 + \dots \\ &= 1 + 1/3 + (1/5 + 1/7) + (1/9 + 1/11 + 1/13 + 1/15) + \dots \\ &> 1 + 1/4 + 2/8 + 4/16 + \dots = 1 + 1/4 + 1/4 + 1/4 + \dots = \infty \end{aligned}$$