SSSAJ

Soil Science Society of America Journal

Vol. 68 No. 2 March–April 2004

SSSJD4 68(2):339–712 (2004)

SSSAJ March–April 2004 Vol. 68, No. 2

Soil Science Society of America Journal

Division S-1—Soil Physics

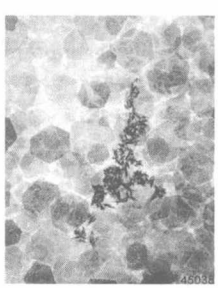
Division S-1—Notes

Division S-2—Soil Chemistry

Division S-3—Soil Biology & Biochemistry

Division S-3—Notes

Continued on page ii



This issue's cover: Transmission electron microscope (TEM) image (50 000× magnification) of a mixture of Fe- and Al-oxide minerals: hexagonal gibbsite [α-$Al(OH)_3$] platelets and aggregated, lathe-shaped goethite (α-FeOOH) particles. Iron- and Al-oxides are considered the most important minerals for retaining PO_4 in soils, but their relative importance as they coexist in soils is unclear. Khare et al., "XANES determination of adsorbed phosphate distribution between ferrihydrite and boehmite in mixtures", p. 460–469, utilized X-ray absorption spectroscopy to quantify how PO_4 distributes itself between Fe- and Al-oxides in a mixture to estimate the relative affinity of each mineral for PO_4.

Continued from page i

SOIL SCIENCE SOCIETY OF AMERICA
JOURNAL

VOL. 68 MARCH–APRIL 2004 No. 2

DIVISION S-1—SOIL PHYSICS

Minimizing Drift in Electrical Conductivity Measurements in High Temperature Environments using the EM-38

D. A. Robinson,* I. Lebron, S. M. Lesch, and P. Shouse

ABSTRACT

The EM-38 is a noninvasive instrument, commonly used for monitoring salinity, mapping bulk soil properties, and evaluating soil nutrient status. Users in the Southwest USA have observed as much as 20% "drift" in the measurement of bulk soil electrical conductivity (EC_a) with this instrument. This drift has usually been ignored or compensated for by statistical procedures. We performed laboratory and field experiments to determine if the drift is due to calibration instability of the instrument or to heating of the instrument by the sun. In laboratory experiments, after a warm-up period, the instrument provided constant readings in the range 25 to 40°C; above 40°C the response of the instrument was unpredictable. In field experiments, where we placed the EM-38 in a fixed location we observed an unexpected response at air temperatures below 40°C. Temperature sensors in different locations on the instrument demonstrated that temperature differences between the instrument's transmitting and receiving coils and the control panel (CP) were as great as 20°C. As the instrument is temperature compensated from this CP, erroneous compensation occurred when the instrument was placed in direct sunlight. In this study, we demonstrate that differential heating of the EM-38 is one cause of drift and erroneous bulk electrical conductivity measurement; shading the instrument substantially reduced this problem, effectively extending the reliable working temperature range by minimizing drift.

THE CONCEPT OF USING induced electromagnetic fields to measure ground conductivity has been applied in the geosciences for more than 50 yr (Belluigi, 1948; Wait 1954, 1955, 1982). Induction methods were used extensively for ore prospecting as metallic ore bodies can have substantial electrical conductivity (Keller and Frischknecht, 1966). They were also used for well logging in the petroleum exploration industry (Keller and Frischknecht, 1966). Noninvasive instruments were first considered for use in agriculture by De Jong et al. (1979). Since then the technique has been used to map a variety of physical quantities with which EC_a correlates (e.g., salinity, moisture, and clay content). Water content has been estimated from measurements of EC_a by Kachanoski et al. (1988) and Sheets and Hendrickx (1995), salinity by a number of authors (Corwin and Rhoades, 1982; Wollenhaupt et al., 1986; Hendrickx et al., 1992; Rhoades, 1993; Lesch et al., 1995a, 1995b; Rhoades et al., 1999), and inferring differences in mineralogy by Triantafilis et al. (2000). Increasingly, applications are being identified in precision agriculture for determining nutrient status and potential yield (Corwin and Lesch, 2003; Corwin et al., 2003).

The EM-38 has been adapted for general mapping in agriculture, an example is the Lower Colorado Region Salinity Assessment Program. This is a network of people and organizations that are committed to improving the assessment of soil salinity in agricultural fields in the Southern Colorado region to guide management decisions (http://www.ussl.ars.usda.gov/lcrsan/LCRhome.htm; verified 7 Oct. 2003). Soil mapping survey units consisting of converted spray rigs, mounted with dual dipole EM-38 units and GPS, have been used to map agricultural fields (Rhoades, 1993; Lesch et al., 1995a, 1995b; Triantafillis et al., 2002). Data has been analyzed using ESAP computer software to produce maps and statistical sampling plans (http://www.ussl.ars.usda.gov/MODELS/esap-95.htm; verified 7 Oct. 2003). As this network of users has developed, large amounts of data have been collected and some anomalous results have been observed.

The term drift has been used to describe disparate values in EM-38 data, collected at different times from the same location, that cannot be accounted for by changes in water content or soil temperature. The causes

D.A. Robinson, I. Lebron, S.M. Lesch, and P. Shouse, USDA-ARS, George E. Brown, Jr. Salinity Lab., 450 W. Big Springs Road, Riverside, CA 92507. Received 26 Feb. 2003. *Corresponding author (darearthscience@yahoo.com).

Published in Soil Sci. Soc. Am. J. 68:339–345 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: CP, control panel; EC_a, bulk soil electrical conductivity; H_i, induced magnetic field; H_p, primary magnetic field; Rx, receiving coil; Tx, transmitting coil.

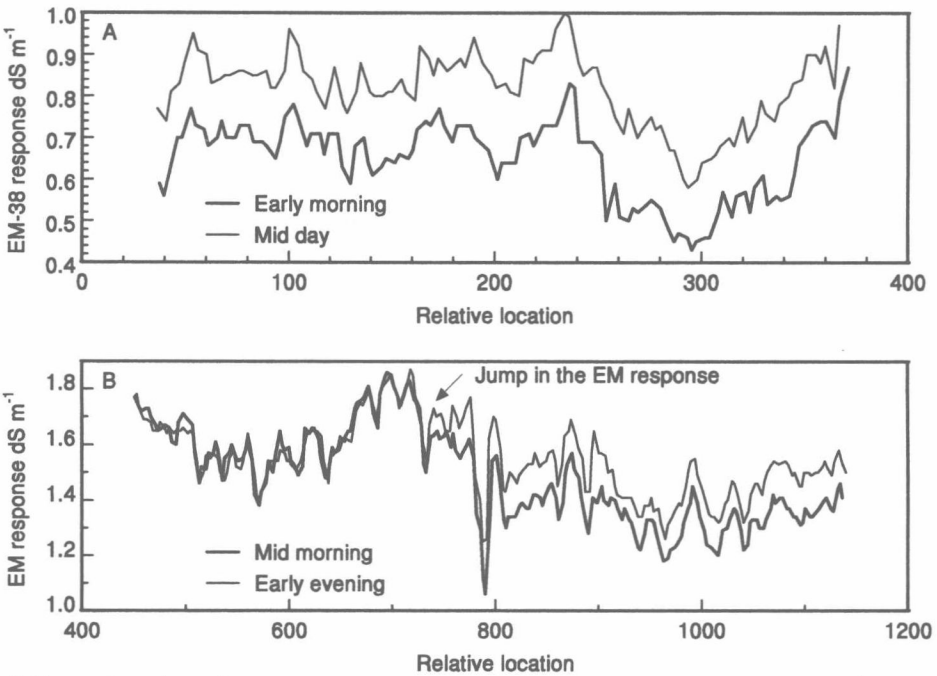


Fig. 1. (A) Data from field mapping where the same row was run in the morning and again later in the day showing responses that follow each other with an offset. (B) A similar data set from another field site run in the morning and again later in the day, where the response jumps at point 730.

of drift are considered to be two-fold, the first arising from the coil spacing and the second due to thermal distortion of the coils. In the case of the EM-38 the spacing between the coils is fixed, the second forms the subject of this piece of work. In an experiment with a static EM-38 Sudduth et al. (2001) noted that as temperature increased, the in-phase I/P reading of the EM-38, which should be zero decreased and as a result the measured EC_a was observed to increase. It is common practice to take measurements along a plough row in a field and then later in the day return to that row to repeat the measurements (Sudduth et al., 2001). This operation is known as running a drift row. Figure 1A shows the values of EC_a along a row where measurements were collected early in the morning and later in the day. The readings correspond to the instrument in the vertical orientation. In Fig. 1A, the readings of the later run are offset by an increase of about 20%. Given the depth of EC_a measurement, this could not be accounted for by changes in soil temperature.

Figure 1B shows another anomaly between EC_a data collected in the morning and in the evening. The data from the evening initially coincides with the data from the morning, then unexpectedly a jump occurs wherein the data follows the same pattern but is shifted upwards by 10%. The cause of these anomalies is unknown and users have suggested that the calibration of the instrument is unstable. If this is the case, maps of soil properties, such as nutrient status, can potentially contain substantial errors.

Our impression from working with the instrument was that drift effects appeared more pronounced on hot sunny days. The Geonics manual suggests that the working range of the instrument is 5 to 50°C. However, in a personal communication with Geonics, we were informed that the instrument is only temperature compensated up to 40°C. Each probe then has its own systematic temperature response characteristic above this temperature. With this information in mind, we designed a set of experiments to determine the effect of environmental temperature on the repeatability of EM-38 measurements. Our objective was to determine the source of drift by: (i) evaluating the stability of EM-38 calibration and (ii) determining the impact of heating on the EM-38 response. These experiments were performed in the laboratory and under field conditions to find a solution for instrument drift and to reduce measurement error.

MATERIALS AND METHODS

Instrument Background

A schematic diagram of the EM-38 is presented in Fig. 2 showing the location of induced magnetic fields during operation. A transmitting coil (Tx) in one end of the instrument creates a primary magnetic field (H_p). This field creates current loops in the ground below and the current loops induce their own magnetic field (H_i). The induced field is superimposed on the primary field and both H_p and H_i are measured in a receiving coil (Rx) at the other end of the instrument (McNeill, 1980). The measured response is a function of ground conductivity, which is linear in the range of soil conductivity of 0 to 10 dS m^{-1}. Measurements of ground conductivity can be made with the instrument in either the vertical or horizontal orientation. In the horizontal orientation the instrument measures to a depth of about 0.75 m with the greatest sensitivity just under the instrument. With the instrument in the vertical orientation it measures to a depth of about 1.5 m with the greatest sensitivity at about 0.4 m.

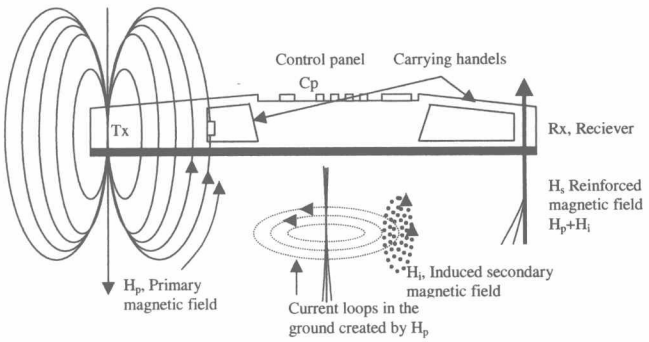


Fig. 2. Schematic diagram of the EM-38, which is 1 m in length. Tx is the transmitting coil and Rx is the receiving coil. Locations CP and Rx are where the temperature sensors were placed.

A standard single dipole EM-38 was used throughout the experiments. A second single dipole EM-38 was used to replicate both the indoor and outdoor experiments. The instruments were calibrated using the described standard method. The probe was placed 1.5 m above the ground on a wooden support; the vertical and horizontal readings were adjusted until the vertical read twice the value of the horizontal. The instruments were calibrated after a warm-up period of 2 h. The calibration was checked for consistency after each experimental run.

During the experiments we measured the temperature of the air, soil, and two parts of the instrument, the CP under which the instrument circuit is located, and at the receiving coil Rx (Fig. 2). Thermocouples, connected to a Campbell CR10x data logger (Campbell Scientific Inc., Logan, UT) were used to record the temperature every minute.

Controlled Experimental Setup Indoors

Indoor experiments were conducted with the EM-38 so that the temperature of the surroundings could be controlled. The first objective was to verify the reliability of EM-38 calibration. The EM-38 was calibrated and placed in a large room where temperature was maintained at 22 ± 1°C. The instrument was placed on a plastic drum, 1 m above the ground and kept in the vertical orientation for all experiments. By doing so the instrument response to the ground, primarily the rebar (iron rods) in the concrete could be evaluated. EM-38 measurements were taken every minute and recorded on a Polycorder located several meters from the instrument.

The second objective was to determine the reliability of the instruments temperature compensation. This was performed by warming the instrument with an electric blanket. Preliminary tests were conducted to ensure that the blanket did not interfere with the response of the EM-38. The response was measured without the blanket, with the blanket wrapped around the central 50 cm of the instrument and with the instrument completely covered. No effect was observed, the same conditions were repeated, but this time with the blanket switched on. Finally, the blanket was switched on and off repeatedly to see if this had any impact on the EM-38 response, again no effect was observed.

We conducted two experiments using the blanket to heat the instrument. The first determined the effect of differential instrument heating and the second determined the effect of uniform instrument heating. In the first of these experiments localized heat was applied to the EM-38 circuit (CP, Fig. 2) in the central 50 cm of the instrument while maintaining the rest of the instrument and the environment at constant temperature and constant electrical conductivity.

The second experiment was used to determine the response of the instrument to uniform warming. This time the entire instrument was wrapped in the blanket and heated. In both experiments the temperature of the instrument was raised to a maximum of 55°C. This is a temperature commonly experienced during summer in the Southern USA.

Outdoor Experiments

Outdoor experiments at the U.S. Salinity Laboratory were conducted on bare soil (Arlington, sandy loam) that was irrigated once per day at 0600 h. Measurements were made on a series of warm sunny days in June and July of 2002 when the weather was similar to that commonly experienced during typical fieldwork. The EM-38 response was recorded continuously over a 10-h period beginning at 0900 h using a Polycorder located several meters from the instrument under shade. High temperatures did not affect the performance of the Polycorder. The experiments were run with the EM-38 in the vertical orientation. This allowed measurements of EC_a to be obtained from a depth where soil is least subject to changes in temperature or water content. The EM-38 was positioned on a 2.5-cm thick piece of wood placed on the ground to prevent heating from the soil and to ensure the same daily location. Soil temperature (10-cm depth) was also monitored at the beginning of each experiment. This was performed around mid-day and in the late afternoon, using a handheld temperature probe. The calibration of the instrument was checked periodically and found to be consistent. A final experiment on an asphalt surface was performed by placing the EM-38 in the vertical position on a 2.5-cm thick wood on top of asphalt. During the first 160 min the instrument was shaded, after that time the shade was removed and EC_a and temperatures at CP and Rx were recorded for 600 min.

RESULTS AND DISCUSSION

Controlled Experiments Indoors

Experiments were conducted indoors in a controlled environment to best define EM-38 response to constant temperatures, differential heating, and elevated temperatures. In the first experiment the instrument was switched on and run continuously at a constant air temperature of 22°C. The readings of the EM-38 were constant during a 12-h uninterrupted time period. This simple experiment was necessary to test the stability of the calibration of the EM-38. Since no jumps or sudden changes in EC_a were recorded and the readings remained constant this demonstrated that the cause of the drift was not unstable calibration.

In the next experiment the central section of the EM-38 containing the instrument circuit board was warmed using an electric blanket, while the transmitting and receiving coils were maintained at the ambient room temperature. The temperature of the receiving coil, floor, and air were monitored and remained constant at 22°C. The response, which was replicated by another single dipole EM-38 (data not shown), showed that as the instrument panel and circuitry warmed up, the instrument electrical conductivity response decreased (Fig. 3). This suggested that the instrument temperature compensation was located at, and controlled by, the instrument circuit board under the black CP (Fig. 2). This also suggests that the temperature compensation is provided for the coils and not the circuit. If the circuit were

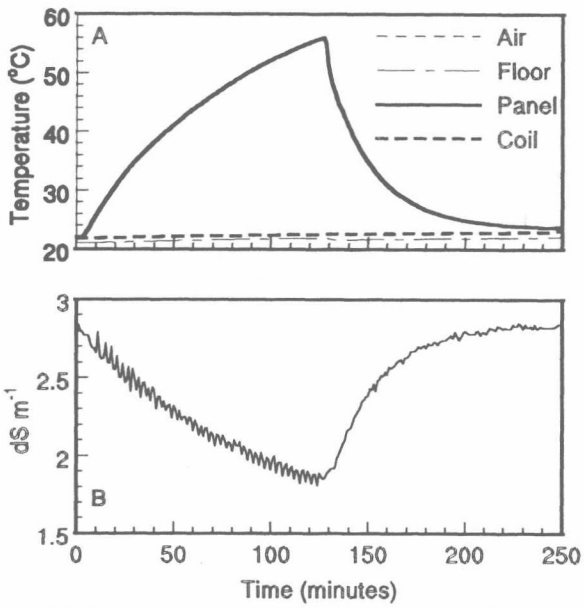


Fig. 3. (A) Temperature response for control panel (CP) and receiving coils (Rx) of the EM-38 with the central 50 cm of the instrument wrapped in an electric blanket and warmed. (B) The soil bulk electrical conductivity (EC_a) response, which reduces as the instrument panel is warmed.

temperature compensated, a constant EC_a value would be expected. The temperature compensation design assumes that the instrument is at a uniform temperature, and that the coils (Rx in Fig. 2) are at the same temperature as the circuit. These findings were corroborated by the third experiment.

The results from the third experiment with the instrument completely wrapped in the blanket and heated uniformly are presented in Fig. 4. This experiment demonstrates more clearly the temperature compensation of the instrument. For the first 120 min, while the temperature is below 40°C, the instrument effectively compensates for temperature changes and provides a constant reading. Above 40°C, when the increase in temperature is more abrupt, the EC_a value measured by the EM-38 increased. The EC_a reading rose from 2.8 dS m^{-1} to a maximum value of 3.2 dS m^{-1} occurring at 46°C (Fig. 4B). As the temperature continued to rise the EC_a reading began to decline (Fig. 4C). When the instrument was allowed to cool down the EC_a response dropped below the initial value. By the time we checked the probe the following day the reading was back to normal and the instrument was in calibration. It is interesting to observe that the temperature of the circuit, under the CP, was slightly higher than the temperature of the receiver (dashed line in Fig. 4). This higher temperature may be due to differing thermal properties of the materials used to construct the instrument.

Outdoor Experiments on Bare Soil Surface

Experiments were conducted on bare soil during a 2-wk period at the end of June and beginning of July 2002, which showed a range of shade temperature. Results for the CP temperature, temperature difference between CP and Rx, and EM-38 response for the five experimental runs are presented in Fig. 5. All the experiments started at 0900 h and finished at 1900 h. The scale in Fig. 5 is relative, showing the time from the beginning of the experiment, when the instrument was switched on, to the end of the experiment. The experiment was designed to mimic a typical data collection day where the instrument had been stored overnight at 20°C, taken outside, and used immediately.

The first four experiments were run with the instrument directly exposed to the sunshine. The EM-38 response for these 4 d is presented in Fig 5C. Day 1 was an extremely hot day with shade temperatures reaching

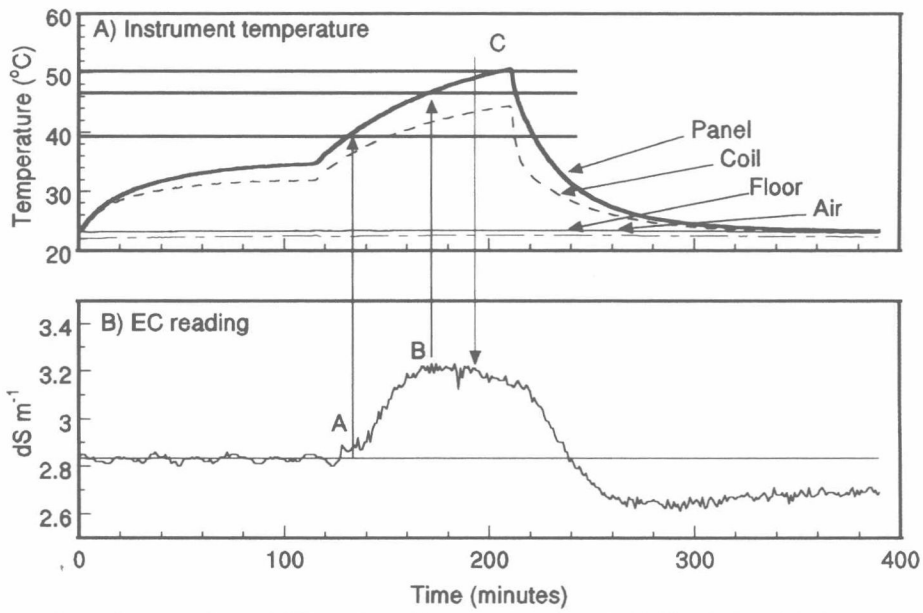


Fig. 4. (A) Temperature response for control panel (CP) and receiving coil (Rx) of the EM-38 with the whole instrument wrapped in an electric blanket and warmed. (B) The soil bulk electrical conductivity (EC_a) response, which demonstrates the temperature compensation up to 140 min and then an unpredictable response as temperature increased above 40°C.

45°C; Days 2 to 4 were progressively cooler. During Days 2 to 4, (i) EC_a data was observed to increase slowly in the first 120 min of the experiment. (ii) A plateau was reached for about 30 min, then (iii) values dropped as temperatures continued to increase. This was similar to the pattern of measured EC_a, increasing, stabilizing and then dropping as the temperature increased observed during the indoor experiments (Fig. 4) The temperatures at which these events occurred (40, 48, and 52°C) were also identical (Fig. 4 and 5). On all days the instrument response showed jumps of increasing EC_a as the instrument began to cool. This occurred between 430 and 550 min, on Days 1 through 4 and showed no consistent pattern.

On Day 1, the temperature of the CP of the instrument at 0900 h was already 40°C. The response of the instrument on this day was displaced with respect to Days 2, 3, and 4 (Fig. 5C). The increase, plateau, and decrease in response observed, when the CP temperature reached 40, 48, and 52°C, occurred at the beginning of the experiment. In fact, by the time we started recording a plateau could be observed which subsequently decreased soon after. This decrease reached a 20% drop below 0.5 dS m^{-1} at the warmest part of the day.

Figure 5B shows the difference in temperature between the instrument CP and the receiver coil during the experiment. In the middle of the day, the temperature of the Rx was 20°C lower than in the CP. It appears that the instrument is temperature compensating the coils for temperatures they are not experiencing. This difference in temperature between the CP, where the temperature sensor and compensation is located, and the receiver coil, for which it compensates, is in part due to the black metallic CP cover. The overestimation of receiver coil temperatures places the instrument out of operational range for most of the day, when it is in direct sun light, resulting in erroneously low values of EC_a.

The last experiment was performed on a day with temperatures equivalent to Day 3. This time the instrument was entirely shaded using a white PVC plastic

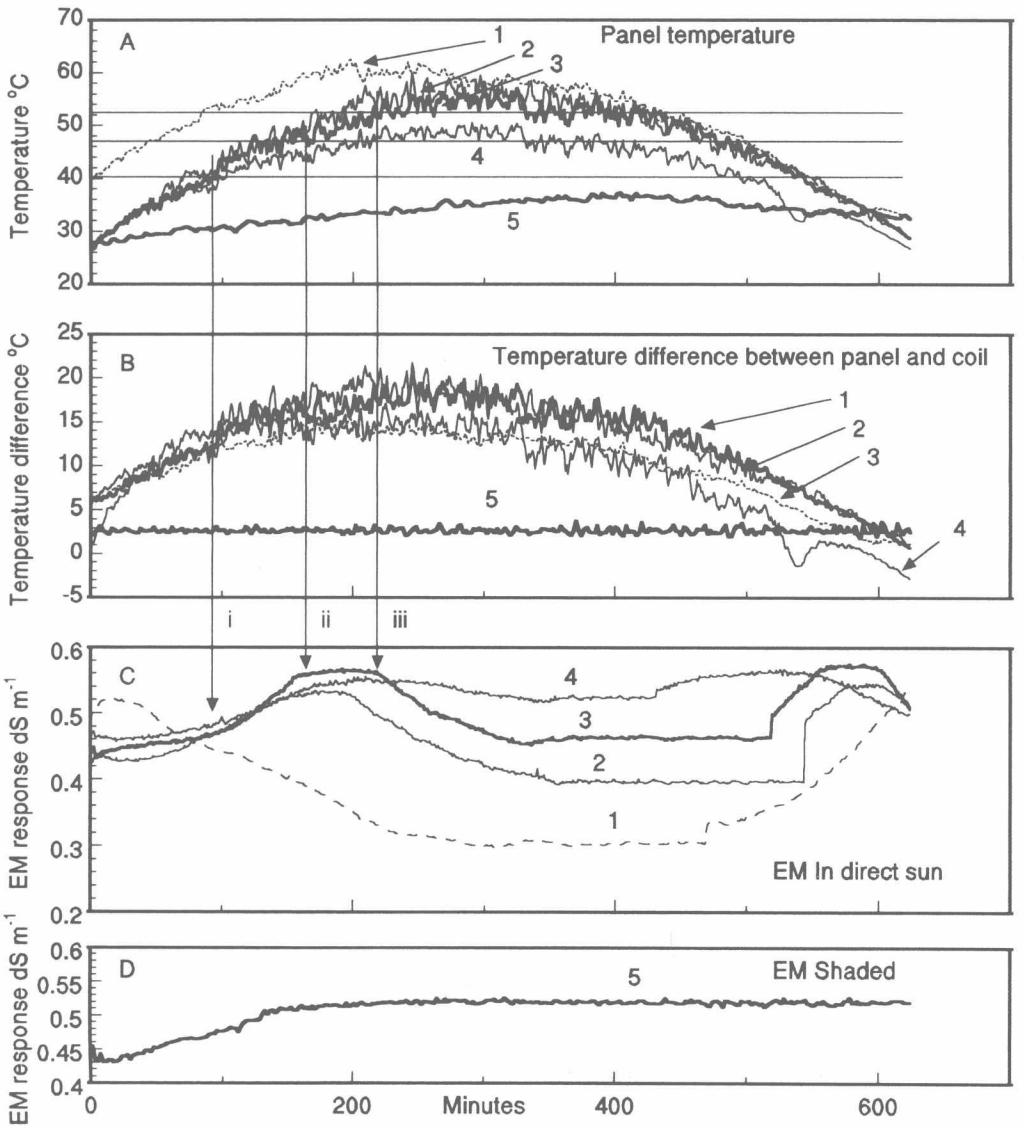


Fig. 5. (A) Panel temperatures (control panel [CP], Fig. 2) for five days. (B) Temperature difference between CP and receiving coil (Rx) for the five days. (C) The instrument responses with increasingly hot panel temperatures. (D) Electromagnetic response with the instrument shaded.

cover. The results for this experimental run became relatively consistent after 120 min, indicating constant soil EC_a, as was expected (Fig. 5D). In the first 120 min of this experiment a 14% increase in measured EC_a was observed coincident with a 2 to 3°C increase in soil temperature (10-cm depth). However, the instrument had been placed in the vertical orientation to avoid response to surface soil heating or water loss. Even if the instrument had responded to a change in soil temperature, it was not sufficient to account for the 14% increase in measured EC_a. This slow increase in instrument response was likely due to a required warm up period. The instrument was stored in the laboratory over night at 20°C and took time to re-equilibrate to the outdoor temperature. We found that the instrument typically required at least 2 h to adjust when the difference between instrument storage temperature and outdoor temperature was 10°C or more. Though not shown these measurements were replicated with a different EM-38 borrowed from colleagues and similar results were observed.

There is always a concern when making measurements on soils that the responses observed are due to changes in water content, or due to changes in the soil temperature and thus EC_a. We have suggested in our argument that the changes in these factors would not be consistent with the observed EM-38 response. However, we conducted a further experiment using the previously described setup but with the instrument located on an impervious asphalt surface. By so doing we could completely rule out a change of water content influencing the results. The heating of the asphalt might be predicted to cause a small if noticeable increase in bulk electrical conductivity measured. Results from this experiment are presented in Fig. 6. During the first 160 min the instrument was covered with a shade and as expected there was a marginal increase in the measured bulk electrical conductivity. The shade was removed at the time denoted by Line A and the panel temperature rose rapidly. The EC_a response showed a small fluctuation rising and reaching a peak value with the panel at 48°C (Fig. 6, Point B). As the temperature continued to rise the EC_a response declined slightly. This decline became steeper once the temperature in the panel reached 52°C. Notice that the temperature under the handle, next to the coil reached 40°C (Fig. 6, Point C). This behavior was identical to the indoor experiment shown in Fig. 3 in which we increased the temperature abruptly in the central part of the EM-38. The signal bottomed out at a value 17% lower than the initial value measured in the morning (Fig. 6, Point D). At Point E the instrument response increased abruptly as temperatures were declining. Although the response increased, it didn't reach an EC_a level similar to the morning until all the temperatures dropped below 40°C (Fig. 6, Point F). The response of the instrument on the asphalt confirmed the instrument sensitivity to heating in direct sunlight.

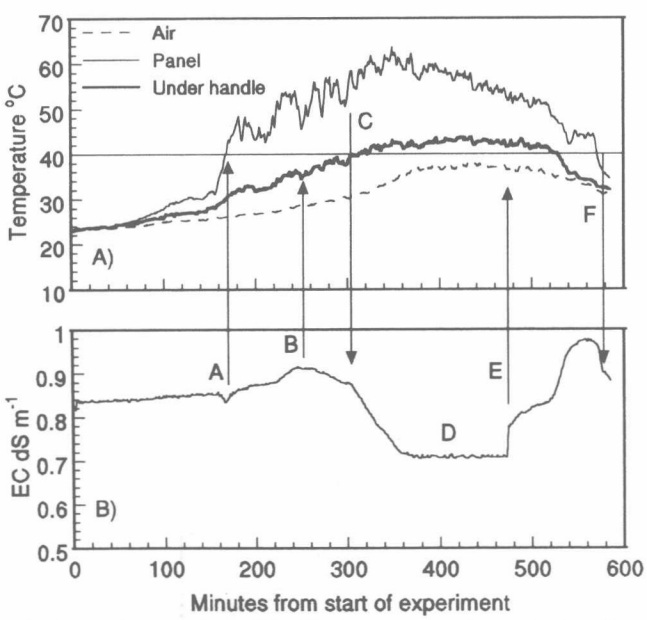


Fig. 6. (A) Temperature of the EM-38 corresponding to the locations under the handle (receiving coil [Rx], Fig. 2) and on the panel (control panel [CP], Fig. 2) of the instrument. (B) The soil bulk density of electrical conductivity (EC_a) response on the asphalt in the vertical orientation without being moved, Points A-E are described in the text.

Drift Observed in Field Mapping Data

Our data indicated that differential heating of the instrument and CP temperatures over 40°C were one of the causes of drift. The cause of this drift comes primarily from the elevated instrument CP temperature (40+°C), or lack of instrument warm up time when taken to a new environment with differing temperature. Our findings appear to be in good agreement with those of Sudduth et al. (2001) who found a strong correlation between temperature increase and EC_a decline for field measurements using a vertically oriented EM-38. We agree with the comments of Sudduth et al. (2001) who suggested that an I/P compensation in the instrument was required. There is only a limited amount that can be achieved by producing field compensations. It appears that the drift is a combination of instrument factors that come down to circuit design, placement of temperature compensation sensors, and coil performance under heating. If the circuit cannot be improved it would be of great use to users in hot areas if information such as I/P and coil temperature could be recorded on the data logger. This means that potentially inaccurate data could be removed from survey data.

In a personal communication with Geonics, we were informed that the effect of temperature on the instrument above 40°C is an absolute value. As an example a 0.05-dS m^{-1} absolute change for a ground conductivity of 0.5 dS m^{-1} is a 10% change. However, at 5 dS m^{-1} this is only a 1% change. The drift highlighted in Fig. 1A can be explained by an insufficient warm-up period of time with the instrument having been taken from an air-conditioned lab or truck before use in the field. This may also explain why some users have trouble calibrating the instrument when arriving at field sites. Calibrating the instrument when it has not been given time to equilibrate with outdoor temperatures and warm-up will give a false calibration, as the instrument subsequently

warms it may go out of calibration. We recommend as for many electronic instruments a 1-h warm up for the electronics (instrument switched on), and time to equilibrate with ambient temperatures before checking instrument calibration (total of 2 h).

The jump observed in the data in Fig. 1B is consistent with jumps in recorded EC_a at around 550 min in Fig. 5C, these jumps happened again toward the end of the day as the instrument cooled down. We suggest that many of these problems might be avoided by shading the instrument as demonstrated in Fig. 5D. Since data collected with clear skies and temperatures above 40°C are likely to cause erroneous measurements, this simple solution of shading the instrument should extend the working range of the EM-38; giving more accurate EC_a responses. Surveys conducted at air temperatures above 40°C are likely to result in erroneous underestimation of EC_a.

CONCLUSIONS

Results presented in this work demonstrate that drift observed in field data collected with the EM-38 is in part due to elevated temperature conditions. When the CP temperature rises above 40°C spurious EC_a measurements occur with EC_a being increasingly underestimated as temperature rises. Results suggest that shading the instrument and keeping its operating temperature below 40°C can substantially improve results. This can reduce the instrument panel temperature by as much as 20°C. This is where the temperature compensation circuitry is located and this can effectively extend the working range of the instrument. We recommend allowing 2 h for instrument warm up before calibration. We suggest that this lack of warm-up time could be one cause of drift often observed when rows are duplicated as a check on readings during a survey. Following these simple steps could improve measurement accuracy by as much as 20% at low conductivity values.

ACKNOWLEDGMENTS

This research was supported in part by a research grant from the USDA-NRI grant program (2002-35107-12507). We would like to especially thank, the staff at Geonics, Cliff Landers (Soil & Water West, Inc.), Jim Rhoades (Agricultural Salinity Consulting), Steve Burch (Imperial Irrigation District), Dennis Corwin and Richard Austin for technical assistance and helpful discussions.

REFERENCES

Belluigi, A. 1948. Coupling of a vertical coil with a homogeneous earth. Geophysics 14:501–507.

Corwin, D.L., and J.D. Rhoades. 1982. An improved technique for determining soil electrical conductivity—Depth relations from above-ground electromagnetic measurements. Soil Sci. Soc. Am. J. 46:517–520.

Corwin, D.L., and S.M. Lesch. 2003. Application of soil electrical conductivity to precision agriculture: Theory, principle, and guidelines. Agron. J. 95:455–471.

Corwin, D.L., S.M. Lesch, P.J. Shouse, R. Soppe, and J.E. Ayars. 2003. Identifying soil properties that influence cotton yield using soil sampling directed by apparent soil electrical conductivity. Agron. J. 95:352–364.

DeJong, E., A.K. Ballantyne, D.R. Cameron, and D.W.L. Read. 1979. Measurement of apparent electrical conductivity of soils by an electromagnetic induction probe to aid salinity surveys. Soil Sci. Soc. Am. J. 43:810–812.

Hendrickx, J.M.B., B. Baerends, Z.I. Raza, M. Sadiq, and M. Akram Chaudhry. 1992. Soil salinity assessment by electromagnetic induction of irrigated land. Soil Sci. Soc. Am. J. 56:1933–1941.

Kachanoski R.G., E.G. Gregorich, and I.J. Van Wesenbeeck 1988. Estimating spatial variations of soil water content using non-contacting electromagnetic inductive methods. Can. J. Soil Sci. 68: 715–722.

Keller, G.V., and F.C. Frischknecht. 1966. Electrical methods in geophysical prospecting. Pergamon Press, New York.

Lesch, S.M., D.J. Strauss, and J.D. Rhoades. 1995a. Spatial prediction of soil salinity using electromagnetic induction techniques: 1. Statistical prediction models: A comparison of multiple linear regression and cokriging. Water Resour. Res. 31:373–386.

Lesch, S.M., D.J. Strauss, and J.D. Rhoades. 1995b. Spatial prediction of soil salinity using electromagnetic induction techniques: 2. An efficient spatial sampling algorithm suitable for multiple linear regression model identification and estimation. Water Resour. Res. 31:387–398.

McNeill, J.D. 1980. Electromagnetic terrain conductivity measurement at low induction numbers. Tech note TN-6, Geonics Ltd., Ontario, Canada. Available online at http://www.geonics.com/ftp-files/geonicstechnicalnotes/tn6/tn6.pdf (verified 7 Oct. 2003).

Rhoades, J.D. 1993. Electrical conductivity methods for measuring and mapping soil salinity. p. 201–251. *In* D. Sparks (ed.) Advances in Agronomy. Vol. 49. Academic Press, San Diego, CA.

Rhoades, J.D., F. Chanduvi, and S.M. Lesch. 1999. Soil salinity assessment: Methods and interpretation of electrical conductivity measurements. FAO report 57. FAO, Rome.

Sheets, K.R., and J.M.H. Hendrickx. 1995. Non-invasive soil water content measurement using electromagnetic induction. Water Resour. Res. 31:2401–2409.

Sudduth, K.A., S.T. Drummond, and N.R. Kitchen. 2001. Accuracy issues in electromagnetic induction sensing of soil electrical conductivity for precision agriculture. Comput. Electron. Agric. 31: 239–264.

Triantafilis, J., G.M. Laslett, and A.B. McBratney. 2000. Calibrating an electromagnetic induction instrument to measure salinity in soil under irrigated cotton. Soil Sci. Soc. Am. J. 64:1009–1017.

Triantafilis, J., M.F. Ahmed, and I.O.A. Odeh. 2002. Application of a mobile electromagnetic sensing system (MESS) to assess cause and management of soil salinization in an irrigated cotton-growing field. Soil Use Manage. 18:330–339.

Wait, J.R. 1954. Mutual coupling of loops lying on the ground. Geophysics 19:290–296.

Wait, J.R. 1955. Mutual electromagnetic coupling of loops over a homogeneous ground. Geophysics 20:630–637.

Wait, J.R. 1982. Geo-electromagnetism. Academic Press, New York.

Wollenhaupt, N.C., J.L. Richardson, J.E. Foss, and E.C. Doll. 1986. A rapid method for estimating weighted soil salinity from apparent soil electrical conductivity measured with an above ground electromagnetic induction meter. Can. J. Soil Sci. 66:315–321.

An Efficient Markov Chain Model for the Simulation of Heterogeneous Soil Structure

Keijan Wu, Naoise Nunan, John W. Crawford,* Iain M. Young, and Karl Ritz

ABSTRACT

The characterization of the soil habitat is of fundamental importance to an understanding of processes associated with sustainable management such as environmental flows, bioavailability, and soil ecology. We describe a method for quantifying and explicitly modeling the heterogeneity of soil using a stochastic approach. The overall aim is to develop a model capable of simultaneously reproducing the spatial statistical properties of both the physical and biological components of soil architecture. A Markov chain Monte Carlo (MCMC) methodology is developed that uses a novel neighborhood and scanning scheme to model the two-dimensional spatial structure of soil, based on direct measurements made from soil thin sections. The model is considerably more efficient and faster to implement than previous approaches, and allows accurate modeling of larger structures than has previously been possible. This increased efficiency also makes it feasible to extend the approach to three dimensions and to simultaneously study the spatial distribution of a greater number of soil components. Examples of two-dimensional structures created by the models are presented and their statistical properties are shown not to differ significantly from those of the original visualizations.

Models of soil physical structure have been developed since the 1950s and used to interpret the impact of structure on function. Childs and Collis-George (1950) introduced the cut-and-rejoin models of soil capillaries, which were modified by Marshall (1958). While many models of soil structure have been developed since then, most relate the structure to physical processes, generally ignoring heterogeneity (for examples see review by Young et al. [2001]), or assume simple pore-size distribution models in an attempt to take some qualitative account of spatial heterogeneity (Young and Ritz, 2000). A number of the more sophisticated approaches exploit the observation that the structure is spatially correlated. For example, Dexter (1976) used a one-dimensional Markov chain model for horizontal soil structure. Moran and McBratney (1997) proposed a two-dimensional fuzzy random model of soil pore structure, which treats the pores as a fuzzy porous set rather than explicitly dealing with geometry. In Vogel (2000) a network model for water retention and permeability is developed where the pore network is geometrically idealized but can be used to predict physical properties from topological parameters determined from thin sections. The geometry of the pore network is explicitly described using a fractal-based approach in Crawford et al. (1995). Finally, Yeong and Torquato (1998) use a combination of the two-point correlation function and the lineal path function to characterize the pore geometry of a broad range of isotropic structures.

The most useful of these models has been used to interpret the impact of structure on physical properties and processes; but comparatively little work has examined the impact on biology. Some attempt has been made to link biological processes with soil structure; these have generally been limited to N transformations (Young and Ritz, 2000). These studies have clearly demonstrated the importance of understanding the relative spatial distribution of the physical and biotic elements of soil structure in determining the larger-scale properties of the resultant biological process (Arah et al., 1997; Rappoldt and Crawford, 1999). Therefore there is a need to further develop models of soil structure that are capable of integrating the physical and biological heterogeneities that occur in most soils. However there are a number of challenges that place constraints on appropriate methodologies, and one of the most significant of these is the limitation of existing imaging technology.

The only reliable methods for visualizing soil in three dimensions are γ and X-ray tomography (Rogasik et al., 1999). While the technology is rapidly improving, it is not a trivial matter to differentiate between pore and solid matrix in these visualizations. Although resolutions of 5 μm or higher are now possible, it is still not possible to directly image soil microbes in situ and in three dimensions over comparable scales. The only method for simultaneously imaging soil structure and the distribution of microbes is by using biological thin sections (e.g., Nunan et al., 2001, 2003). Therefore any model of structure must be capable of being parameterized from two-dimensional data and extrapolated to three dimensions.

The requirements of a useful model capable of describing the heterogeneity of physical and biological elements in soil are three-fold. First, the models must be able to describe the spatial structure of multiphase media (matrix, pore, microbe etc.) at the scale of individual pores and microbes. Second, the model must also accommodate any spatial anisotropy inherent in soil. Finally, the model should be capable of using parameters determined from two-dimensional sections. Currently, no method exists that has been demonstrated to simultaneously satisfy these requirements. Yeong and Torquato (1998) state that their method can be developed to satisfy these constraints, although to date no such modification has appeared. The development of their method based on correlation functions, to multi-

K. Wu and N. Nunan, Soil-Plant Dynamics Unit, Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, UK; J.W. Crawford and I.M. Young, SIMBIOS, University of Abertay Dundee, Kydd Building, Bell Street, Dundee DD1 1HG, Scotland; K. Ritz, National Soil Resources Institute, Cranfield University, Silsoe, Bedfordshire MK45 4DT, UK. Received 4 June 2002 *Corresponding author (j.crawford@simbios.ac.uk).

Published in Soil Sci. Soc. Am. J. 68:346–351 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: MCMC, Markov chain Monte Carlo; MRF, Markov random fields.

phase anisotropic media is far from a trivial extension of the existing methodology.

In this paper, we introduce an efficient method for modeling the architecture of soil, based on two-dimensional images obtained from soil thin sections. The algorithm is based on a Markov chain process that permits the macroscopic (centimeter scale) properties to be determined from local (pore scale) conditions. A Monte Carlo implementation is used to reproduce the stochastic properties of the architecture and to estimate features of the posterior or predictive distribution of interest by using samples drawn from images derived from real soil.

Markov chain Monte Carlo methods usually employ an iterative scheme to obtain the final spatial description (Besag and Green, 1993), and where correlated structures of the order of the image size exist (such as macropores in soil thin sections) the iterative schemes are slow to converge or may even fail. The development of MCMC presented in this paper avoids these problems by a novel choice of neighborhood for the Markov chain and the use of a scanning scheme, based on Elfeki and Dekking (2001) but without the requirement for preconditioning, in which the model image is constructed from a single-pass raster. The resulting structure models are quantitatively compared with the original images to test the validity of the approach. A brief discussion of the applicability of the method to extrapolate to three-dimensions is presented, although a full treatment is deferred to a forthcoming paper. Finally, the trivial extension to simultaneously modeling the biological and physical components of soil architecture in three-dimensions is outlined.

MATERIALS AND METHODS

Methodology

A detailed description of Markov random fields (MRF) can be found in Besag (1974), Geman and Geman (1984) and Cressie (1993). Markov theory lends itself to modeling soil precisely because the structure of soil is spatially correlated. This means that the structural state at any particular point in space is conditionally dependent on the state in the vicinity. Formally, the neighborhood where such dependency prevails is predefined and these dependencies are expressed in the form of conditional probabilities.

In using the MRF method to model visualizations such as those in the current application, a central assumption is that the state of the structure at some point conditionally depends on only a relatively small number of points in a predefined neighborhood. Implementation of the algorithm commences with the derivation of the conditional probabilities from direct measurements of the probabilities of different neighborhood structures in an image of soil structure. The generation of model structures starts with an initial estimate of the spatial distribution of states (e.g., a spatially random distribution). The structure at each point is then updated in accordance with the conditional probabilities derived from the image. This update is then repeated by successive applications of the conditional probabilities at each point, until the statistical properties of the resulting spatial distribution converge (i.e., do not change significantly between successive iterations). Larger-scale correlations emerge as a consequence of the local dependency built into the conditional probabilities, but only after a number of iterations of the algorithm. Indeed a major limitation of the application of the conventional MRF method is the tendency to a convergence slowly and an associated high computational demand. This limits the application to systems where the number of states of each point is low, and where there are only relatively short-range spatial correlations in the structure. Furthermore, the standard implementation of the method cannot reproduce nonisotropic structures, nor can it recognize concavity or convexity of shapes (Tjelmeland and Besag, 1998). Thus, this renders the method inappropriate for modeling most soil structures.

Elfeki and Dekking (2001) employed a raster scheme within a MCMC approach to simulate geological strata. Their model was parameterized from data collected from a set of wells, and they assumed that the state of a particular point in the strata was dependent on the points immediately to one side and to the top. Therefore if the state of the points are known along a transect at the surface and in a vertical direction at a single well location, it is possible to use the model to predict the state of the remaining points in the transect-well plane. The model calculates the state of a particular point on the basis of the state of the point to the left and above, and the conditional probabilities associated with the Markov process. This is repeated in a step-wise way as the chain moves from left to right across the domain in a raster fashion. Crucially, however, to get an accurate representation the chain must be conditioned on several wells across the transect. Therefore the probabilities are adjusted to guarantee agreement with the well data (Elfeki and Dekking, 2001). This approach is appropriate for data on geological strata since the number of wells that can be drilled always limits one, and the aim is to reproduce the actual structure as closely as possible. The purposes are quite different to those of the algorithm presented here. In our case, we have complete information about the state of the points in a two-dimensional domain (i.e., soil thin section), and we aim to reproduce the functionally important statistical properties of the structure rather than a literal copy of the structure itself. The method should also be extendable to three dimensions, and for the reasons outlined above, we are restricted to methods that can extrapolate from two-dimensional data. The method of Elfeki and Dekking (2001) is then inappropriate.

Multidimensional Markov Chain Model

We consider an image made up of pixels arranged in a rectangular array. The standard implementation of the MRF method assumes that the state of a particular point in an image depends on the state of an isotropic neighborhood centered about the point. As this is unsuitable for modeling anisotropic soil structures, we proceed by removing this constraint.

The Potential Function

The standard MRF method calculates the required conditional probability from a potential function defined in an isotropic neighborhood. The fundamental framework contains the following components:

(i) a set $S = \{x | x \in S\}$ of (pixel) sites;
(ii) a set $N = \{n_x | n \in N\}$ attached to the site;
(iii) a probability model, for the joint distribution of the S.

$$p(x) \propto \exp\left[\sum_{x \in S} G_x(n_x)\right] \qquad [1]$$

Here we redefine the neighborhood as one based on five pixels

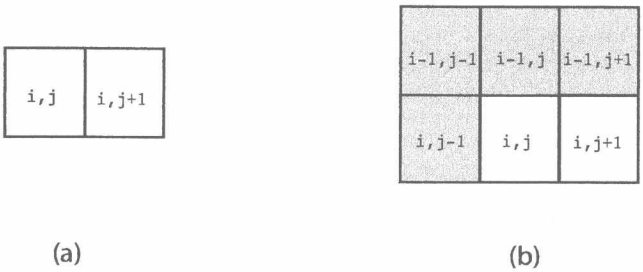


Fig. 1. Notation relating to neighborhood system adopted in this study. (a) Two-pixel neighborhood; (b) five-pixel neighborhood.

forming an anisotropic neighborhood as illustrated in Fig. 1. Thus, the state of any given pixel is assumed to depend on the two pixels immediately to its left, and on the three immediately above it.

The conditional probabilities are then determined from the relation:

$$P(X_i = x_i|x_{i-1},...,x_1) = p(x_i|x_{i-1})$$
$$\propto \exp\left\{\sum_{j=1}^{N} g_{ij}(x_{ij}) + \sum_{j=1}^{N-1} [G_{ij}(x_{ij}, x_{i,j-1}) + G_{ij}(x_{ij}, x_{i,j+1})]\right\} \quad [2]$$

where N is the number of neighbors, g_{ij} and G_{ij} are the potential functions, and x_{ij} represents the state of the pixels at positions (i,j). Because of the difficulty in parameterizing the potential function, these models are usually generalized to involve an arbitrary structure of pairwise pixel interactions (Gimel'farb, 1999). There have only been a few attempts to broaden the class of these models by introducing potential functions with more complex neighborhoods, including the Gibbs model (Moussouris, 1974) and region maps (Derin and Elliot, 1987). The associated potentials for the formulae have only two values: $G_x(n_x) = \theta_x$, if all states x in the neighborhood n_x are equal, and $-\theta_x$ otherwise. For higher-order interactions, the potential functions are assumed to be a linear function of parameters, which can be derived from an image only under such simplified assumptions.

The advantage of determining the probabilities from the potential function is that the neighborhood interactions of an entire image can be represented by relatively few parameters. However, the apparent compactness is a consequence of the assumption that the potential functions can be expressed as a linear function of the parameters. Because of this, the parameters can be estimated using maximum-likelihood methods (Qian and Titterington, 1991). This appears to work for many kinds of images, but its applicability to images of soil, where relatively long-range correlations exist, has not been verified. We attempted to model the structure of our soil samples by calculating the potential function. However, the resulting model structures were inadequate. For moderate-sized images, the modeled structure showed substantial departures from the original (parent) structure, and as the size of the modeled domain increased, the agreement deteriorated further (Fig. 2). Thus, the simplifying assumptions underlying the linear formulation of the potential functions are incompatible with the image structures associated with soil. We therefore developed an alternative method based on direct measurement of the 64 conditional probabilities associated with all possible configurations of the five-pixel neighborhood (Fig. 1).

Markov Chain Monte Carlo Method

While the definition of the neighborhood as defined in Fig. 1 provides the potential to treat anisotropy, this change alone is clearly insufficient, and an alternative to the determination of conditional probabilities from the potential functions must be found. To this end, we replace the potential functions with the full set of conditional probabilities that define all possible combinations of states for the neighborhood. In the case studied here, we aim to model the relative position of pore space and solid, and so each pixel can be in one of two possible states.

The standard methods for implementing MCMC are iterative, for example, the Gibbs sampler version of the Metropolis Hastings algorithm (Geman and Geman, 1984). However, these suffer from long convergence times as discussed above, and are not suitable for modeling large correlated structures such as are found in soil. Here, we use a more efficient method based on the scanning scheme algorithm proposed by Qian and Titterington (1991), modified to cope with long-range correlations in the structural heterogeneity found in soil images. The modification replaces the potential function with a more explicit determination of the transition probabilities as detailed below.

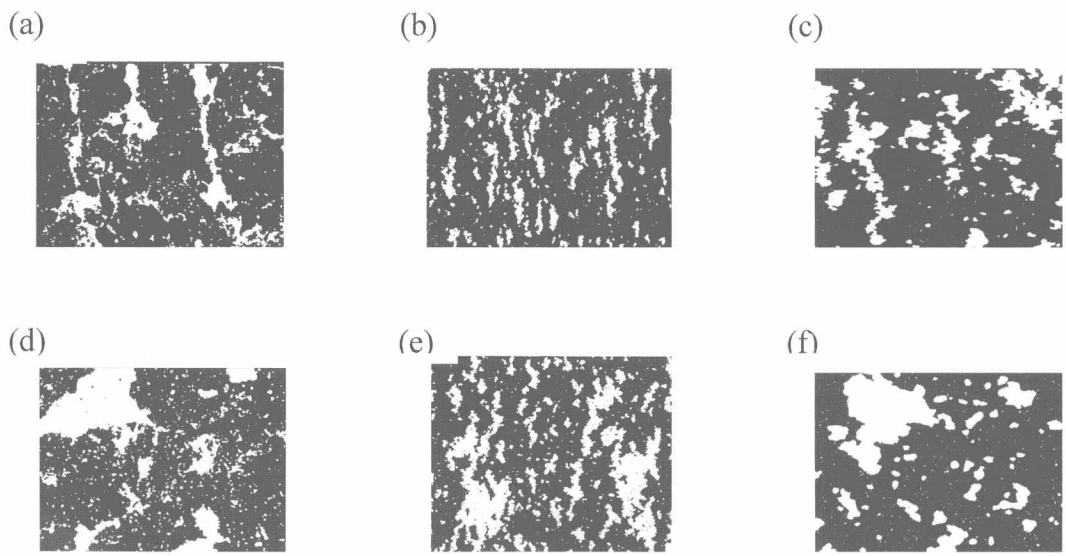


Fig. 2. Comparison of simulated soil images with real images. The size of each image is 1.6 by 1.2 cm. (a) Original soil image with anisotropic, linear pore structure. (b) Simulated image using the potential function; (c) simulated image using the scan scheme. (d) Original soil image containing pores of the order of the image size; (e) simulated image using the potential function; (f) simulated image using the scan scheme.

The Scanning Scheme Algorithm

After considering several different forms for the neighborhood, we determined that the neighborhood given in Fig. 1 was the smallest capable of reproducing the observed soil properties. The modeling proceeds in two steps. First, the state of the pixel located at the point (i,j) is determined from knowledge of those at $(i,j-1)$, $(i-1,j-1)$, $(i-1,j)$, and $(i-1,j+1)$ using the associated four-neighborhood conditional probability. Second, the state of the pixel at $(i,j+1)$ is obtained from knowledge of the new state at (i,j) together with the state of the those at $(i,j-1)$, $(i-1,j-1)$, $(i-1,j)$, and $(i-1,j+1)$ using the associated five-neighborhood conditional probability. These probabilities are obtained from the original image by sampling the four- and five-neighborhoods, and enumerating the different realizations of the state of the point (i,j) and $(i,j+1)$ respectively for each configuration of the neighborhood.

To initiate the model we need to assign states to all the cells in the first row, and the first cell of the next row. The cells in the first row are obtained using a two-neighborhood Markov chain where the state of a cell is conditionally dependent on the state of the cell to its left. The parameters for this Markov chain are obtained as above, but using the smaller two-cell neighborhood. The state of the first cell of the next row is determined using the same two-neighborhood conditional probability. Using these boundary values, the chain runs from the left-hand corner of the image and progresses in raster fashion across the image to the right-hand side. First, the state of cell (i,j) in the neighborhood is evaluated, followed by the state of cell $(i,j+1)$. The neighborhood is then advanced two cells to the left and the process is repeated. This continues until the last cell on the right-hand side is reached and its state is evaluated. At this point, the cells in the first two rows have been evaluated. Next, the neighborhood remains at the right-hand side but moves one row down. The chain now reverses direction, and instead of deriving the states of the (i,j) and $(i,j+1)$ cells in terms of the state of the others, it is the states of cells (i,j) and $(i,j-1)$ that are determined. However, before the chain can proceed leftwards, the state of the first cell on the right-hand side of the third row must be evaluated. This is done in the same way as when the neighborhood was at the left-hand side of the domain, using the two-neighborhood conditional probability. The chain can now advance leftwards until the left-hand side of the domain is reached. The neighborhood then moves down one row, and the chain reverses as before. Thus, the whole domain is scanned in a raster-like fashion on this basis, until the states of all the required cells are obtained.

The scanning scheme algorithm converges rapidly. In the examples reported here, we observed the transition kernel (i.e., the matrix of conditional probabilities for all possible neighborhood configurations), calculated from the reconstructed image as the chain progressed. Almost all the probabilities had become stable after the chain had completed 200 rows, which is equivalent to a depth of 4.0 mm in the original soil section and takes about 10 s of computing time on a 1.7-GHz Pentium IV computer. In other words, the minimum size of a simulated representative soil image should be 0.6 by 0.4 cm^2, and it takes only a few minutes to generate an image covering several squared.

Validation

The method was validated using images obtained from soil thin sections and selected to represent a broad contrast in structural properties.

Sampling

Soil cores were collected from an arable field and thin sections were produced using the method described in Nunan et al. (2001). Soil pore maps were obtained by subtracting images obtained with cross-polarized light from images captured using transmitted bright-field light. The resultant images were then segmented into solid and void. The images employed in this study were binary pore maps of dimension 760 by 570 pixels, representing an area of 1.6 by 1.2 cm. Images were selected to represent a range of characteristic soil structural properties, as shown below.

Comparison of Real and Simulated Images

To compare the simulated and real images we selected a range of quantitative metrics that characterize the heterogeneity and connectivity of the structures under investigation. The most obvious of these is the porosity and this is readily determined from both the real and simulated structures. The corresponding values are listed in Table 1 and range from 7 to 24% in the real structures. There was no significant difference between the porosities in the simulated and real structures ($P > 0.05$, paired t test).

The second metric adopted was the mass fractal dimension, which essentially characterizes the degree of aggregation of the solid matrix. The mass (solid) fractal dimension was determined by the box counting method (Hastings and Sugihara, 1993). The calculated values are listed in Table 1, and again there was no significant difference between the simulated and real structures ($P > 0.05$, paired t test).

The third and fourth metrics chosen characterize the pore space, where visually obvious differences between the samples were present. The third is the variance of the porosity as measured in a 0.4 by 1.6 mm sampling window placed at 50 random locations in each image. For a given porosity, this is a measure of the connectivity of the pore space (Mandelbrot, 1985). We used the Chi-Square goodness-of-fit test, and tested the null hypothesis that the variances in each pair of simulated and original images were different. This could be rejected at the 95% confidence level indicating that this aspect of the structure of the pore space was not significantly different in the real and simulated images. The fourth metric measured the spatial correlation of the pore space by determining the semivariogram. Figure 3 shows the variograms for the different soil samples used in this study, and there are clear differences in these between the different soil sections. The figure shows the comparison between the variograms for the real and simulated structures. There is no formal statistical way of comparing the properties of semivariograms, however the high degree of correspondence between the curves for the measured and simulated structures is good, adding further support for the modeling methodology.

Table 1. Comparison of soil properties between real image and simulated image

		Real soil image	Simulated image
Sample 1	Porosity, %	7	7
	Mass fractal dimension, D_m	1.9642	1.9654
Sample 2	Porosity, %	17	16
	Mass fractal dimension, D_m	1.9382	1.9322
Sample 3	Porosity, %	24	22
	Mass fractal dimension, D_m	1.9204	1.9366
Sample 4	Porosity, %	12	11
	Mass fractal dimension, D_m	1.9599	1.9503

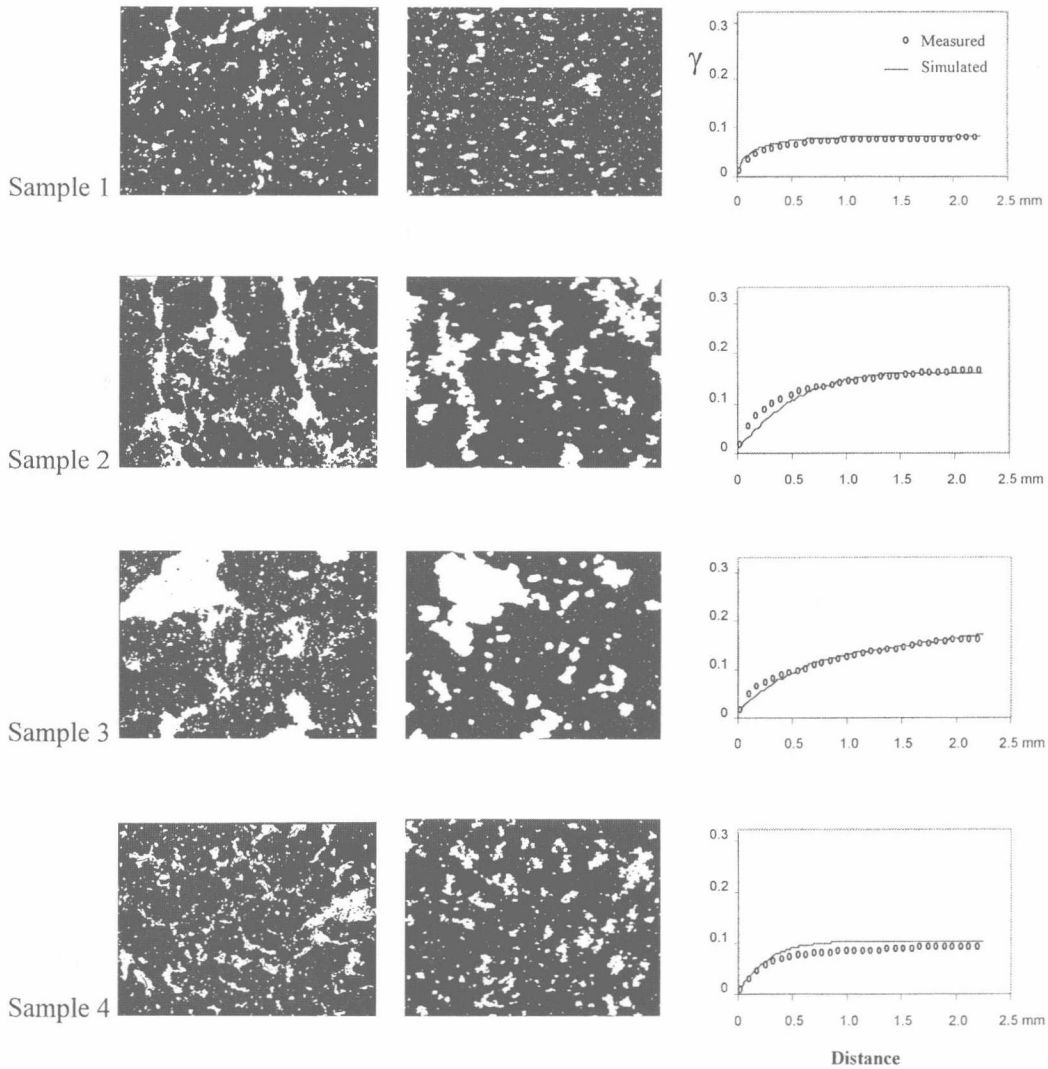


Fig. 3. Some examples of soil thin section images and associated simulated images (each image size is 1.6 by 1.2 cm) (a) original images (i.e., soil thin sections); (b) associated simulated image; (c) semivariograms of pore space in original and simulated images.

CONCLUSIONS AND DISCUSSION

We describe a new method for modeling the complex architecture of soil. The method is based on a MCMC approach, but incorporates a novel neighborhood definition and scanning scheme to model large-scale spatial correlations with rapid convergence. In this approach the state of a pixel in the model image is conditionally dependent on the state of the pixels in the predefined neighborhood. The associated conditional probabilities are calculated directly from a segmented image obtained from a thin section of soil. The method reproduces the mean and spatial variance of the porosity, and fractal dimension of the matrix as well as the spatial variogram of the pore space, as estimated from original thin section images. Agreement was obtained with images of widely contrasting soil structures. As well as being capable of modeling broad-scale heterogeneity, the method is also more efficient and therefore faster than previous approaches. As a consequence, we can simulate larger structures than has previously been possible and so link pore-scale to core-scale. By linking with suitable models for physical processes, it should now be possible to search for scaling laws from first principals that relate the impact of microscopic detail on macroscopic behavior.

The method is easily extendable to simultaneously model the spatial distribution of a variety of components of soil architecture. In the current paper, we verified the approach by modeling the physical elements hence each pixel could be in one of two states—pore or solid. In a forthcoming publication, we have extended the approach to model the relative spatial distribution of microbes in soil, parameterized directly from soil thin sections that have been prepared in a manner that preserves the microbes in situ (Nunan et al., 2003).

While it is of interest to be able to produce two-dimensional models of the different components of soil architecture that have the same statistical properties as

real soil, the most important advantage of the current approach is that is has the ability to extrapolate to three dimensions. The efficiency of the algorithm together with a Markov process that is based on local neighborhood, without the need to condition the chain on existing data, makes it possible to consider modeling large three-dimensional structures. This means it is possible to produce models for the three-dimensional distribution of microbes in structured soil for the first time. This can be combined with efficient algorithms for modeling unsaturated flow through porous media (e.g., Zhang et al., 2002) and the distribution of O_2 (Rappoldt and Crawford, 1999) to begin to understand how the physical and biological processes in soil interact. Again, the ability to relate pore- to core-scale means we can approach the issue of possible scaling laws from first principals. By linking microbes and their microhabitats directly in this way, the potential for theoretical and experimental soil ecology is significantly advanced.

ACKNOWLEDGMENTS

This work was part-funded by a UK Department of Trade and Industry Award in the Biological Treatment of Soil and Water LINK Programme, and is carried out in association with Aventis and QuantiSci. The Scottish Crop Research Institute receives grant-in-aid from the Scottish Executive Environment and Rural Affairs Department.

REFERENCES

Arah J.R.M., J.H.M. Thornley, P.R., Poulton, and D.D. Richter. 1997. Simulating trends in soil organic carbon in long-term experiments using the ITE (Edinburgh) Forest and Hurley Pasture ecosystem models. Geoderma. 81:61–74.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). J. Royal Stat. Soc. Ser. B 36:192–236.

Besag, J., and P.J. Green. 1993. Spatial statistics and Bayesian computation. J. Royal Stat. Soc. Ser. B 55:25–37.

Childs, E.C., and N., Collis-George. 1950. The permeability of porous materials. Proc. R. Soc. London, Ser. A 201:392–405.

Crawford, J.W., N. Matsui, and I.M. Young. 1995. The relation between the moisture-release curve and the structure of soil. Eur. J. Soil Sci. 46:369–375.

Cressie, N.C. 1993. Statistics for Spatial Data, rev. ed. Wiley-Interscience, Hoboken, NJ.

Derin, H., and H. Elliot. 1987. Modelling and segmentation of noisy and textured images using Gibbs random fields. IEEE Trans. Anal. Machine Intell. PAMI-9:39–55.

Dexter, A.R. 1976. Internal structure of tilled soil. J. Soil Sci. 27: 267–278.

Elfeki, A., and M. Dekking. 2001. A Markov chain model for subsurface characterization: Theory and applications. Math. Geol. 33: 569–589.

Geman, S., and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE T. Pattern Anal. Machine Intell. PAMI 6:721–741.

Gimel'farb, G.L. 1999. Image textures and Gibbs random field. Kluwer Academic Publishers, Dordrecht.

Hastings, H.M., and G. Sugihara. 1993. Fractals: A user's guide for the natural sciences. Oxford University Press, Oxford.

Mandelbrot, B.B. 1985. The fractal geometry of nature. W.H. Freeman & Co., New York.

Marshall, T.J. 1958. A relation between permeability and size distribution of pores. J. Soil Sci. 9:1–8.

Moran, C.J., and A.B. McBratney. 1997. A two-dimensional fuzzy random model of soil pore structure. Math. Geology 29:755–777.

Moussouris, J. 1974. Gibbs and Markov random systems with constraints. J. Stat. Phys. 10:11–33.

Nunan, N., K. Wu, K. Ritz, I., Young, and J.W. Crawford. 2001. Quantification of the in situ distribution of soil bacteria by large-scale imaging of thin-sections of undisturbed soil. FEMS Microbiol. Ecol. 36: 67–77

Nunan, N., K. Wu, I., Young, J.W. Crawford, and K. Ritz, 2003. Spatial distribution of bacterial communities and their relationships with the micro-architecture of soil. FEMS Microbiol. Ecol. 44: 203–215

Qian, W., and D.M. Titterington. 1991. Multidimensional Markov-chain models for image-textures. J. Roy. Stat. Soc. Ser. B 53: 661–674.

Rappoldt, C., and J.W. Crawford. 1999. The distribution of anoxic volume in a fractal model of soil. Geoderma 88:329–347.

Rogasik, H., J.W. Crawford, O. Wendroth, I.M. Young, M. Joschko, and K. Ritz. 1999. Discrimination of soil phases by dual energy x-ray tomography. Soil Sci. Soc. Am. J. 63:741–751.

Tjelmeland, H., and J. Besag. 1998. Markov random fields with higher-order interactions. Scand. J. Stat. 25:415–433.

Vogel, H.J. 2000. A numerical experiment on pore size, pore connectivity, water retention, permeability, and solute transport using network models. Eur. J. Soil Sci. 51:99–105.

Yeong, C.L.Y., and S. Torquato. 1998. Reconstructing random media. Phys. Rev. E. 57:495–506.

Young, I.M., and K. Ritz. 2000. Tillage, habitat space and function of soil microbes. Soil Tillage Res. 53:201–213.

Young, I.M., J.W. Crawford, and C. Rappoldt. 2001. New methods and models for characterizing structural heterogeneity of soil. Soil Tillage Res. 61:1–13.

Zhang, X.X., A.G. Bengough, J.W. Crawford, and I.M. Young. 2002. A lattice BGK model for advection and anisotropic dispersion equation. Adv. Water Resour. 25:1–8.

Millimeter-Scale Spatial Variability in Soil Water Sorptivity: Scale, Surface Elevation, and Subcritical Repellency Effects

P. D. Hallett,* N. Nunan, J. T. Douglas, and I. M. Young

ABSTRACT

Recent evidence suggests that reduced water infiltration may be linked to small scale microbial and/or chemical processes that cause subcritical water repellency. We measured water sorptivity on the surface of a large intact block of soil (0.9 m wide, 1.3 m long, 0.25 m deep) taken from a grassland site and examined the effects of surface elevation and water repellency on water sorptivity at the millimeter scale. The soil block was partially dried to 0.22 $mm^3\ mm^{-3}$, appeared to wet readily, and is not severely water repellent at any water content. Water sorptivity varied from 0.1 to 0.8 mm $s^{-1/2}$ across the sampling grid with a coefficient of variation (CV) of 0.57. Water repellency, determined by comparing water and ethanol sorptivities, also varied considerably (CV = 0.47). Geostatistical analyses of water sorptivity and repellency measurements found little evidence of spatial autocorrelation, suggesting a high degree of local variability. These data were compared to larger scale measurements obtained with conventional infiltrometers under tension conditions (40 mm contact radius), and ponded conditions (37 and 55 mm radius rings) where macropores influence infiltration heterogeneity. Larger scale tension infiltrometer measurements were less variable with a CV of 0.22, whereas ponded infiltrometer measurements were more variable, CV > 0.50, presumably because of the influence of macropore flow. Data collected on surface elevation showed that ponded infiltration but not tension infiltration was influenced by surface topography. The results suggested that repellency can induce levels of spatial variability in water transport at small scales comparable to what macropores induce at larger scales.

Low levels of water repellency have been observed in many soils (Hallett et al., 2001; Wallis et al., 1991; Tillman et al., 1989). Although water appears to readily infiltrate these soils, it has been postulated that the slight, yet significant, reduction in infiltration rates through repellency can cause an increase in soil aggregate stability, and in the heterogeneity of overland flow and water infiltration at the field scale (Hallett and Young, 1999). Most studies on the heterogeneity of water infiltration and overland flow have concentrated on the influence of macropores and other dominant soil pore structure features serving as preferential flow pathways (Smettem, 1987; Heuvelmann and McInnes, 1997). We hypothesize that low levels of repellency will also exhibit high levels of spatial variability across the surface of soil, and that this will add additional heterogeneity, particularly under conditions of tension infiltration when macropores are less important.

The concept of low level or subcritical water repellency is not new. Soil physicists are taught the importance of soil–water contact angles early on in their undergraduate syllabus and Philip (1957) recognized the importance of repellency in his original work on sorptivity, but despite this knowledge, it is widely ignored in current research as soil is assumed to be completely nonrepellent. Tillman et al. (1989) developed a simple technique for quantifying repellency and with these data suggested that most soils exhibit subcritical water repellency where despite the soil appearing to uptake water readily, partially hydrophobic soil particle surfaces impede the rate of infiltration. Hallett and Young (1999) combined Tillman et al.'s (1989) approach with a miniaturized infiltrometer developed by Leeds-Harrison et al. (1994) to allow for water repellency to be measured on individual soil aggregates at millimeter resolution. Subsequent work using this new technique showed that repellency had a biological origin controlled by organism type (White et al., 2000), nutrient levels (Hallett and Young, 1999) and exudate chemistry (Czarnes et al., 2000).

The biological origin of repellency suggests that it will have a high spatial and temporal variability at very small scales, because of the submillimeter spatial variability of organic matter, organisms and the microbial environment in soil (Nunan et al., 2002). Using the miniaturized infiltrometer, we measured water sorptivity on the surface of a large intact block of soil to determine its spatial heterogeneity at the microscale and the effect of surface elevation and subcritical water repellency. These data were compared with larger scale measurements obtained with conventional infiltrometers (Logsdon and Jaynes, 1996; Shouse et al., 1994) under ponded conditions where macropores influence infiltration heterogeneity, and under tension conditions where heterogeneity would be expected to be less severe because measurements were above a size threshold where repellency variability is detectable. Data were also collected on surface topography since depressional storage may affect measurements and can operate over a range of scales, perhaps smaller than the size of the infiltrometer (Kamphorst et al., 2000). Geostatistics were applied to measure spatial variability and to detect any potential spatial pattern and dependency in water infiltration at the different spatial resolutions examined.

This work is highly relevant to describing physical and biological phenomena that may impart heterogeneity to the overland flow and infiltration of water in soil at

P.D. Hallett and N. Nunan, Plant Soil Interface Programme, Scottish Crop Research Institute, Dundee, DD2 5DA, Scotland; J.T. Douglas, Environment Division, Scottish Agricultural College, Bush Estate, Penicuik, EH26 0PH, Scotland; I.M. Young, Scottish Informatics Mathematics Biology & Statistics SIMBIOS (Centre), University of Abertay Dundee, Bell Street, DD1 1HG, Scotland. This research was partly funded by grant-in-aid support from the Scottish Executive Environment and Rural Affairs Department. Received 24 Mar. 2003. *Corresponding author (p.hallett@scri.sari.ac.uk).

Published in Soil Sci. Soc. Am. J. 68:352–358 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: CV, coefficient of variation.