JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2019, Vol. 64, No. 5
doi: 10.1111/1556-4029.14060
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Patrice de Caritat,[1,2,†] *Ph.D.; Timothy Simpson*,[1] *B.Sc. (Hons); and Brenda Woods*,[1,2] *Ph.D.*

Predictive Soil Provenancing (PSP): An Innovative Forensic Soil Provenance Analysis Tool*

ABSTRACT: Soil is a common evidence type used in forensic and intelligence operations. Where soil composition databases are lacking or inadequate, we propose to use publicly available soil attribute rasters to reduce forensic search areas. Soil attribute rasters, which have recently become widely available at high spatial resolutions, typically three arc-seconds (~90 m), are predictive models of the distribution of soil properties (with confidence limits) derived from data mining the inter-relationships between these properties and several environmental covariates. Each soil attribute raster is searched for pixels that satisfy the compositional conditions of the evidentiary soil sample (target value ± confidence limits). We show through an example that the search area for an evidentiary soil sample can be reduced to <10% of the original investigation area. This Predictive Soil Provenancing (PSP) approach is a transparent, reproducible, and objective method of efficiently and effectively reducing the likely provenance area of forensic soil samples.

KEYWORDS: forensic science, soil forensics, soil attribute grid, Geographic Information System, raster operation, search area

Forensic geology can be described as the use of geological methods and materials in the analysis of samples and places that may be connected with criminal behavior or disasters (1). In this definition, geological methods encompass geophysics, geochemistry, mineralogy, petrography, microscopy, and micropalaeontology, while geological materials refer to soil, sediment, and rock. Biological methods, such as palynology and DNA analysis for instance, also contribute significantly to the success of soil forensics (e.g., [2]). The use of geological material, such as soil, in forensic investigations is increasing in police forces around the world, including at the Federal Bureau of Investigation (FBI), the Royal Canadian Mounted Police (RCMP) and the Australian Federal Police (AFP), as well as in other forensic agencies (e.g., [3-11]). Forensic soil provenancing, a sub-discipline within forensic geology, can be defined as the capability to spatially constrain the likely region of origin of an evidentiary sample of earth-related material (12,13). Rawlins et al. (12) characterized the prediction of the provenance of a sample of earth-related material as "one of the most difficult and challenging tasks for analytical earth scientists." After briefly reviewing the current approach to soil provenancing, we develop an alternative approach that may circumvent some of these challenges.

Empirical Soil Provenancing

Typically, forensic soil provenancing is implemented in an empirical way, in which the multidimensional information contained in the evidentiary soil's geochemistry, mineralogy (including grain morphology and mineral chemistry), biology, bulk properties, etc., is compared to either purposely acquired or pre-existing knowledge (7,12–15). We call this "Empirical Soil Provenancing" (ESP) in Fig. 1. Such knowledge typically is derived from soil geochemical surveys and stored in databases containing this same or similar multidimensional information over the region of interest at an appropriate density (14).

Geochemical surveys come in many guises (e.g., [16,17]) and although many already exist at a range of spatial scales (continental to local), sampling densities (1 sample per 1000's of km^2 to 100's of samples per km^2), and sampling media (materials) selections (topsoil, C horizon, sediment, ...), forensic applications have specification requirements that may not have been the primary focus of the original surveys (18). Despite that fact these pre-existing surveys and associated databases have their use in forensic applications, as long as their limitations are understood.

Once a database is selected, a number of statistical and visualization analysis tools are typically implemented, including univariate, bivariate and multivariate statistical analysis, exploratory data analysis, analysis of variance, compositional data analysis, spatial interpolation and smoothing, cluster analysis, supervised or unsupervised classification, and data mining (e.g., [19-29]).

The next step in the ESP workflow is the comparison of the evidentiary soil sample's composition with the selected database.

---

[1] Australian Federal Police, GPO Box 401, Canberra, ACT 2601, Australia.
[2] National Centre for Forensic Studies, University of Canberra, Bruce, ACT 2617, Australia.
[†] Present address:Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia.
Corresponding author: Patrice de Caritat, Ph.D. E-mail: Patrice.deCaritat@ga.gov.au

*Presented in part at the 2018 International Symposium of the Australian and New Zealand Forensic Science Society (ANZFSS), September 11, 2018, in Perth, Australia.

Received 28 Jan. 2019; and in revised form 26 Mar. 2019; accepted 27 Mar. 2019.

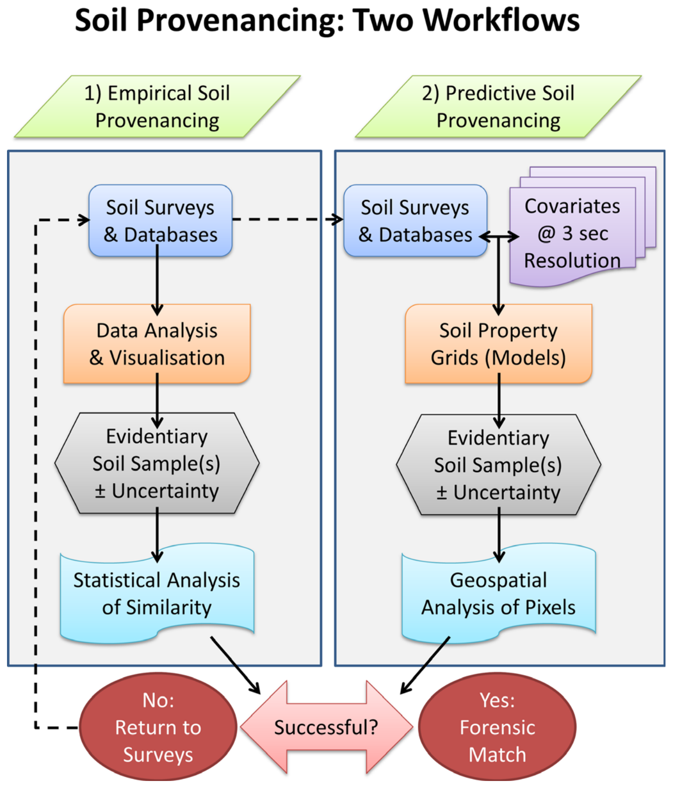


FIG. 1—*Empirical and predictive soil provenancing workflows developed in this study. [Color figure can be viewed at wileyonlinelibrary.com]*

This analysis of similarity can be done using a few or many compositional characteristics (including chemical element abundances, isotopes, mineral abundances), their ratios or other calculated indexes, correlation analysis, and/or factor or principal component analysis (e.g., [30-36]), among others.

Finally, the evidentiary sample is ascribed to a particular region of origin and conventional forensic investigation can proceed there. If unsuccessful or inconclusive (e.g., too large an area), more data and better data may need to be collated (if pre-existing) or collected (if not), which may imply undertaking a more refined soil survey, or rely on other, independent information.

Information from geochemical surveys, however, may not always be available at a suitable density and/or quality over a region of (potential) investigation, though it may be available in some cases and certainly can be obtained with appropriate resourcing and prioritization (37). Designing and implementing a geochemical survey can take time and commitment, commensurate with the size of the investigation area and required sampling density. Thus, if soil provenancing is to rely solely on this empirical approach, it will likely be of limited applicability in the short term, unless a crime is committed in an area covered by a suitable geochemical survey. This should not prevent, in the medium term, the continued effort of developing suitable empirical databases (e.g., [38,39]), which ultimately deliver "ground-truthed" results.

Predictive Soil Provenancing

If empirical, quantitative databases of desirable soil properties are not available at a suitable spatial resolution and analytical quality over many/all potential regions of interest, an alternative approach dubbed "Predictive Soil Provenancing" (PSP) is presented here (Fig. 1). In this approach, we propose that use be made of rasters or grids released in the last decade or so by the digital soil mapping community (e.g., [40,41]). The soil grids are models of the geographic distribution of selected soil and landscape attributes generated by high-dimensional correlation analysis with several environmental variables, or "covariates," available at high spatial resolutions. The preparation of the soil rasters is heavily reliant on data mining and advanced analytics.

To implement PSP, several soil grids are combined or intersected to identify a subset of pixels (or cells) that satisfy the set of conditions reflecting the composition of the evidentiary soil sample, i.e., the target value and its confidence limits, which define the search range ([target value − confidence limit, target value + confidence limit]). This step is carried out using a Geographic Information System (GIS), an analytical tool poised to gain increasing practical applicability in forensic geoscience (18,42). Thus, a region that is more likely to be the source of an evidentiary soil sample with given target values (and search ranges) can be identified and prioritised for further forensic investigation. PSP can be regarded as a fully quantitative implementation of the concept of "predictive geolocation" of Pirrie et al. (43) using the mathematical and spatial rigor of GIS to narrow down a search area. Other soil forensic evidence, such as pollen, presence of anthropogenic constituents or DNA, if available, should of course be considered too, and this could further narrow down the search area. PSP is a rapid and objective area-reduction step that could precede ESP using conventional soil forensic tools.

A few recent publications have proposed quantitative spatial methods that have similarities with the PSP method proposed here (44,45), highlighting this as an area of active research and ongoing development in soil forensics.

Materials and Methods

Recently, digital soil grids have become available for whole countries or even continents, such as in Europe (46), the United States (47), Africa (48), or indeed as global compilations (49). In Australia, national grids of soil attributes have been released by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and partner organizations (50) within the Terrestrial Ecosystem Research Network (TERN), which is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS). The innovative idea developed in the present paper is that in the absence of suitable empirical databases of soil properties on which to base soil provenancing, these grids offer a possible alternative. This approach may provide a transparent, reproducible, and objective framework for guiding investigational priorities, for example, searching areas for evidence in a criminal case.

Raster Derivation

CSIRO produced various soil and landscape attribute rasters by the method described below (from [50]): The Soil and Landscape Grid of Australia (SLGA)'s Australia-wide Soil Attribute Maps were generated using measured soil properties from existing databases and laboratory spectroscopic measurements. The spatial modeling was performed using decision trees with piecewise linear models and kriging of residuals. Fifty environmental covariates that represent climate, biota, terrain and soil parent material were used in the modeling.

Uncertainty was derived using a bootstrap (Monte Carlo type) approach to derive for each pixel a probability density function (pdf), from which the Estimated Value (EV; 50th percentile) and 90% confidence limits (5th percentile and 95th percentile) were derived. The approach is described in detail in CSIRO (50) and references therein.

Available Attributes

Soil attributes provided by the TERN initiative are as follows: bulk density (whole earth), organic carbon, sand, silt, clay, pH water, pH $CaCl_2$, available water capacity, total nitrogen, total phosphorus, effective cation exchange capacity, depth of regolith, depth of soil, and coarse fragments. Landscape attributes provided are as follows: slope, slope relief classification, aspect, relief 1000 m radius, relief 300 m radius, Topographic Wetness Index (TWI), Topographic Position Index (TPI), partial contributing area, Multi-Resolution Valley Bottom Flatness (MRVBF), plan curvature, profile curvature, Prescott Index, Solar Radiation Model (SRAD) net radiation January, SRAD net radiation July, SRAD total shortwave sloping surface January, and SRAD total shortwave sloping surface July (50). Those attributes are available for several depth layers.

Used Attributes

The attributes used for developing and illustrating the PSP method here are as follows: total phosphorus (P), total nitrogen (N), pH ($CaCl_2$), sand, silt, and clay for the 0–5 cm depth layer. They were selected because they are (i) mainstream soil parameters collected by many geochemical surveys, (ii) generally robust (i.e., total element contents), (iii) fundamental soil attributes (e.g., pH, texture), and/or (iv) could be readily exploited in a forensic context. These attributes are easily measured in forensic or commercial laboratories. The metadata provided on the CSIRO Data Portal (https://data.csiro.au/dap/) for these attributes is given in Appendix 1 for completeness.

Other potentially useful soil and landscape attributes are expected to be made available from various organizations in the future, notably more chemical elements, mineral maps (including from satellite-borne ASTER technology), and isotopic landscapes ("isoscapes"). For the purposes of developing, illustrating, and demonstrating the potential of the PSP concept, however, the above selection of variables will suffice.

Data

The data rasters were downloaded from the CSIRO Data Portal (https://data.csiro.au/dap/) using an internet browser. The large downloads occur via the WebDAV protocol, which necessitates a request being sent and download credentials (username and password) being received by automated return email. The EV, 5th percentile and 95th percentile national grids for total P, total N, pH, sand, silt, and clay in the top 5 cm of soil (each ~4 GB in size after compression for two billion pixels) were downloaded. For the purposes of this publication, the national grids were "cookie-cut" to the State boundary of New South Wales (NSW; area 800,642 km^2). Each final grid cropped to NSW had a size of ~660 MB for ~110 million pixels. Figure 2 shows the raw rasters for these six attributes cut to the NSW border.

Data from often-small evidentiary (forensic) samples need to match the selected attributes discussed above. Soil total N can be determined on as little as 1 mg of sample (modified Pregl–Dumas method), total P requires 0.6 g of sample (fusion X-ray fluorescence), soil pH can be determined on <0.1 g of sample (colorimetry or microelectrode potentiometry), and soil texture requires <0.2 g of sample (Coulter counter/Multisizer method) (51–55). Thus, a minimum sample size of 1 g is sufficient to provide a determination of the selected properties, though a larger amount would allow replicates to be analyzed, thus improving precision. The fusion XRF method is a destructive method, which has implications for workflow prioritisation of forensic analyses.

Software

All raster operations were performed in the free, open-source GIS software QGIS (v. 2.18), with any additional calculation, tabulation, and graphing being done in Excel from the Microsoft Office Professional Plus 2010 suite. The GIS raster operation for six conditions as used here over ~110 million pixels takes only a few minutes on a standard networked personal computer.

Results

Evidentiary Soil Sample

We test the PSP method by determining the provenance area for a hypothetical evidentiary soil sample having the property (target) values listed in Table 1. We explicitly account for confidence limits around those values by applying a search range, instead of a single value, for each soil value, as also shown in Table 1. These search ranges are considered to reflect conservative (i.e., wide) confidence limit brackets, which we adopted for four reasons. Firstly, they are meant to include sample collection, preparation, and analysis errors as well as the modeling uncertainty (provided as the 5th percentile and 95th percentile soil attribute raster grids). Secondly, they include the error associated with comparing data from representative soil geochemical survey samples with often imperfect forensic evidentiary samples. Thirdly, it is considered that allowing wide search ranges (or "slack" conditions) will deliver a more convincing test for, and demonstration of, the utility of this novel technique. Fourthly, this conservative approach minimizes the occurrence of false negatives.

Soil Rasters

Figure 3 shows the raster maps for the six selected soil attributes, with pixels that satisfy each search range condition as shown in Table 1 represented in black (score of one), and those that do not represented in white (score of zero).

Equation 1 below shows the script using inequalities and the logical operator "AND" for the cumulative selection of pixels in the NSW-clipped rasters:

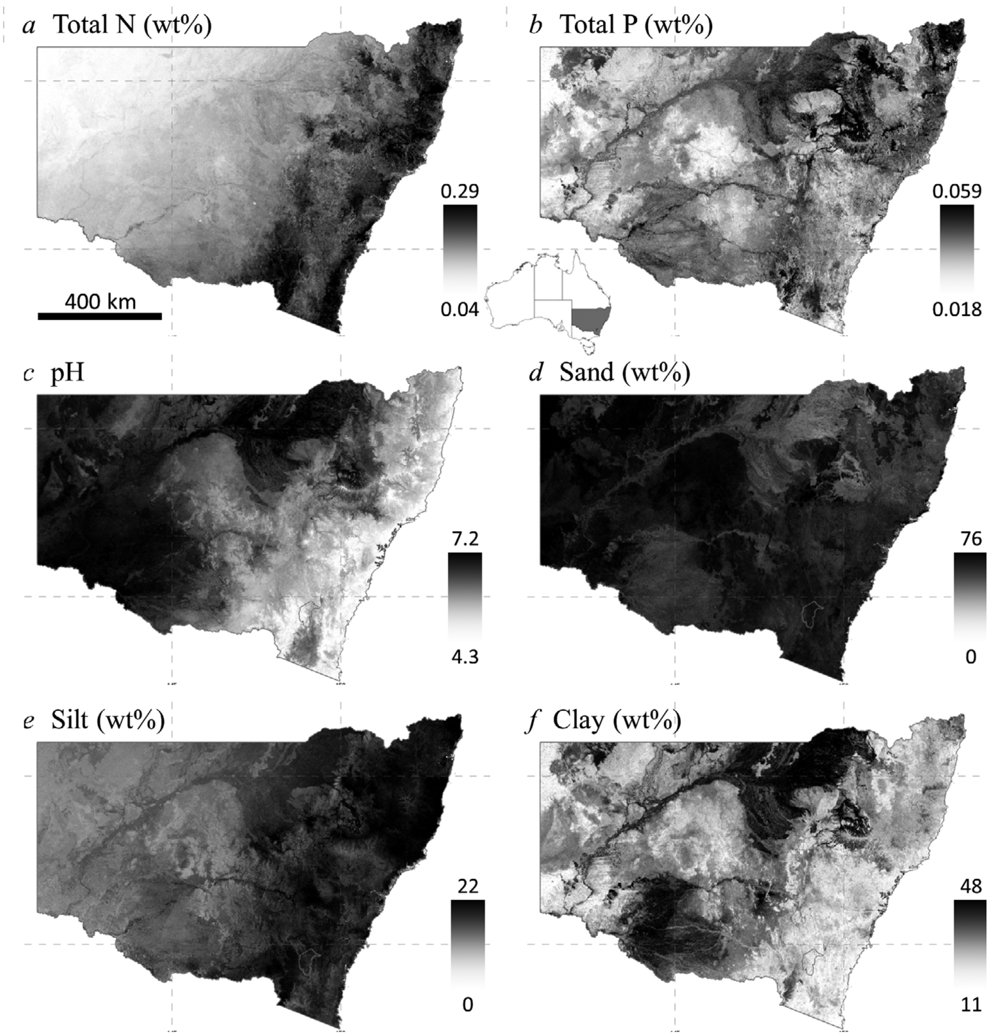


FIG. 2—*Original soil attribute rasters for NSW: (a) Total N (full range, white to black: 0.04–0.29 wt%); (b) Total P (0.018–0.059 wt%); (c) pH (4.3–7.2); (d) Sand (0–76 wt%); (e) Silt (0–22 wt%); and (f) Clay (11–48 wt%). Inset shows location of NSW (gray) in Australia.*

```
( ( 0.05 <= "NTO_EV_NSW_clip@1" AND "NTO_EV_NSW_clip@1" <= 0.2 ) + (
0.020 <= "PTO_EV_NSW_clip@1" AND "PTO_EV_NSW_clip@1" <= 0.040 ) + ( 4.5
<= "PHC_EV_NSW_clip@1" AND "PHC_EV_NSW_clip@1" <= 6.5 ) + ( 30 <=
"SND_EV_NSW_clip@1" AND "SND_EV_NSW_clip@1" <= 60 ) + ( 10 <=
"SLT_EV_NSW_clip@1" AND "SLT_EV_NSW_clip@1" <= 20 ) + ( 25 <=
"CLY_EV_NSW_clip@1" AND "CLY_EV_NSW_clip@1" <= 55 ) )
```

(1)

TABLE 1—*Soil property target values for the hypothetical evidentiary soil sample as well as the search ranges (target values ± confidence limits) for pixel selection.*

Soil Property	Unit	Target Value	Search Range
Total N	wt%	0.1	0.05–0.20
Total P	wt%	0.030	0.020–0.040
pH	N/A	5.5	4.5–6.5
Sand	wt%	45	30–60
Silt	wt%	15	10–20
Clay	wt%	40	25–55

where NTO = Total N; PTO = Total P; PHC = pH; SND = Sand; SLT = Silt; CLY = Clay; EV = Estimated Value. This equation can be directly copied and pasted into the QGIS raster calculation tool. At each pixel, a score of 1 or 0 is ascribed for each of the six conditions in Eq. 1 depending on whether it is satisfied or not satisfied, respectively. Consequently, each pixel where all six conditions are satisfied gets a score of six.

Figure 4a shows the final results of applying the PSP method, where each pixel is colored according to the number of conditions

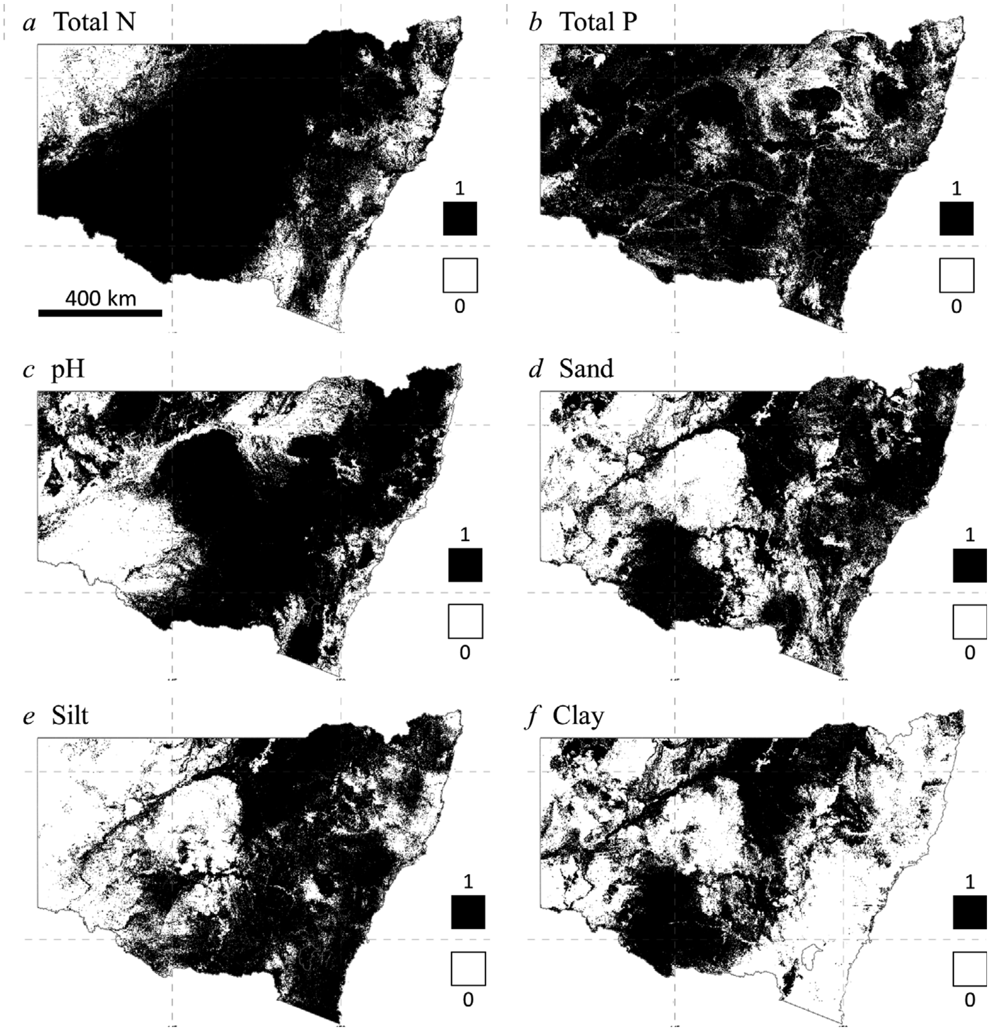


FIG. 3—*Final soil attribute rasters for NSW (0 = search range not satisfied; 1 = search range satisfied): (a) Total N (search range = 0.05–0.20 wt%); (b) Total P (0.020–0.040 wt%); (c) pH (4.5–6.5); (d) Sand (30–60 wt%); (e) Silt (10–20 wt%); and (f) Clay (25–55 wt%).*

that are satisfied at that point (red = all six conditions are satisfied, to dark blue = none of the conditions are satisfied). Figure 4b shows in black only the pixels with a full score of six.

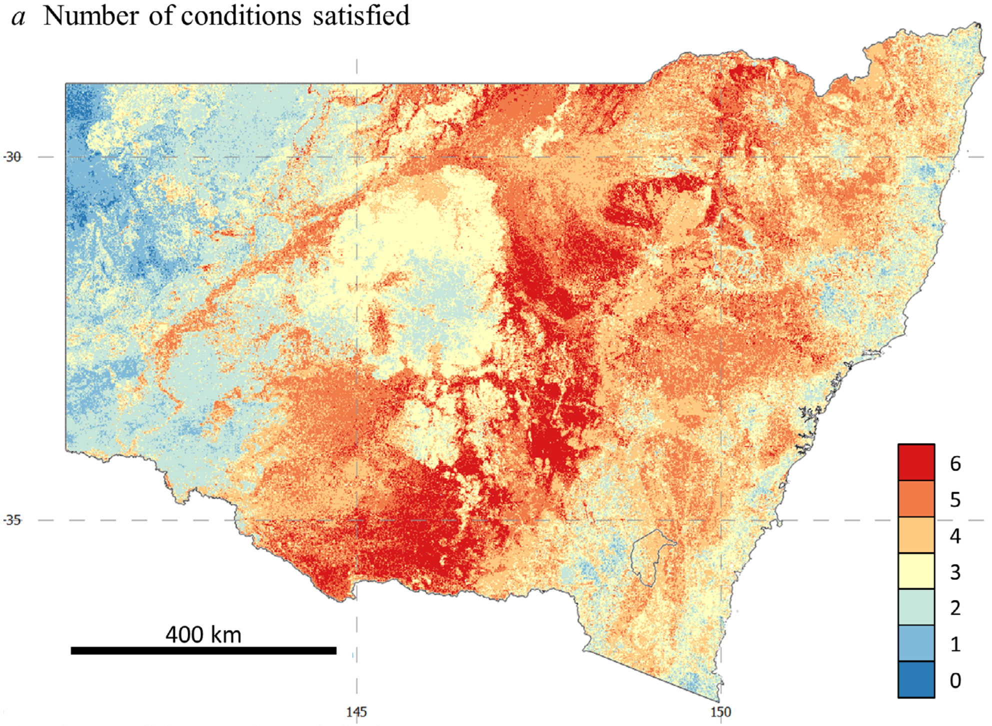
Discussion

Table 2 shows the number of pixels satisfying each conditional statement and the area they represent. Any condition taken individually only reduces the search area to between 620,000 km^2 and 320,000 km^2 of the initial area of NSW (800,642 km^2). When combined, though, the search area decreases significantly to only 67,000 km^2, or about 8.4% of the initial area.

The rasters of individual search conditions, considered individually, reduce the search area to between only 77% to 40% of NSW (Fig. 5). This is still a very large area, impossible to survey practically in detail for forensic evidence such as a cache or a grave. When all six of these conditional statements are considered together, however, the resulting search area is reduced to 8.4% of the initial investigation area (Fig. 5).

A validation test was performed on three widely spaced soil samples from NSW (Table 3), which reflect a range of soil types and textures (vertosol to sodosol; loam to silt loam). The samples come from the uppermost part of the soil profiles to mimic the material that would be transferred to a shoe, tire, or digging implement, and also to be compatible with the 0-5 cm depth range represented by the soil attribute rasters used in the test. The samples were collected as three to five subsamples composited together to give a sample more representative of each area. In each case, a dry bulk sample of 100 g or more was homogenized and riffle split down to the size required for analysis (a few g). Importantly, these samples were independent of the dataset used to create the soil attribute grids. To compensate for the closure (constant sum) effect in the textural data, the sand values were ignored; also, total N concentration was not measured, leaving four attributes available for the PSP analysis validation. The lower and upper limits of the search ranges were set to reflect average uncertainties for the area of interest. The match criteria (raster operations) for the validation samples' search ranges relative to the rasters' uncertainty envelopes (5th and 95th percentiles, see above) were calculated according to the following raster operation (for sample #11636):

a Number of conditions satisfied



b Six conditions only satisfied

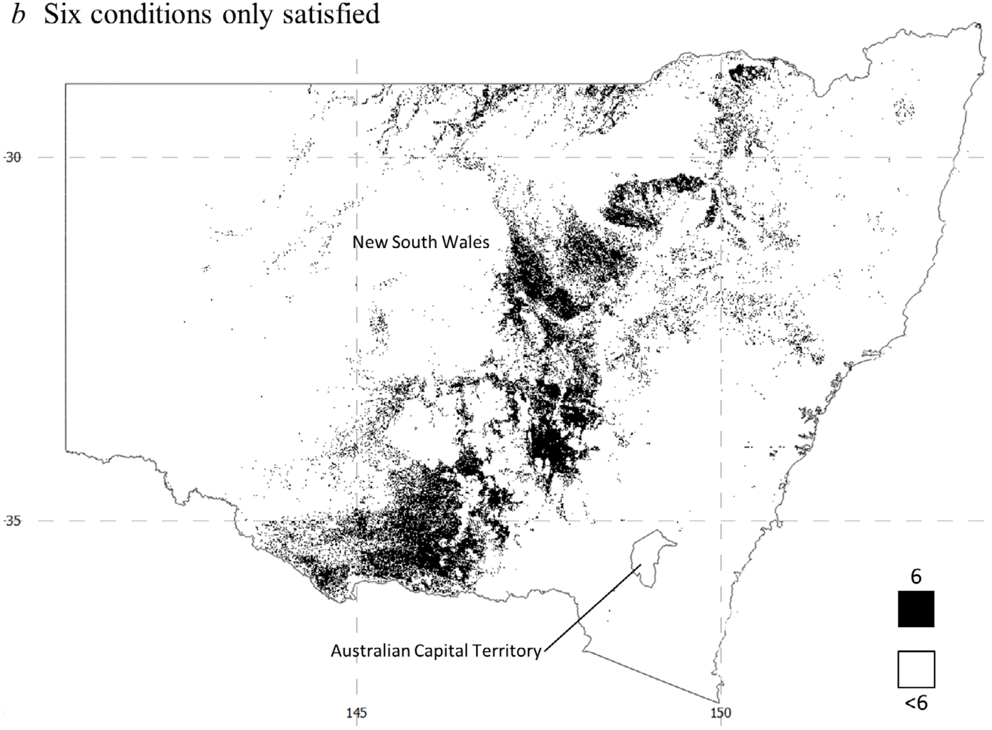


FIG. 4—*Predictive Soil Provenancing maps of NSW for evidentiary soil sample satisfying the conditions stated in Table 1: (a) Pixels satisfying from zero (dark blue) to all six (red) conditions; (b) Only pixels satisfying all six conditions (black) are shown to the exclusion of all others. Graticule in degrees of longitude and latitude. [Color figure can be viewed at wileyonlinelibrary.com]*

```
( ( 0.031 <= "PTO_95_NSW_clip@1" AND "PTO_05_NSW_clip@1" <= 0.058 ) + (
5.9 <= "PHC_95_NSW_clip@1" AND "PHC_05_NSW_clip@1" <= 7.5 ) + ( 14 <=
"SLT_95_NSW_clip@1" AND "SLT_05_NSW_clip@1" <= 78 ) + ( 14 <=
"CLY_95_NSW_clip@1" AND "CLY_05_NSW_clip@1" <= 21 ) )
```

(2)

TABLE 2—*Number of pixels and area satisfying the six conditional statements separately as well as all six conditions simultaneously.*

Condition	Pixels	Area (km^2)	Area (%)
None	110,506,516	800,642	100%
$0.05 \leq$ Total N ≤ 0.2	84,733,831	613,914	76.7%
$0.020 \leq$ Total P ≤ 0.040	85,374,187	618,553	77.3%
$4.5 \leq$ pH ≤ 6.5	71,838,022	520,481	65.0%
$30 \leq$ Sand ≤ 60	59,133,194	428,432	53.5%
$10 \leq$ Silt ≤ 20	58,571,993	424,366	53.0%
$25 \leq$ Clay ≤ 55	44,452,196	322,065	40.2%
All 6 conditions together	9,312,416	67,470	8.4%

where PTO = Total P; PHC = pH; SLT = Silt; CLY = Clay; 05 = 5th percentile lower confidence limit; 95 = 95th percentile lower confidence limit. In all cases, the pixel where each sample originated from was correctly identified by PSP for all four variables. Figure 6 shows the spatial search results for these three validation samples. Considering that only four variables were used and that total P was determined on a finer grain-size fraction (<180 μm) than that represented by the soil attribute rasters (<2 mm), the results are deemed very encouraging.

The PSP method is very effective at narrowing down a search area to those regions of the initial investigation area that are most likely to match, within analytical confidence limits and modeling uncertainty, the properties of the evidentiary soil sample. It does, however, rely on the availability of sufficient material for the relevant analyses (of the order of 1 g for methods used here), some of which are destructive. If additional forensic evidence comes to hand, such as for instance knowledge of a limited travel radius of a suspect from a given township, or DNA suggesting a specific ecosystem environment, the area to be searched for forensic evidence could be reduced even further.

When applying the PSP technique to some forensic cases, it may be necessary to consider the integrity of the evidentiary soil sample. For instance, an evidentiary soil sample collected from a car tire or a shoe sole may be altered compositionally due to the variable transferability/persistence of the different grain-size fractions of soils (e.g., [56,57]). Therefore, in future applications it may be more appropriate to use the silt and clay fractions, possibly ratioed to each other, rather than the sand fraction if loss of coarse particles during transfer is suspected. Other than in the latter case, we recommend that, when applied to forensics cases, all soil attribute rasters for which an analytical value has been obtained (or can be obtained) from the evidentiary soil sample be used. Therefore, there should be no need to pick and choose a selection of rasters, which could potentially bias the outcome. Note that the size of a search area cannot be increased by including another attribute (raster) and its associated search conditions; it can only be preserved or reduced. Therefore, by using as many soil properties as possible as the recommended procedure, we not only avoid bias but also get the most effective reduction in size of the search area.

The PSP tool is transparent because of the fully documented method developed here, including the raster operation script,

TABLE 3—*Soil property target values for three validation soil samples as well as the search ranges (target values ± confidence limits) for pixel selection.*

Sample	Lat	Long	Soil Property	Unit	Target Value	Search Range
#1103	−29.97	148.16	Total P	wt%	0.056	0.037–0.075
			pH	N/A	8.7	7.4–10.1
			Silt	wt%	54	17–91
			Clay	wt%	25	0–69
#1143	−31.17	141.90	Total P	wt%	0.043	0.028–0.057
			pH	N/A	8.6	7.3–9.9
			Silt	wt%	47	15–78
			Clay	wt%	9	0–25
#11636	−35.91	145.94	Total P	wt%	0.045	0.031–0.058
			pH	N/A	6.7	5.9–7.5
			Silt	wt%	46	14–78
			Clay	wt%	17	14–21

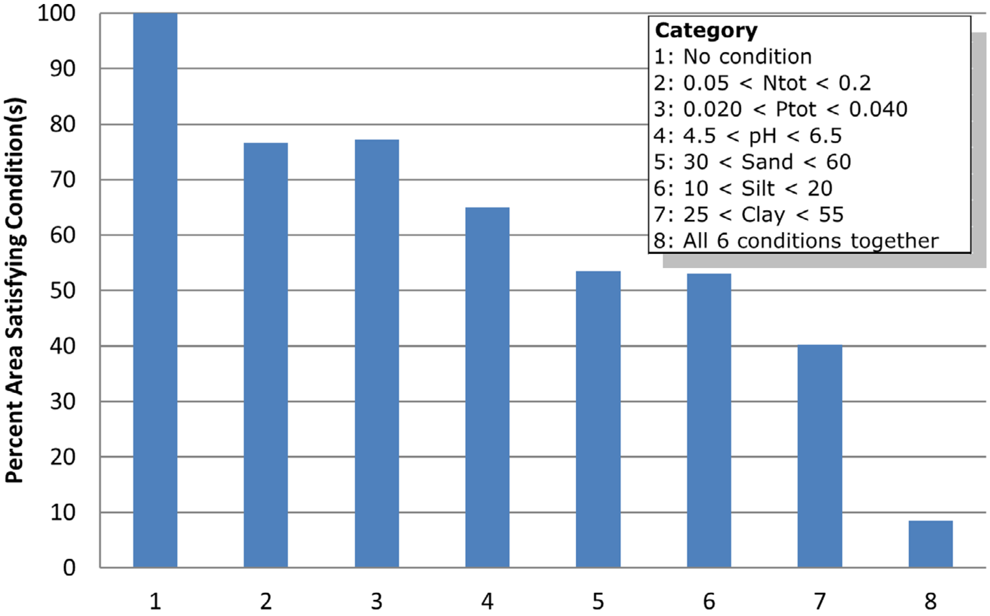
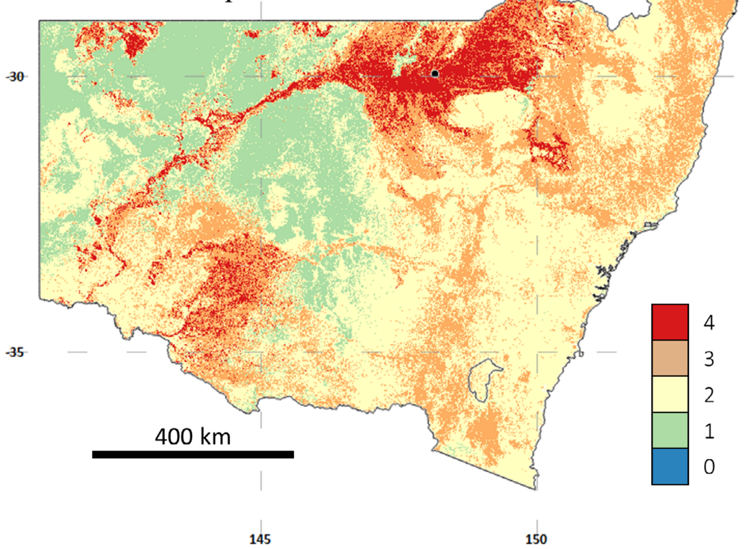
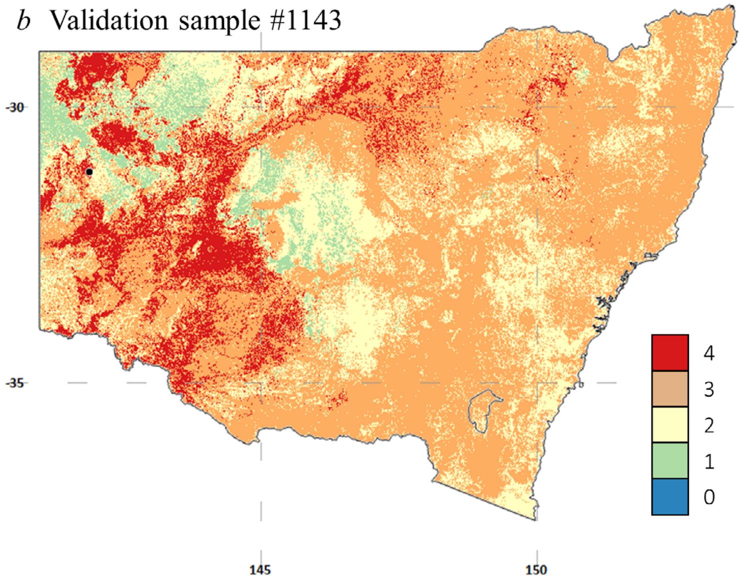


FIG. 5—*Bar graph showing the size reduction achieved by each search condition considered individually (bars 2–7; 40–77%) compared to the original investigation area of NSW (bar 1). Bar (8) shows the reduction in search area (8.4%) achieved when all six conditions are considered together. [Color figure can be viewed at wileyonlinelibrary.com]*

a Validation sample #1103



b Validation sample #1143



c Validation sample #11636

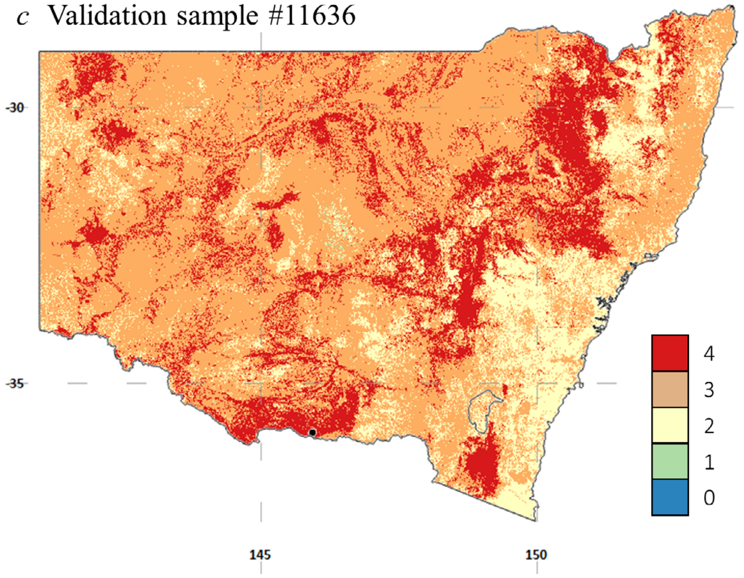


FIG. 6—*Predictive Soil Provenancing maps of NSW for three validation soil samples satisfying the conditions stated in Table 3. Pixels are colored according to number of conditions satisfied from zero (dark blue) to all four (red): (a) Sample #1103 located in northern NSW (black dot); (b) Sample #1143 located in western NSW (black dot); Sample #11636 located in southern NSW (black dot). [Color figure can be viewed at wileyonlinelibrary.com]*

which can be copied and pasted into a GIS. It is also reproducible because the soil rasters are public domain, open-source GIS software is used and the pixel selection is fully deterministic (as opposed to stochastic), so that anyone with an Internet connection and a personal computer will obtain the same results. To ensure that the PSP method is objective, we recommend that all available soil attribute rasters be used together, rather than selecting certain rasters over others, which could influence the outcome.

Compared to the traditional empirical approach, the PSP method presented here has the advantages of (i) being applicable to a whole country or even continent, (ii) having a high spatial resolution, and (iii) being transparent, reproducible, and objective. Its main limitations, on the other hand, are that rasters are (i) currently available only for selected bulk soil attributes (i.e., not many chemical or mineralogical properties), which may differ from those measured on typical forensic samples, (ii) the product of modeling, which although conducted in as rigorous a manner as possible, still only delivers a representation of reality based on data mining, and (iii) partially reliant on analytical methods that may require relatively large samples (1 g or more) and/or be destructive.

Future work developing PSP should test the robustness of the method (e.g., by applying it to other geographic areas, by using different combinations of soil attributes in the search conditions, or by altering the scoring method for pixels close to the target value from a binary to a continuous scale score). Application to a real forensic case would highlight any limitations such as the amount and composition of evidentiary soil samples available to forensic analysis (e.g., [43]) in comparison with samples collected during geochemical surveys, and perhaps come up with further recommendations for the application of PSP to forensic soil provenancing. New soil attribute rasters for major soil oxides are in preparation by various organizations. When these become available, additional elements all available from a single, parsimonious analysis (XRF) will add significant discriminatory power to the PSP method. Collaboration with the digital soil mapping community could result in new soil attribute rasters being generated that have direct applicability to forensic work, such as properties that can be obtained on small samples and/or by nondestructive analytical techniques. Finally, the continued collection and analysis of soil samples at various spatial densities is of course encouraged both pre-emptively as the core business of various agencies and as follow-ups of search areas for forensic purposes.

Conclusions

The novel Predictive Soil Provenancing (PSP) method for determining the likely origin of an evidentiary soil sample is introduced. PSP is underpinned by the high-resolution soil attribute grids recently released for instance in Australia for the whole continent at ~90 m resolution pixels. These grids were derived by CSIRO using data mining (decision trees with piecewise linear models and kriging of residuals). Simple raster operations in an open-source Geographic Information System allow for the rapid selection of pixels that satisfy any number of compositional conditions. The conditions are entered as a script of logical operators comprising for each variable a range of values (search range) that includes the target (evidentiary soil) value bracketed by its confidence limits. It is shown using New South Wales, Australia, as an exemplar that a search area can be reduced to as little as 8.4% of the original investigation area by considering only six soil attributes. In the absence of geochemical/soil surveys at a sufficient density and/or of a sufficient quality, the PSP method offers a transparent, reproducible, and objective alternative approach based on best available scientific data, process, and knowledge.

References

1. Ruffell A. Forensic pedology, forensic geology, forensic geoscience, geoforensics and soil forensics. Forensic Sci Int 2010;202:9–12.
2. Young JM, Rawlence NJ, Weyrich LS, Cooper A. Limitations and recommendations for successful DNA extraction from forensic soil samples: a review. Sci Justice 2014;54:238–44.
3. Zala K. Dirty science: soil forensics digs into new techniques. Science 2007;318:386–7.
4. Dalpe C, Blanchard C, Chartrand M, St-Jean G, Wojtyk J. Trace element analysis of Canadian surface soil and associated quartz by LA-ICP-MS: preliminary development of a geo-location database for forensic investigations. Proceedings of the International Annual Meetings of the American Society of Agronomy (ASA), Crop Science Society of America (CSSA) and Soil Science Society of America (SSSA); 2010 Oct 31-Nov 4; Long Beach, CA. Madison, WI: ACSESS Digital Library, 2010;63392.
5. Woods B, Lennard C, Kirkbride KP, Robertson J. Soil examination for a forensic trace evidence laboratory – Part 1: spectroscopic techniques. Forensic Sci Int 2014a;245:187–94.
6. Woods B, Kirkbride KP, Lennard C, Robertson J. Soil examination for a forensic trace evidence laboratory – Part 2: elemental analysis. Forensic Sci Int 2014b;245:195–201.
7. Di Maggio RM, Donnelly L, Al Nuaimi KS, Barone PM, Da Silva Salvador FA, Dawson L, et al. Global developments in forensic geology. Episodes 2017;40:120–31.
8. Webb JB, Bottrell M, Stern LA, Saginor I. Geology of the FBI lab and the challenge to the admissibility of forensic geology in US court. Episodes 2017;40:118–9.
9. Fitzpatrick R. Getting the dirt – the value of soil in criminal investigations. Royal Canadian Mounted Police Gazette 2011;73:22–3.
10. Dawson L, Fitzpatrick R. The use of soil in criminal investigations. Geophys Res Abs, EGU Gen Assembly, 2013;15:11558.
11. Santillana E, Cordero JC, Alamilla F. Forensic soil analysis: case study of looting at a Roman-Visigothic burial vault. In: Kars H, van den Eijkel L, editors. Soil in criminal and environmental forensics. Cham, Switzerland: Springer International Publishing, 2016;45–60.
12. Rawlins BG, Kemp SJ, Hodgkinson EH, Riding JB, Vane CH, Poulton C, et al. Potential and pitfalls in establishing the provenance of earth-related samples in forensic investigations. J Forensic Sci 2006;51:832–45.
13. Ritz K, Dawson L, Miller D, editors. Criminal and environmental soil forensics. Berlin/Heidelberg, Germany: Springer Science + Business Media B.V., 2009.
14. Lark RM, Rawlins BG. Can we predict the provenance of a soil sample for forensic purposes by reference to a spatial database? Eur J Soil Sci 2008;59:1000–6.
15. Bowen AM, Caven EA. Forensic provenance investigations of soil and sediment samples. In: Pirrie D, Ruffell A, Dawson LA, editors. Environmental and criminal geoforensics. London: GSL Special Publications, 2013;384:9–25.
16. Darnley AG, Björklund A, Bølviken B, Gustavsson N, Koval PV, Plant JA, et al. A global geochemical database for environmental and resource management. Recommendations for international geochemical mapping, final report of IGCP Project 259. Paris, France: UNESCO Publishing, 1995.
17. Garrett RG, Reimann C, Smith DB, Xie X. From geochemical prospecting to international geochemical mapping: a historical overview. Geochem Explor Environ Anal 2008;8:205–17.
18. McKinley J. The application of geographic information system (GIS) in forensics geoscience. Episodes 2017;40:166–71.
19. Kürzl H. Exploratory data analysis: recent advances for the interpretation of geochemical data. J Geochem Explor 1988;30:309–22.
20. Johnson CC. Within site and between site nested analysis of variance (ANOVA) for geochemical surveys using MS Excel. British Geological Survey Report 2002;IR/02/043. http://nora.nerc.ac.uk/id/eprint/8364 (accessed March 27, 2019).
21. Templ M, Filzmoser P, Reimann C. Cluster analysis applied to regional geochemical data: problems and possibilities. Appl Geochem 2008;23:2198–213.
22. Campbell GP, Curran JM, Miskelly GM, Coulson S, Yaxley GM, Grunsky EC, et al. Compositional data analysis for elemental data in forensic science. Forensic Sci Int 2009;188:81–90.

23. Grunsky EC. The interpretation of geochemical survey data. Geochem Explor Environ Anal 2010;10:27–74.
24. Caritat P de, Grunsky EC. Defining element associations and inferring geological processes from total element concentrations in Australian catchment outlet sediments: multivariate analysis of continental-scale geochemical data. Appl Geochem 2013;33:104–26.
25. Aitkenhead MJ, Coull MC, Dawson LA. Predicting sample source location from soil analysis using neural networks. Environ Forensics 2014;15:281–92.
26. Bonetti J, Quarino L. Comparative forensic soil analysis of New Jersey State Parks using a combination of simple techniques with multivariate statistics. J Forensic Sci 2014;59:627–36.
27. Cracknell MJ, Reading AM, Caritat P de. Multiple influences on regolith characteristics from continental-scale geophysical and mineralogical remote sensing data using Self-Organizing Maps. Remote Sens Environ 2015;165:86–99.
28. Harris JR, Grunsky EC. Predictive lithological mapping of Canada's North using Random Forest classification applied to geophysical and geochemical data. Comput Geosci 2015;80:9–25.
29. Zuo R, Carranza EJM, Wang J. Spatial analysis and visualization of exploration geochemical data. Earth Sci Rev 2016;158:9–18.
30. Kment P, Mihaljevič M, Ettler V, Šebek O, Strnad L, Rohlová L. Differentiation of Czech wines using multielement composition – a comparison with vineyard soil. Food Chem 2005;91:157–65.
31. Feng J-L, Zhu L-P, Zhen X-L, Hu Z-G. Grain size effect on Sr and Nd isotopic compositions in eolian dust: implications for tracing dust provenance and Nd model age. Geochem J 2009;43:123–31.
32. Feng J-L, Hu Z-G, Ju J-T, Zhu L-P. Variations in trace element (including rare earth element) concentrations with grain sizes in loess and their implications for tracing the provenance of eolian deposits. Quat Int 2011;236:116–26.
33. Pye K, Blott SJ. Development of a searchable major and trace element database for use in forensic soil comparisons. Sci Justice 2009;49:170–81.
34. Frei R, Frei KM. The geographic distribution of Sr isotopes from surface waters and soil extracts over the island of Bornholm (Denmark) – a base for provenance studies in archeology and agriculture. Appl Geochem 2013;38:147–60.
35. Woods B, Lennard C, Kirkbride KP, Robertson J. Soil examination for a forensic trace evidence laboratory – Part 3: a proposed protocol for the effective triage and management of soil examinations. Forensic Sci Int 2016;262:46–55.
36. Caritat P de, Mann A. An improved method for assessing the degree of geochemical similarity (DOGS2) between samples from multi-element geochemical datasets. Geochem Explor Environ Anal 2019;19:58–73.
37. McKinley J. How useful are databases in environmental and criminal forensics? In: Pirrie D, Ruffell A, Dawson LA, editors. Environmental and criminal geoforensics. Special Publication, Book 384. London: Geological Society of London, 2013;109–19.
38. Johnson CC, Breward N, Ander EL, Ault L. G-BASE: baseline geochemical mapping of Great Britain and Northern Ireland. Geochem Explor Environ Anal 2005;5:347–57.
39. Smyth D. Methods used in the Tellus geochemical mapping of Northern Ireland. British Geological Survey, Report 2007;OR/07/022. http://nora.nerc.ac.uk/14008/ (accessed March 27, 2019).
40. Lagacherie P, McBratney AB, Volz M, editors. Digital soil mapping: an introductory perspective (Developments in soil science, book 31). Amsterdam, Netherlands: Elsevier, 2007;31.
41. Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, et al. Digital soil map of the world. Science 2009;325:680–1.
42. Ruffell A, McKinley J. Geoforensics. Chichester: Wiley-Blackwell, 2008.
43. Pirrie D, Dawson L, Graham G. Predictive geolocation: forensic soil analysis for provenance determination. Episodes 2017;40:141–7.
44. Menchaca PR, Graham RC, Younglove T. Developing and testing a soil property database for forensic applications in southern California. J Forensic Sci 2018;63:1043–52.
45. Stern LA, Webb JB, Willard DA, Bernhardt CE, Korejwo DA, Bottrell MC, et al. Geographic attribution of soils using probabilistic modeling of GIS data for forensic search efforts. Geochem Geophys Geosyst 2019;19(20):913–32.
46. ESDAC (European Soil Data Centre). Topsoil physical properties for Europe. European Commission Joint Research Centre, 2015. https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data (accessed March 27, 2019).
47. NRCS (Natural Resources Conservation Service). Gridded Soil Survey Geographic (gSSURGO) database – user guide (version 2.2). United States Department of Agriculture, 2016. https://www.nrcs.usda.gov/wps/PA_NRCSConsumption/download?cid=nrcs142p2_051847&ext=pdf (accessed March 27, 2019).
48. ISRIC (International Soil Reference and Information Centre). Soil property maps of Africa at 250 m resolution. https://www.isric.org/projects/soil-property-maps-africa-250-m-resolution (accessed March 27, 2019).
49. SoilGrids. Global gridded soil information. https://soilgrids.org (accessed March 27, 2019).
50. CSIRO (Commonwealth Scientific and Industrial Research Organisation). Soil and landscape grid of Australia, 2017. http://www.clw.csiro.au/aclep/soilandlandscapegrid/index.html (accessed March 27, 2019).
51. Shah GD, Pansare VS, Mulay VN. A modified micro-Dumas method for rapid determination of nitrogen. Microchim Acta 1956;44:1140–3.
52. Hutton JT, Elliott SM. An accurate XRF method for the analysis of geochemical exploration samples for major and trace elements using one glass disc. Chem Geol 1980;29:1–11.
53. Dudley RJ. A simple method for determining the pH of small soil samples and its use in forensic science. J Forensic Sci Soc 1976;16:21–7.
54. Dudley RJ. The particle size analysis of soils and its use in forensic science – the determination of particle size distributions within the silt and sand fractions. J Forensic Sci Soc 1976;16:219–29.
55. McTainsh GH, Lynch AW, Hales R. Particle-size analysis of aeolian dusts, soils and sediments in very small quantities using a Coulter Multisizer. Earth Surf Process Landf 1997;22:1207–16.
56. Chisum W, Turvey B. Evidence dynamics: Locard's exchange principle and crime reconstruction. J Behav Prof 2000;1:1–15.
57. Fitzpatrick RW, Raven MD, Forrester ST. A systematic approach to soil forensics: criminal case studies involving transference from crime scene to forensic evidence. In: Ritz K, Dawson L, Miller D, editors. Criminal and environmental soil forensics. Dordrecht, Netherlands: Springer, 2009;105–27.

Appendix 1: Predictive Soil Provenancing (PSP): An Innovative Forensic Soil Provenance Analysis Tool

The metadata provided on the CSIRO Data Portal (https://data.csiro.au/dap/) for the attributes used in this publication are listed below for completeness (from 50).

Total Phosphorus

This is version 1 of the Australian Soil Total Phosphorus product of the Soil and Landscape Grid of Australia. The Soil and Landscape Grid of Australia has produced a range of digital soil attribute products. Each product contains six digital soil attribute maps, and their upper and lower confidence limits, representing the soil attribute at six depths: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These depths are consistent with the specifications of the GlobalSoilMap.net project (http://www.globalsoilmap.net/). The digital soil attribute maps are in raster format at a resolution of three arc-seconds (~90 × 90 m pixels). These maps are generated by combining the best available Digital Soil Mapping (DSM) products available across Australia. Attribute Definition: Total phosphorus; Units: %; Period (temporal coverage; approximately): 1950–2013; Spatial resolution: three arc-seconds (~90 m); Total number of gridded maps for this attribute: 18; Number of pixels with coverage per layer: 2007M (49200 * 40800); Total size before compression: about 8GB; Total size after compression: about 4GB; Data license : Creative Commons Attribution 3.0 (CC By); Target data standard: GlobalSoilMap specifications; Format: GeoTIFF.

Total Nitrogen

This is version 1 of the Australian Soil Total Nitrogen product of the Soil and Landscape Grid of Australia. The Soil and

Landscape Grid of Australia has produced a range of digital soil attribute products. Each product contains six digital soil attribute maps, and their upper and lower confidence limits, representing the soil attribute at six depths: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These depths are consistent with the specifications of the GlobalSoilMap.net project (http://www.globalsoilmap.net/). The digital soil attribute maps are in raster format at a resolution of three arc-seconds (~90 × 90 m pixels). These maps are generated by combining the best available Digital Soil Mapping (DSM) products available across Australia. Attribute Definition: Total nitrogen; Units: %; Period (temporal coverage; approximately): 1950–2013; Spatial resolution: three arc-seconds (~90 m); Total number of gridded maps for this attribute: 18; Number of pixels with coverage per layer: 2007M (49200 * 40800); Total size before compression: about 8GB; Total size after compression: about 4GB; Data license : Creative Commons Attribution 3.0 (CC By); Target data standard: GlobalSoilMap specifications; Format: GeoTIFF.

pH

This is version 1 of the Australian Soil pH–$CaCl_2$ product of the Soil and Landscape Grid of Australia. The Soil and Landscape Grid of Australia has produced a range of digital soil attribute products. Each product contains six digital soil attribute maps, and their upper and lower confidence limits, representing the soil attribute at six depths: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These depths are consistent with the specifications of the GlobalSoilMap.net project (http://www.globalsoilmap.net/). The digital soil attribute maps are in raster format at a resolution of three arc-seconds (~90 × 90 m pixels). These maps are generated by combining the best available Digital Soil Mapping (DSM) products available across Australia. Attribute Definition: pH of 1:5 soil/0.01M calcium chloride extract; Units: None; Period (temporal coverage; approximately): 1950–2013; Spatial resolution: three arc-seconds (~90 m); Total number of gridded maps for this attribute: 18; Number of pixels with coverage per layer: 2007M (49200 * 40800); Total size before compression: about 8GB; Total size after compression: about 4GB; Data license : Creative Commons Attribution 3.0 (CC By); Target data standard: GlobalSoilMap specifications; Format: GeoTIFF.

Sand

This is version 1 of the Australian Soil Sand product of the Soil and Landscape Grid of Australia. The Soil and Landscape Grid of Australia has produced a range of digital soil attribute products. Each product contains six digital soil attribute maps, and their upper and lower confidence limits, representing the soil attribute at six depths: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These depths are consistent with the specifications of the GlobalSoilMap.net project (http://www.globalsoilmap.net/). The digital soil attribute maps are in raster format at a resolution of three arc-seconds (~90 × 90 m pixels). These maps are generated by combining the best available Digital Soil Mapping (DSM) products available across Australia. Attribute Definition: 200 μm - 2 mm mass fraction of the less than 2 mm soil material determined using the pipette method; Units: %; Period (temporal coverage; approximately): 1950–2013; Spatial resolution: three arc-seconds (~90 m); Total number of gridded maps for this attribute: 18; Number of pixels with coverage per layer: 2007M (49200 * 40800); Total size before compression: about 8GB; Total size after compression: about 4GB; Data license : Creative Commons Attribution 3.0 (CC By); Target data standard: GlobalSoilMap specifications; Format: GeoTIFF.

Silt

This is version 1 of the Australian Soil Silt product of the Soil and Landscape Grid of Australia. The Soil and Landscape Grid of Australia has produced a range of digital soil attribute products. Each product contains six digital soil attribute maps, and their upper and lower confidence limits, representing the soil attribute at six depths: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These depths are consistent with the specifications of the GlobalSoilMap.net project (http://www.globalsoilmap.net/). The digital soil attribute maps are in raster format at a resolution of three arc-seconds (~90 × 90 m pixels). These maps are generated by combining the best available Digital Soil Mapping (DSM) products available across Australia. Attribute Definition: 2–200 μm mass fraction of the less than 2 mm soil material determined using the pipette method; Units: %; Period (temporal coverage; approximately): 1950–2013; Spatial resolution: three arc-seconds (~90 m); Total number of gridded maps for this attribute: 18; Number of pixels with coverage per layer: 2007M (49200 * 40800); Total size before compression: about 8GB; Total size after compression: about 4GB; Data license : Creative Commons Attribution 3.0 (CC By); Target data standard: GlobalSoilMap specifications; Format: GeoTIFF.

Clay

This is version 1 of the Australian Soil Clay product of the Soil and Landscape Grid of Australia. The Soil and Landscape Grid of Australia has produced a range of digital soil attribute products. Each product contains six digital soil attribute maps, and their upper and lower confidence limits, representing the soil attribute at six depths: 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm. These depths are consistent with the specifications of the GlobalSoilMap.net project (http://www.globalsoilmap.net/). The digital soil attribute maps are in raster format at a resolution of three arc-seconds (~90 × 90 m pixels). These maps are generated by combining the best available Digital Soil Mapping (DSM) products available across Australia. Attribute Definition: 2 μm mass fraction of the less than 2 mm soil material determined using the pipette method; Units: %; Period (temporal coverage; approximately): 1950–2013; Spatial resolution: three arc-seconds (~90 m); Total number of gridded maps for this attribute: 18; Number of pixels with coverage per layer: 2007M (49200 * 40800); Total size before compression: about 8GB; Total size after compression: about 4GB; Data license : Creative Commons Attribution 3.0 (CC By); Target data standard: GlobalSoilMap specifications; Format: GeoTIFF.