Preprint of
Schmidt, A. & G. Tsetskhladze 2013. Raster was yesterday: using vector engines to process geophysical data. *Archaeological Prospection* 20(1): 59-65. DOI: 10.1002/arp.1443

See http://www.GeodataWIZ.com/armin-schmidt

Raster was yesterday: using vector engines to process geophysical data

by Armin Schmidt[a] and Gocha Tsetskhladze[b]

(a) GeodataWIZ (UK); A.Schmidt@GeodataWIZ.com

(b) University of Melbourne, Classics and Archaeology, School of Historical and Philosophical Studies (Australia)

Abstract

Tools for the processing of raster data are well developed, but noisy data still pose considerable challenges. If anomalies are broken up into isolated individual readings, for example due to high noise levels, it may still be possible for a human interpreter to recognise the isolated readings as being part of a single anomaly. However, such a concept of neighbourhood is difficult to implement with raster tools and an alternative, vector based approach is presented here. By converting the measured raster data into polygons, it is possible to undertake shape and neighbourhood analysis to process the data. This allows discriminating, reshaping and merging of the anomalies based on their spatial location relative to each other (neighbourhood) and with respect to the size of each anomaly. The added advantage of this approach is the possibility to use the processed vector data as a basis for interpretation and visualisation diagrams in two- and three dimensions. This method is applied to the GPR survey of a necropolis at Pessinus, showing several types of grave monuments.

Keywords

raster data, vector data, polygon, neighbourhood analysis, GPR, Pessinus

Introduction

Results from archaeological geophysical surveys are nowadays mostly represented as raster data, since the majority are collected on a raster of measurement positions, or close to a raster using GNSS guidance. In the beginning, though, when archaeological geophysics started as a sub-discipline of geological geophysical surveying, data collection was usually carried out on a very sparse grid, or even just with random sampling locations. This was partially a consequence of the very slow survey speed, but also because the surveys' objective simply was to find anomalies (e.g. from kilns) that could later be excavated, an approach derived from geological applications. Data were mostly visualised by creating contour diagrams, thereby interpolating between the sparse data points (Bevan 2000), or simply focusing on single traverses of measured data.

However, with the realisation that the majority of archaeological features exhibit a geophysical contrast (Le Borgne 1955; Clark 1968) came a shift towards detailed spatial characterisation of buried archaeological remains through geophysical surveys (Levels 2 and 3 of investigation, according to the classification by Gaffney & Gater (2003)). To facilitate this, measurements have to be collected with sufficient spatial density to represent the details of buried structures. Raster data are best suited to achieve a uniform data coverage of a survey area (Schmidt & Ernenwein 2011) and to facilitate processing and visualisation with pixel based computer systems. The fixed resolution avoids under-sampling of parts of a survey area that otherwise could result in the misinterpretation of features. It is for this reason that even un-gridded surveys with motorised or

manually pushed sensor arrays usually rely on GNSS guidance to achieve data recording as close to a predetermined grid as possible. In most cases raster data are subsequently formed from such guided measurements using a gridding algorithm (Li & Götze 1999), although slight artefacts may be introduced into the gridded data where gaps between survey lines have to be filled.

The processing of raster data is well supported and most software tools have useful filters. They are usually implemented as kernel convolution in the spatial domain (Schmidt 2003; Aspinall *et al.* 2008), very similar to image processing software. Since Fourier transformation of the data is fast on modern computers filtering can also be undertaken in the frequency domain, which is particularly useful if very broad background variations are to be removed with a high-pass filter (Aspinall & Haigh 1988). All raster tools have been refined over the years and can be considered mature.

Nevertheless there are some remaining challenges that raster based data processing cannot easily address. While a human interpreter may be able to identify adjacent pixels of high measurement values as belonging to the same anomaly, raster based processing is agnostic to the concept of 'neighbourhood'. This would not be a problem if all archaeological geophysics data were well above the noise level created by repeatable small-scale soil variations and random instrument fluctuations (called coherent and incoherent noise, respectively, by Scollar *et al.* (1990)). However, in many situations such noise breaks up anomalies that belong to the same feature into small isolated patches. This situation is sometimes also encountered in time-slices derived from single-channel GPR surveys, especially when the data were collected bidirectionally (i.e. zig-zag). This can be caused by polarisation effects when rotating antennas at the end of a line, by the different location of transmitter and receiver boxes relative to a target, or by a slight 'role' of the antennas on rough ground. With raster processing one of the few methods available to address such a problem is low-pass filtering between adjacent lines to even-out variations. However, this has the considerable disadvantage of smoothing the shape of anomalies and thereby degrading the spatial information content that can be gained from high-resolution surveys.

As an alternative approach we propose to convert the rasterised measurements into vector data (polygons) and process them with shape and neighbourhood analysis tools. This allows to discriminate, reshape and merge anomalies based on their spatial location relative to each other (neighbourhood) and with respect to the size of each anomaly. The added advantage of this approach is the possibility to use the processed vector data as a basis for interpretation and visualisation diagrams in two- and three dimensions.

Archaeological Data

The vector processing scheme will be illustrated with GPR measurements from one of the necropoleis at Ballıhisar-Pessinus in central Anatolia. Pessinus is the sacral place of Cybele, who became the Magna Mater, goddess of the Roman Empire. The city is situated in Central Phrygia, about 150 km south-southwest of Ankara. The Phrygian settlement developed into an important Roman city, embellished by successive Roman emperors with a temple for the Imperial Cult, a theatre, and a processional street ending in a monumental arch. In previous investigations of this site (Schmidt *et al.* 2011) most attention was focussed on the monumental buildings, located in a dry river valley cut into the Neogene/Pliocene high-plateau pediment (Brackman *et al.* 1995). However, 'shadow settlements' exist on these plateaus overlooking the city's valleys, formed by several extended necropoleis that were built at least from the Roman to the middle Byzantine period. Since there is a considerable and real threat of looting any targeted excavation would have drawn attention to the burial monuments and a non-invasive geophysical investigation was hence deemed necessary. For the same reason, the data reported here are not presented in a georeferenced form.

A high-resolution GPR survey was started in 2011 on one of the necropoleis (Schmidt & Tsetskhladze 2012) over an area of 60 m × 40 m that appeared entirely featureless apart from scatters of fine Roman pottery (Samian ware) as well as thick and coarse ceramics, probably from

roof tiles. As the plateau has been heavily eroded by severe spring rains and the survey area lies on a slight slope the actual location of these surface finds was not deemed to be directly correlated to subsurface structures. The annual erosion processes will have reshaped the plateau considerably since classical times, which is demonstrated by the large stones and limestone blocks that are deposited on the steep slopes that lead from it into the dry valleys. The spatial layout of the necropolis is of considerable archaeological interest, including the placing of burials in relation to the plateau's current edges, as well as the visibility of the city of Pessinus from the location of individual graves. This information will allow new insights into the relationship between the city and its cemeteries.

The GPR survey was undertaken with a Malå Ramac system and a 500 MHz antenna manually pulled on a sledge, recording traverses bidirectionally with a separation of 0.25 m and a trace separation of 0.04 m. The data were processed with bandpass filtering (trapezoidal Ormsby filter defined by 5, 300, 700, 900 MHz) and migration so that clear timeslices could be generated. The site's fairly uniform electromagnetic velocity was determined through migration tests to be 0.07 m/ns. The data were collected over data grids of 20 m × 20 m and the timeslices were processed with a Zero Median Traverse filter in these data grids to remove imbalances between survey lines. The timeslices are clear but suffer in some areas from a break-up of anomalies between adjacent survey lines so that neighbourhood processing was considered advantageous. Figure 1a shows the timeslice at 1.5 m depth.

Method

To facilitate the processing of measured data with neighbourhood vector tools they first have to be converted to polygons from their original raster format. Many methods have been suggested to delineate geophysical anomalies in raster data, most notably based on the calculation of gradients, for example the Normalised Gradient (Ma & Li 2012) or the Analytical Signal (Nabighian 1972; Tabbagh *et al.* 1997). Other methods are linked to the specific properties of certain geophysical data, like Euler Deconvolution of magnetometer surveys (Reid *et al.* 1990) or other complex parameters (Stampolidis & Tsokas 2012). However, all these methods become unreliable when the noise level in the data (the signal-to-noise ratio) becomes very poor. Since the proposed vector processing will mainly be applied to such noisy data these delineation tools were considered unsuitable and it was decided to use a simple thresholding method instead. The dynamic range of the necropolis survey in Pessinus is reasonably even across each individual GPR timeslice and a single threshold was therefore used for each timeslice. An alternative approach would have been to use a Wallis filter first to even out differences in contrast and intensity of the data (Scollar *et al.* 1990), but this was found to be unnecessary.

The timeslices were exported as GeoTIFF data files, which are well suited to represent the full data range, including dummy readings and spatial information, and then imported into an Open-Source PostGIS 2.0.1 database. PostGIS is a spatial extension to the extremely robust PostgreSQL database and follows the 'Simple Features for SQL' specification from the Open Geospatial Consortium (OGC). Far beyond being just a database for spatial data, PostGIS is a highly optimised processing engine that provides very fast routines for the complex manipulation of vector data (e.g. intersections, buffering), similar and sometimes even superior to stand-alone desktop GIS packages. Initially developed to handle vector data, raster support is directly included in PostGIS since version 2 (Obe & Hsu 2011). The result of a PostGIS vector processing SQL-statement can conveniently be visualised through OpenJUMP or QGIS.

After importing the georeferenced raster data into PostGIS, a processing sequence was implemented as a parameterised function that could be called repeatedly with slightly varied parameters to obtain processing results that were 'looking good'. In the following explanation, the parameters chosen for this example are provided in brackets.

- **Raster to vector conversion**: by using the descriptive statistics of the raster data, a threshold was defined in multiples of the data's standard deviation and all values above the threshold were returned as the initial anomaly-polygon (1.7 standard deviations; Figs. 1a & b).
- **Size threshold 1**: to reject very small anomalies that were assumed to be noise an area-threshold was applied to the anomaly-polygons (0.05 m^2, the diameter of an equivalent circle would be 0.25 m; Fig. 1c).
- **Neighbourhood merge**: all anomaly-polygons within a certain neighbourhood distance from each other were merged into a single polygon. For this all polygons were buffed with a neighbourhood distance and polygons that then overlapped were merged. It was found that a buffering distance of 0.2 m produced best results as it allowed to join even anomalies on adjacent lines (0.25 m spacing) that were 0.3 m shifted against each other along the lines (Fig 1d).
- **Size threshold 2**: to ensure that the neighbourhood merge led to sufficiently substantial anomalies, a second area-threshold was applied. While the first threshold was small to reject noise spikes (see above), the second threshold had to be considerably larger to select only fairly extensive anomaly-polygons (1.8 m^2, corresponding to a diameter of 1.5m; Fig. 1e).
- **Shrink with fringe**: as the neighbourhoods were determined by buffering the data, the resulting polygons were larger than the original anomalies. To reduce them to their original size, a negative buffer was applied to the neighbourhood-merged data. The ability to use a negative buffer radius is a particularly useful feature of PostGIS. However, to maintain a reasonably contiguous appearance of neighbouring anomalies the shrinking was not quite as large as the preceding buffer expansion, creating a thin fringe around the original anomalies (0.15 m fringe; Fig. 1f).
- **Generalisation**: to use the anomaly-polygons for simplified interpretation diagrams they were 'generalised' by reducing the number of vertices that define the corners of the polygons (Douglas-Peucker parameter of 0.2; Fig. 1g & h).

When applying this processing workflow to GPR timeslices it proved useful to store the depth and thickness of each raster timeslice as an attribute of the resulting anomaly-polygons. This allowed to display them in a three-dimensional data viewer as 'pancake slices' (Fig. 2).

Discussion and Conclusion

The extraction of anomalies from the GPR survey of the necropolis and a comparison with the timeslices allowed to distinguish three different types of anomalies.

- **Type A**: an elaborate grave monument built of stone blocks with a structural layout that changes considerably with depth, in the shape of a miniature house (Fig. 3a).
- **Type B**: a sarcophagus-like rectangular block that extends over a limited depth range, over which it remains a well defined isolated feature (Fig. 3b).
- **Type C**: a cut into the limestone geology, which manifests itself as a persistent low-reflection GPR anomaly that does not vary with depth, while the geological features around it do (Fig. 3c). There is no GPR evidence for a feature inside the cut and it could hence be either an excavation for a simple grave or a rectangular quarry pit for the extraction of building materials.

This classification, derived from GPR data alone, will in a later stage be compared to historic archaeological excavation records to test its wider applicability. So far there is no evidence for a consistent correlation between the location of the burials and their type, although they appear to be mainly aligned with the contour lines of the current surface topography. This could indicate that the

orientation relative to Pessinus is of lesser importance. It will be essential to extend the survey area to test this archaeological hypothesis.

The manipulation of archaeological geophysics data after their collection can be subdivided into two phases, namely data improvement and data processing (Schmidt & Ernenwein 2011). Data improvement rectifies errors in the data that were introduced during the measurement process, either caused by instrument imperfections or operator errors. The main purpose of subsequent data processing is to enhance anomalies that are only noticed by an experienced operator on a high-contrast computer monitor, or to reshape anomalies so that the causative features can be interpreted more easily (e.g. by applying reduction-to-the-pole to magnetometer data). Parameters of data processing functions are therefore adjusted in such a way that the results 'look good' and most closely resemble even faint anomalies that were visible before. The procedure described here for creating anomaly-polygons enhances and reshapes anomalies that may already be perceived by a well-trained human interpreter and the proposed procedure is hence considered to be a data processing workflow. In addition, it also serves to extract feature shapes from the anomalies and is hence a suitable precursor for subsequent data interpretation.

The main advantage of the proposed framework is the very small number of parameters that are needed to specify thresholds and buffers, and reasonable estimates can already be determined from site conditions and survey parameters (e.g. minimum expected anomaly sizes, traverse spacing). As with all data processing, the parameters ultimately have to be chosen based on the experience of the operator until a desired output is achieved. However, the values chosen from one single timeslice were shown to create a convincing data representation for all depth ranges in the form of 'pancake slices' (Fig. 2), which indicates that they are meaningful for the whole data set. The more traditional rendering of the data volume as iso-surfaces would have been problematic since the site's varied features created anomalies with different reflectivity values. Although the presented vector processing also starts with a threshold (i.e. an iso-value), the subsequent neighbourhood processing adjusts the shapes considerably so that they represent the individual anomalies well. Another processing approach that generates generalised representations of three-dimensional GPR data was developed by Leckebusch *et al.* (2008), fitting boxes around GPR anomalies using the three-dimensional data gradient. However, the generalisation created by this method is considerable and may work best with very rectangular anomalies. A manual approach that uses interpretation polygons from only few timeslices that are then extruded up- and downwards (Verdonck *et al.* 2012) generates results that are similar to the presented vector processing method, but is considerably more time consuming.

One of the main application areas of the proposed method is the isolation of anomalies in noisy data. Although inspired by GPR data (see above) the method is applicable to all geophysical data. Fig 4a shows noisy magnetometer data from Pessinus, collected in an area close to the main temple precinct using a FM36 fluxgate gradiometer bidirectionally along transects with a separation of 0.5 m. Due to the poor magnetic contrast of the soil the anomalies are very weak (± 2 nT) and close to the noise level of instrument and soil variations. Nevertheless the processing workflow successfully joined together isolated high readings (Fig. 4b) that were perceived by an experienced operator as belonging to the same anomalies. The random layout of these anomalies makes an archaeological interpretation unlikely and they are hence attributed to soil variations.

The proposed processing scheme is currently defined in two dimensions, as the three-dimensional neighbourhood analysis is not yet as well developed, even in PostGIS. However this situation is likely to improve in the near future so that the scheme can then be expanded. The time savings when analysing vector data instead of three-dimensional voxel data will be even more significant than in two-dimensions. In the meantime the two-dimensional anomaly polygons can be analysed in terms of overlap between one layer (i.e. timeslice) and the next so that sufficiently overlapping two-dimensional anomalies (e.g. 80%) are assigned to the same three-dimensional anomaly. This is akin to the visual inspection of anomalies from neighbouring timeslices that is being used by several authors (e.g. Tolnai 2012).

Acknowledgements

Useful discussions with Neil Linford, James Lyall and Benno Zickgraf at the NSGG day-conference (London, December 2012) helped to highlight aspects of this processing scheme that were of particular interest to users and these were therefore discussed in more detail in this paper.

Bibliography

Aspinall A, Gaffney CF, Schmidt A. 2008. *Magnetometry for Archaeologists* Geophysical methods for archaeology. Lanham: AltaMira Press.

Aspinall A, Haigh JGB. 1988. A review of techniques for the graphical display of geophysical data. In S. P. Q. Rahtz (ed.) *Computer Applications and Quantitative Methods in Archaeology 1988*: 295-307 BAR International Series 446(ii). Oxford: Tempus Reparatum.

Bevan BW. 2000. An early geophysical survey at Williamsburg, USA. *Archaeological Prospection* **7**(1): 51-58. DOI: 10.1002/(SICI)1099-0763(200001/03)7:1.

Brackman P, De Dapper M, Devreker J, Vermeulen F. 1995. The Use of Geomorphology, Remote Sensing and GIS Techniques for Geoarchaeological Purposes in the Pessinus Area, Central Anatolia (Turkey). *Natuurwetenschappelijk Tijdschrift* **75**(1): 3-34.

Clark AJ. 1968. A square array for resistivity surveying. *Prospezioni Archeologiche* **3**: 111–114.

Gaffney C, Gater J. 2003. *Revealing the Buried Past: Geophysics for Archaeologists*. Strout: Tempus Publishing Ltd.

Le Borgne E. 1955. Susceptibilité magnétique anormale du sol superficiel. *Annales de Géophysique* **11**: 399-419.

Leckebusch J, Weibel A, Bühler F. 2008. Semi-automatic feature extraction from GPR data. *Near Surface Geophysics* **6**(2): 75–84. DOI: 10.3997/1873-0604.2007033

Li X, Götze H-J. 1999. Comparison of some gridding methods. *The Leading Edge* **18**(8): 898-900. DOI: 10.1190/1.1438401.

Ma G, Li L. 2012. Edge detection in potential fields with the normalized total horizontal derivative. *Computers & Geosciences* **41**: 83-87. DOI: 10.1016/j.cageo.2011.08.016.

Nabighian MN. 1972. The analytical signal of two-dimensional magnetic bodies with polygonal cross-section: its properties and the use for automated anomaly interpretation. *Geophysics* **37**(3): 507-517. DOI: 10.1190/1.1440276

Obe R, Hsu L. 2011. *PostGIS in action*. Greenwich, Conn., London: Manning.

Reid AB, Allsop JM, Grasner H, Somerton AJMAIW. 1990. Magnetic interpretation in three dimensions using Euler deconvolution. *Geophysics* **55**(1): 80-91.

Schmidt A. 2003. Remote Sensing and Geophysical Prospection. *Internet Archaeology* 15. http://intarch.ac.uk/journal/issue15/schmidt_index.html .

Schmidt A, Ernenwein E. 2011. Guide to Good Practice: Geophysical Data in Archaeology. *Archaeology Data Service / Digital Antiquity Guides to Good Practice*. http://guides.archaeologydataservice.ac.uk/g2gp/Geophysics_Toc .

Schmidt A, Parkyn A, Tsetskhladze G. 2011. Pessinus: A City without Contrasts? Geophysical Challenges from Anatolian Soils. In M. G. Drahor and M. A. Berge (eds) *Extended abstracts, Archaeological Prospection, 9th International Conference on Archaeological Prospection, September 19-24, 2011, Izmir - Turkey*: 78-80. Istanbul: Archaeology and Art Publications.

Schmidt A, Tsetskhladze G. 2012. *They may all be dead but they still ain't equal ...* Recent Work in Archaeological Geophysics: 16-17. London, UK: Near-surface Geophysics Group.

Scollar I, Tabbagh A, Hesse A, Herzog I. 1990. *Archaeological Prospecting and Remote Sensing* Topics in Remote Sensing 2. Cambridge: Cambridge University Press.

Stampolidis A, Tsokas GN. 2012. Use of Edge Delineating Methods in Interpreting Magnetic Archaeological Prospection Data. *Archaeological Prospection* **19**(2): 123-140. DOI: 10.1002/arp.1424.

Tabbagh A, Desvignes G, Dabas M. 1997. Processing Z gradiometer magnetic data using linear transforms and analytical signal. *Archaeological Prospection* **4**(1): 1-13.

Tolnai K. 2012. *GPR in-depth - a chronological interpretation of high resolution GPR data from Visegrad, Sibrik, Hungary*. Recent Work in Archaeological Geophysics: 39-41. London, UK: Near-surface Geophysics Group.

Verdonck L, Vermeulen F, Corsi C, Docter R. 2012. Ground-penetrating radar survey at the Roman town of Mariana (Corsica), complemented with fluxgate gradiometer data and old and recent excavation results. *Near Surface Geophysics* **10**(1): 35-45. DOI: 10.3997/1873-0604.2011034.

Figures

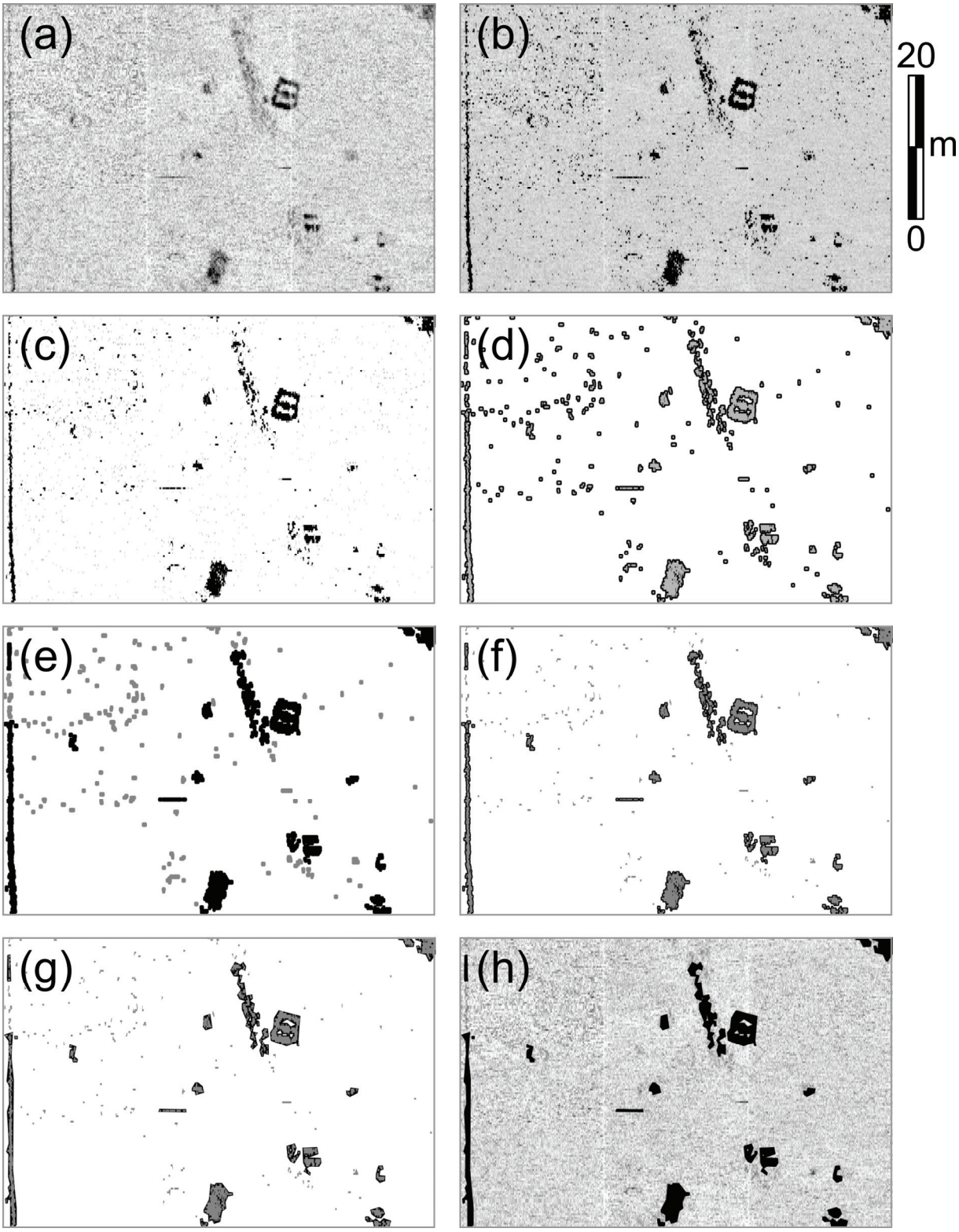

Figure 1: GPR timeslice from a necropolis in Pessinus with several stages of data processing. The respective processing results are shown in black, with polygons from a previous step contrasted in grey: (a) timeslice at 1.5 m depth, black represents high reflectivity; (b) polygons of clipped data overlaid onto greyscale; (c) size threshold 1 applied; (d) neighbourhood merge through buffering; (e) size threshold 2 applied; (f) shrink with fringe, overlaid with threshold 1 polygons; (g) generalised polygons, overlaid with threshold 1 polygons; (h) generalised polygons overlaid onto greyscale.

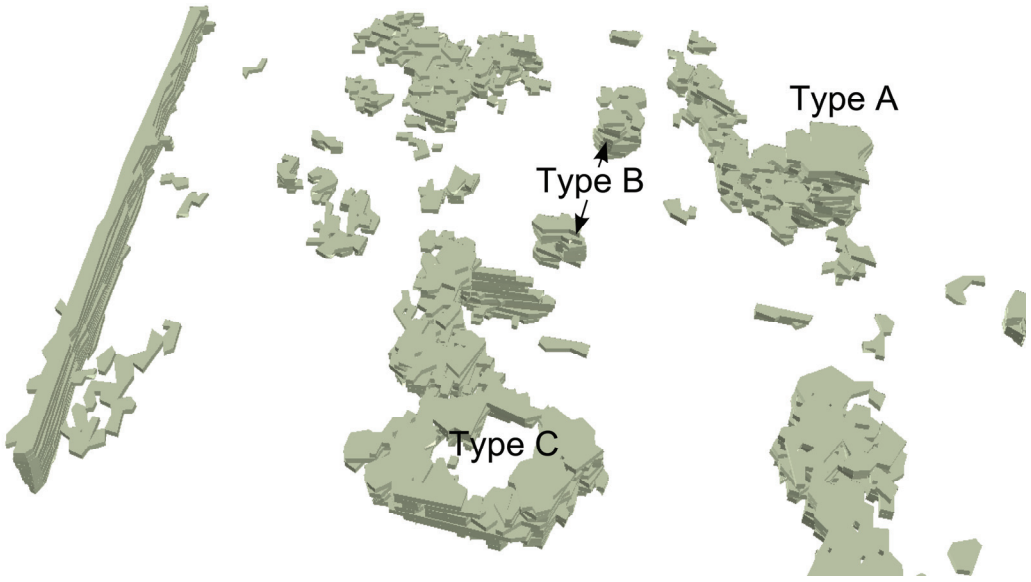

Figure 2: Three-dimensional visualisation of data from a depth of 0.8 m downwards, using the anomaly-polygons to generate 'pancake slices'. See text for explanation of anomaly types A-C.

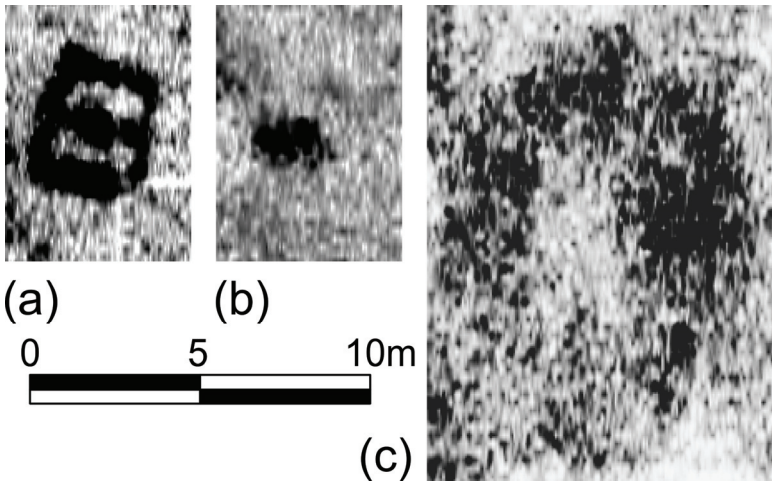

Figure 3: Three types of anomalies identified in the GPR data (black represents high reflectivity): (a) elaborate grave monument (1.5 m depth); (b) sarcophagus-like rectangular block (1.1 m depth); (c) cut into the limestone geology (0.7 m depth).

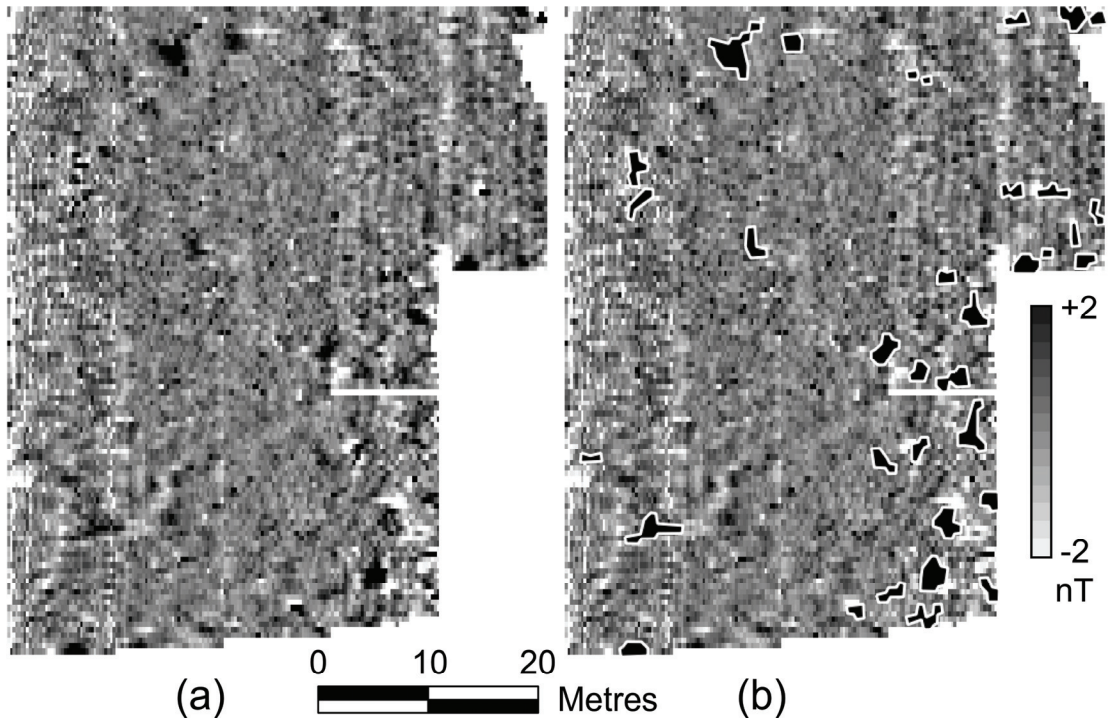

Figure 4: Magnetometer data from Pessinus, close to the temple precinct: (a) measurements showing high noise levels; (b) overlay of processed anomalies, automatically created by merging several small areas of higher readings together. The processing parameters were: clip at 0.6 standard deviations, threshold 1 was 0.2 m^2, threshold 2 was 3.0 m^2, neighbourhood merge 0.4 m, fringe of 0.2 m, Douglas-Peucker parameter of 0.2.