Introduction

Archaeological surface survey is one of the principal methods of archaeological prospection and research on a regional level, but despite the long development of this method its capabilities and limitations are still often not well understood, especially its effectiveness as a discovery method (see for e.g. Ammerman 1981: 81-82; Wandsnider and Camilli 1992; Burger et al. 2004; Burger et al. 2008: 216-218, 228). This is a problem in any archaeological landscape and settlement research but it comes to the forefront especially when the survey is used as a tool for evaluating the presence of archaeological resources before large-scale development projects. In such cases survey results may be the basis for decision making about the presence or absence of archaeological resources, needs for their protection and management, needs for excavation and in this case also for determining the size, time and cost of excavations (see Hey and Lacey 2001; Hey 2006; Medlycott 2017). In such cases, it is crucial that survey procedures are precise, reliable and accurate and that decisions based on their results are valid (see Banning et al. 2017: 468). But standard survey procedures generally do not allow for a realistic evaluation of the precision, reliability and
Multiple levels determining visibility

Visibility is generally accepted as an important aspect of archaeological surveys, which must be considered in the analysis and interpretation of survey results. In most surveys, however, dealing with visibility is mostly limited to documenting the amount of exposed ground surface or amount of vegetation cover in the field during survey and then using this information in the analysis to correct the raw data (e.g. Cherry et al. 1991: 27-28; Gaffney et al. 1991: 61; Terrenato 1996: 223; Terrenato 2000: 60, 66). However, the problem of visibility is much more complex and should be considered both in survey design and in the analysis of survey results on multiple levels:

1. Visibility determined by geomorphic, pedogenic and other post-depositional formation processes.
2. Visibility determined by the nature of the archaeological record.
3. Visibility determined by techniques and strategies of the survey method.
4. Visibility determined by surface and other environmental conditions during the survey.
5. Visibility determined by the human factor.

Within each of these levels, a multitude of factors operate to influence surface visibility and detectability of archaeological record on the surface and consequently survey precision, reliability and accuracy by introducing biases into survey results. In the following text, each of these levels is defined and briefly explored by using information from a selection of relevant literature dealing with factors operating on the levels as they are defined in this paper.

For each level, an example of a study presenting as well as some methodological solutions, proposed so far by different authors, for dealing with problems these levels present for survey archaeology.

1 The five levels have already been briefly defined and touched upon in Gruškovnjak 2017a. In the present paper considerations on some of these levels have been expanded, especially in the case of the 1st and 5th level.
surface materials. That is why soil scientists often realize that the process of burial alone simultaneously erodes the upper soil horizons, often leaving only B horizon as the only proof of soil surface existence prior to burial (Foley 1981: 170; Butzer 1982: 100; Schaetzl and Anderson 2005: 50). Thus, on a buried or exposed land surfaces, movement and transport of archaeological artefacts on the surface must be taken into account, as well as their incorporation into and movement with the burying or eroding sediment (e.g. Barton et al. 2002: 169-170). In any landscape we are surveying we must, therefore, be aware of what kind of geomorphological surfaces we are dealing with and what kind of debris flux, involving erosional and depositional processes, operated on them through the evolution of the landscape, especially during the Pleistocene and Holocene, up to the present day. Geomorphic surfaces can be erosional, constructive (aggrading) or a combination of both (see Schaetzl and Anderson 2005: 467, 471; also see Johnson 1993: 76; Van Nest 2002: 57, fig. 2).

Pedoturbation is another crucial, but rarely considered, factor affecting surface visibility as well as the nature of the stratified subsurface archaeological record. It is a ubiquitous and continuous process synonymous with soil mixing. From an archaeological perspective, pedoturbation is very important because it can result in burial of archaeological artefacts and architectural features even without the additions of mineral matter, as well as in sorting, mixing and moving of the larger size fractions of soil or sediment, like gravel, stone, and crucially archaeological artefacts. Bioturbation, i.e. pedoturbation by soil fauna (faunalturbation) and plants (floralturbation), can, for example, cause surface aggradation and thus burial of archaeological record through surface mounding of fine soil fractions by animals. It can also cause a downward movement of coarse fragments via loosening and within-horizon jostling and undermining (by root growth and decay, and animal burrowing). This eventually causes settling of the coarse fraction at the maximum depth of burrowing and formation of so-called stone-lines and artefact-lines, or layers, overlain by a biomantle, i.e. a layer of material sorted and brought to the surface by animals. With enough time elapsed the entire artefact size range archaeologists typically recover and study will be concentrated into a subsurface stone layer, only the smallest debris remaining mixed in the upper fine textured biomantle. Besides rendering archaeological record invisible on the surface through pedoturbation processes that can move coarse fragments and representative sources on the topic see Schaetzl and Anderson 2005: tab. 10.2.

In such a case the principle of stratigraphic superposition does not apply as the stone line and biomantle are synchronous. Also a false impression of paleo land surface is created in this way, and in some cases the amount of displacement can be sufficient to alter stratigraphic relationships and may potentially cause significant errors in dating of archaeological features. However, if not enough time has elapsed for artefacts of different periods to reach the maximum depth of bioturbation, the relative stratigraphic relationships are preserved (Atkinson 1957: 222-224, 226; Johnson 1989: 383; Balek 2002: 46-48; Van Nest 2002: 77; Schaetzl and Anderson 2005: 242).

**FIGURE 1.** Scheme of post-depositional natural and cultural formation processes (Source: altered after Foley 1981b: fig. 6.5; also see Gruškovnjak 2017a: fig. 4).
burial, the faunal activity may also seriously affect the reliability of surface-collection data by causing a disproportionate amount of small-sized artefacts (artefact size dependent on animal size) to be present on the surface, while large-sized artefacts are buried. Similar sites with differential amounts of faunal activity and different animal species involved in it may thus display very different contents in their surface assemblages. Also, because the biomantle slowly moves downslope (rates of mass transport vary spatially and temporally with local conditions), it is both sedentary in the short term and transported in the long term. In addition the stone-line often acts as a lateral aquifer or lateral subsurface throughflow zone on sloping surfaces and is thus subjected to aquaturbation, which mainly affects the fine soil fraction, but may potentially cause movement, abrasion, degradation etc. of artefacts in the stone-line (Atkinson 1957; Erlando 1984; Bocek 1986; Johnson 1989; Johnson 1993: 72, 74-76; Balek 2002; Johnson 2002; Peacock and Fant 2002; Van Nest 2002: 57, 59, 62-63, 77-79, figs. 2-4; Schaezl and Anderson 2005: 239-255, 543-546, tab.10.1-2; Araujo 2013; Ozán 2017: 2-4).
Other forms of bioturbation can also bring larger clasts, including archaeological artefacts, to the surface. Tree uprooting, for example, is a major cause for this and as the fine materials are washed away, coarse fragments can become concentrated as a surface lag. Tree uprooting can move particles in any direction, bringing buried materials up as well as introducing surface materials to lower levels of the soil while soil horizons are disrupted and mixed together. Repeated uprooting results in so-called “natural ploughing” or “tree ploughing” and through long periods of time this localised and spatially discontinuous process may affect very large portions of the landscape, often creating the so-called cradle-and-knoll or pit-and-mound micro-topography. Deep burrowers may also bring deeply buried materials into the near-surface environment, and larger burrowers such as rodents may cause considerable lateral movements of artefacts. Larger clasts can also be brought to the surface and sorted via cryoturbation (freeze-thaw processes) and argilliturbation (Bocek 1986; Johnson 1993: 74; Balek 2002: 42-48; Johnson 2002: 8-9; Peacock and Fant 2002: 91–92; Van Nest 2002: 57; Schaeztl and Anderson 2005: 243-244, 259-262, 501; Pawlik 2013).

Thus, it must be realized that biomechanical processes of soil formation are inherent to all soils and that archaeological contexts, be it in the bottomlands or uplands, have been universally altered by biomechanical processes. Case-by-case assessments of the formation processes affecting sites are required to discern which patterns can be attributed to human activities and witch to other processes, though this task is inescapably complicated by equifinality (Van Nest 2002: 78; Burger et al. 2008: 205; Ozán 2017: 12-13).

The above-mentioned processes are only a part of the dynamic denudation model (Fig. 2) developed by Donald Johnson (1993; 2002), which must be considered in any archaeological landscape survey, analysis and interpretation. As summarized by Johnson (1993: 76-77): “Dynamic denudation theory provides a rational explanation for the evolution of tropical, subtropical and temperate landscapes with three-tiered soils that may or may not bear stone-lines. The dynamic processes and conditions are driven by gravity, water and biotic agents. The framework is a synthesis of tripleplanation, soil evolution, biomantle, soil thickness, etchplanation, and mass transport theory fused with /.../ A, E, B, C soil horizon designations, and hydrological principles of lateral throughflow. /.../. It explains soil-slope systems with or without stone-lines in a variety of landscapes underlain by variable rock types, including those that are saprolitized. Dynamic denudation principles should have wide applications in archaeology, ecology, forestry, geomorphology, mineral exploration, and pedology.”

A case study of archaeological record’s surface visibility considering dynamic denudation principles is provided in western Illinois, U.S.A., by J. Van Nest (2002). Among archaeologists working in this area the prevailing view has been that before a late prehistoric establishment of bison into the region, Illinois prairies were mostly uninhabited. Seemingly confirming this view was the scarcity of surface archaeological materials away from the river valleys. Surface surveys of upland terrain in the region discovered only ribbon-like distributions of sites on the slope shoulders and upper backslopes along headwater valleys. For a long time, it has been presumed that during the Holocene the uplands underwent massive hillslope erosion, causing nearly all archaeological materials to become lag deposits, which are now incorporated into historic plough zones and are thus all visible on the surface. However, Van Nest observed that erosion in western Illinois uplands is spatially restricted and that no massive hillslope erosion occurred during the Holocene. Large areas of this landscape retain soil profiles with biomantles and buried archaeological remains. She examined the interaction of biomantle–stone zone formation, soil creep on a hillslope transect (catena) and processes related to vegetation type (forest vs. prairie) (Fig. 3). At the hilltops where soil loss by creep is effectively zero, large artefacts have been readily buried by soil fauna to depths below the plough zone. On the contrary, backslope positions are so steep that the rate of soil creep exceeds the rate of artefact burial by soil fauna and thus they remain at the surface. At intermediate slope steepness positions on slope shoulders, burial by soil fauna is rapid enough for the artefacts to start sinking, but erosion is also rapid enough to allow only shallow burial. Thus, artefacts are protected from surface disturbances but may be incorporated into the historic plough zone. At the bases of slopes, artefacts may be buried by sediment washed from upslope, or by alluvium (Fig. 3A). Besides the position on the hillslope transect, vegetation also proved to be an important factor in soil development and artefact burial. Tree ploughing in forests can overwhelm the downward movement of artefacts by soil fauna and in addition, the organic-rich layer with soil fauna activity of many forest soils is so thin that it is now entirely incorporated into historic plough zones. Contrary, in prairie soils the organic-rich A horizons extend to considerable depths and artefacts become buried in biomantles below the depth of ploughing (Fig. 3B). There is probably also a link between prehistoric Indian

—

8 For historical background and comments see Schaeztl and Anderson 2005: 324-338, 537-546.
vegetation burning practices and the position of the forest-prairie boundary on upper slopes, and consequently the distribution of surface and buried habitation scatters in this landscape. The process of burial by soil fauna has been documented for Archaic period sites (>3500 B.P.), while not enough time has elapsed for Woodland period (<2500 B.P.) artefacts to become buried to sub-plough zone depths. Thus, the once prevailing assumption that upper slopes along valleys are an eroded, degraded landscape where all cultural remains occur at the surface, in the plough zone, or occasionally in pits reaching below the plough zone is now obsolete. The biomantle–stone zone hypothesis now predicts that almost all pre-Woodland sites occurring across the upland prairie regions will be buried and consequently not detectable by surface surveys (Van Nest 2002: 79-83, figs. 8-9; Schaeztl and Anderson 2005: 254).

This case study clearly demonstrates that consideration of pedogenic and geomorphic processes along a catena is crucial for decision making about which prospecting method is appropriate for a particular surface while any analysis and interpretation of survey results as well as settlement and land use patterns, which does not consider these, cannot be regarded as valid (see e.g. Bettis and Mandel 2002: 141-142, 149-152; Peacock and Fant 2002: 92, 95; Burger et al. 2008: 205-211; Banning et al. 2017: 469; Ozán 2017). For example, the surface survey will be pointless in a deeply buried landscape where the subsurface survey is called for instead, and no surface finds in such a case do not mean that there is no subsurface archaeological record present in the area (see e.g. Brookes et al. 1982; Stafford and Creasman 2002). Only archaeological record affected by exposure or disturbance processes might be expected to be visible on the surface, while buried sites will not be. Finds might also be dispersed or even drastically moved by post-depositional processes and the location of finds in a colluvium, for example, may be far removed from the location of their primary deposition and incorrectly interpreted if colluviation process is not identified (e.g. Foley 1981: 166-174; Butzer 1982: 98, 100-117; Ebert et al. 1987: 165-166; Burger et al. 2008: 221-227).

Often overlooked, however, is that the patterning of these natural processes, which affect the visibility, preservation and integrity of the archaeological record, are of a very local nature. They are controlled by local microtopography and other small-scale factors and are thus often of an even smaller scale than might be assumed to fall within the boundaries of culturally-caused clusters of artefacts, or sites. Consequently the existing regional scale maps and data on these processes in any given area are usually too general and not precise enough for archaeological purposes which is why all survey projects should incorporate localized small-scale soil geomorphological mapping and other geoarchaeological methods into the initial (and if necessary subsequent) phases of their survey design (Ebert et al. 1987: 173; Stafford and Creasman 2002: 120-121; Johnson 2002: 11; Schaeztl and Anderson 2005: 501-506; Ozán 2017: 2, 7, 12, 13).

In the context of burial an exposure processes it must be mentioned that globally speaking in temperate zones, most of the surfaces are generally affected by constant aggradation and consequently most of the archaeological record is buried, thus displaying very low artefact densities and poor visibility on the surface, while in arid zones most of it is exposed, thus displaying much higher artefact densities (lag concentrate) and better visibility on the surface (Bintliff and Snodgrass 1988: 508-512, fig. 2). That is why the surface survey in temperate zones is almost exclusively applied to ploughed surfaces, where subsurface materials are potentially being brought to the surface. Mechanical ploughing as a form of pedoturbation or anthropoturbation can be considered as a large-scale formation process unique to modern surfaces. Though more severe its effects are similar as in the case of many other natural and anthropogenic pedoturbation processes. Also similar are factors affecting it, but their effects are drastically increased. When studying archaeological record in the plough-zone we must consider at least (1) lateral displacement and its effect on spatial patterning; (2) vertical displacement or the circulation of finds in the plough-zone and the functioning of the surface as a sampling process determining the relation between the whole population of finds in the plough-zone and artefacts brought to the surface; (3) changes in state of preservation of various classes of material in specific conditions; (4) duration, direction and depth of ploughing as well as ploughing equipment; and (5) local characteristics of soil, relief and post-depositional burial of the archaeological record. (see Lewarch 1979; Jermann 1981; Lewarch and O’Brien 1981a: 308; Lewarch and O’Brien 1981b; Ammerman 1985: 34-35; Odell and Cowan 1987; Reynolds 1988; Boismier 1989; Yorston et al. 1990; Dunell and Simek 1995; Schaeztl and Anderson 2005: 292-293). Various experimental studies have shown that what we see on the ploughed surface is only the tip of the iceberg and a random one at that (Ammerman 1985: 37, 39).
FIGURE 3. (A) A slope transect model showing where buried and surface-exposed archaeological remains are expected in the western Illinois uplands. (B) Influence of soil type on the expected depth of stone-lines in fully formed faunal mantles of western Illinois (Source: Van Nest 2002: figs. 8-9).

2nd level determining visibility: nature of the archaeological record

On the second level, visibility is determined by the nature of the archaeological record itself. Among key factors here, are obtrusiveness, clustering and density of artefacts (Schiffer et al. 1978; Wandsnider and Camilli 1992; Banning et al. 2017: 472-473).

Obtrusiveness of artefacts is conditioned especially by their size, shape and colour as well as by the relationship of these properties to the natural material of the soil surface. More the archaeological material differs from the natural background noise of the soil matrix more obtrusive it is. This means that the effect of artefact properties is specific to specific circumstances and must always be evaluated in light of local conditions. Generally as the
size of the artefact increases so too does its obtrusiveness, but there are variations in this relation conditioned by local circumstances. For example, if pebble stones of intermediate size class are predominating, intermediately sized artefacts might be less obtrusive than small sized artefacts. Also, the more the artefact’s shape differs from the shapes of the natural material in the background, more obtrusive it will be and so in some circumstances small unnaturally shaped artefacts will be more obtrusive than larger more naturally shaped artefacts. Similarly, in soils of different colours and in the presence of natural stones of different colours differently coloured artefacts will contrast the natural background differently and will thus have different obtrusiveness. Thus, in different natural backgrounds artefacts of different sizes, shapes and colours will have different levels of obtrusiveness. That is why the natural matrix and any changes or variations in it during the survey should be described in detail in order to evaluate the effect of artefact obtrusiveness and biases it incorporates into survey results (Wandsnider and Camilli 1992: 174, 176, 177-179; see also Banning et al. 2006: 726, 732; Banning et al. 2010). Artefact clustering and density also affect the recovery rate during the survey. More clustered than isolated artefacts will be discovered and the higher the density of the cluster the higher percentage of artefacts present in it will be discovered. The interaction between clustering, density and obtrusiveness is very important but it is not simple and straightforward. Obtrusiveness will especially affect the differential artefact recovery at low densities while its effect at higher densities will be lower. In this case, the size of the artefacts plays the most important role for large artefacts are quite consistently recovered at high as well as low densities while small artefacts are mostly recovered in cases of higher densities (Wandsnider and Camilli 1992: 174, 180-182).

Besides the effects post-depositional processes (1st level determining visibility) have on artefact surface density, artefact density is also conditioned with the duration of the past occupation or activity, the intensity of activities involving discard behaviour, focus of such activities to a specific location and integration of durable cultural materials into these activities, such as stone and pottery, while discovery of other types of activities is severely limited. Similar factors also apply to periods for we are more likely to discover remains of periods characterised by locally concentrated long-lasting activities or occupation, higher population densities and production of durable cultural materials, which are more resistant to destruction. Regarding preservation, time or progressive degradation and destruction of materials also plays a role, as well as types of soil, which have different effects on the degradation and destruction of cultural materials. However, visibility is also conditioned by the resolution of our dating or relative archaeological visibility, meaning that periods with highly diagnostic material will have higher visibility in surface assemblages than periods with less diagnostic material, and if finds from a particular period are not recognised it will stay invisible even though present in the collected material (see Hope-Simpson 1984: 116; Bintliff and Snodgrass 1985: 138; Gallant 1986: 415; Schofield 1989: 460-462, 466-468; Barker 1996: 167; Bintliff 2000: 205-206, 212-213; Banning 2002: 226; Novaković 2003: 145; Hey 2006; Vermeulen and Mlekuž 2012: 209).

For dealing with problems of artefact obtrusiveness, clustering and density, incorporation of seeding experiments into survey design has been proposed as a way to enable the evaluation of biases incorporated into survey results by these properties of the archaeological record (Wandsnider and Camilli 1992: 183, 185). Such an experiment would be done on a smaller plot or plots of the survey area, representative of the local conditions, by seeding a known quantity of non-archaeological artefacts and mapping their distribution in order to control the most relevant attributes such as colour, shape and size of artefacts as well as different degrees of their clustering and density. These experimental plots would be surveyed using the same procedure as throughout the survey area and the results would provide a quantitative measure for evaluating the effects of these characteristics on the recovery of archaeological artefacts in the specific circumstances of the survey area and with the specific procedure used in the survey project (e.g. Wandsnider and Camilli 1992: 173-176; also see Banning et al. 2017: 475-476).

An example is provided by the Seedskadee (Green River, Wyoming, U.S.A.) seeding experiment, presented by L. Wandsnoder and E. L. Camilli (1992), in which an intensive (5 m transect interval) distributional survey technique was used. With the distributional technique, discovery is done in two phases, first systematically by the discovery crew, and then unsystematically by the encoding crew. Regarding shape and colour of seeded artefacts (washer and nails of black and buff colour) more “unnaturally” shaped washers (71%) were recovered and slightly lower number of larger but less unnaturally shaped (stick-like) nails (61%), while more black artefacts (70%), which were more contrasted with the soil surface, were recovered than buff artefacts (62%). Regarding isolated vs. clustered seeded artefacts the discovery crew recovered 10% of all isolated and 69% of all clustered artefacts, while the encoding crew recovered additional 6% of all isolated and 12% of all clustered artefacts. In total,
dramatically more clustered (82%) than isolated (16%) artefacts were recovered, with the inspection of 20–40% of the ground surface (5 m transect intervals and inspection of 1–2 m transect strips) (Fig. 4A). Also, as the cluster density increased, so did the percentage of total seeded artefacts discovered in the cluster (Fig. 4B). Furthermore, of all the isolated seeded artefacts discovered (systematically and unsystematically) the discovery crew recovered 62.5% and the encoding crew 37.5% as opposed to clustered artefacts, the majority of which were recovered by the discovery crew (85%) as compared with the encoding crew (15%) (Fig. 4A). This observation has important implications especially regarding the discovery rate of clustered and unclustered surface distributions by standard survey procedures with only one (systematic) discovery phase (Wandsnider and Camilli 1992: 174, fig. 2, tab. 1).

Because of two discovery phases, the distributional survey technique also allows some evaluation of recovery rates in the case of archaeological distributions, where total quantity and distributional pattern of the sampling universe are not known (Wandsnider and Camilli 1992: 174, fig. 2, tab. 1).
With standard survey procedures, where discovery is done only in one phase, this is not possible and only seedling experiments incorporated into survey design would allow evaluation of artefact obtrusiveness, clustering and density. In the case of standard surveys with transects in 15 m intervals, a lot of the same high-density clusters might be discovered as with the more intensive distributional survey. However, because only 6–13% of the ground surface is inspected, presuming that only 1–2 m away from the transect lines are inspected, this allows for only 6–13% of surface artefacts to be recovered. But because of the biases incorporated by all levels of surface survey procedure is neither an off-site survey, neither a full/total-coverage survey. The population of isolated artefacts and rarely features, and therefore, by detecting artefacts alone, however, cannot be a sufficient criterion for discovering the presence and for determining characteristics of the archaeological record, considering that it constitutes of (1) artefacts, (2) features, (3) anthropogenic soil horizons, (4) organic materials, and (5) chemical and geophysical anomalies. In different types of sites or remains, these constituents are present in different ratios (McManamon 1984: 226-228). With the surface survey, we are detecting only artefacts and rarely features, and therefore, by detecting only one or rarely two types of constituents, the surface survey is an inherently biased discovery method. That is why the use of multiple survey methods, each detecting a different kind of archaeological remains or constituents of archaeological record are called for, at least if our intention is to discover various types of archaeological remains in the landscape and not only large high-density distributions of artefacts. It is not possible to blindly rely on the presuppositions that high densities of surface artefacts correspond to high densities of subsurface artefacts and features as well as areas of most intensive past human activities. Many studies show that surface distributions do not or only partly reflect subsurface distributions, that important sites with large number of features may contain very low numbers of subsurface artefacts and even less surface artefacts, as well as that many sites manifest themselves on the surface with lower densities than off-site distributions and therefore cannot be quantitatively recognized (Shott 1987: 361-362; Schofield 1989; Bankoff et al. 1989: 70-72; Bintliff 1996: 252; Bintliff 2000: 206-209, 212; Fentress 2000: 48-49; Hey and Lacey 2001; Medlycott 2017).

3rd level determining visibility: techniques and strategies of the survey method

Visibility is also determined by the techniques and strategies of the survey method. Here term method is understood as the basic method such as surface survey, subsurface survey, geophysical survey etc., while the technique determines the basic procedures used such as systematic fieldwalking or surface collection (e.g. Cherry et al. 1991), distributional survey (e.g. Ebert et al. 1987), point-sampling (e.g. van de Velde 2001), probability survey (e.g. Plog 1976) etc. in the case of surface survey method. Strategy refers to survey intensity and the shape, size and spatial layout of the survey grid and/or collection units used (e.g. transects, quadrats), or in the case of probability survey to a simple random, stratified random, systematic or stratified systematic unaligned sampling strategy used.

Generally, the more intensive the survey and slower the pace, more surface material will be discovered (e.g. Wandsnider and Camilli 1992: 177, 183, fig. 3; Banning et al. 2006; 2010; 2011; but see below). For example,
When using transects recovery will be primarily determined by transect intervals, which will determine the size of phaenomena the survey is capable of discovering. Phaenomena smaller than the transect interval will thus be discovered only due to coincidence (e.g., Cherry et al. 1991: 18-20). The capability of different shapes and sizes of aggregate units and their layout to discover sites according to their size and artefact density can be calculated by mathematical formulas based on search theory or recovery theory (e.g., Miller 1989; Sundstorm 1993). However, these must generally presume that if the aggregate unit intersects the area of a certain phenomenon it will be discovered, though the reality is not as simple. The chance of discovery is also determined by all other factors of all of the five levels determining visibility and all of these cannot be taken into account in such calculations, which is why they are only capable of evaluating the effectiveness of discovery in ideal circumstances. Besides, such calculations are mostly limited to evaluation of discovery capability for discrete distributions, i.e., sites, and not off-site distributions.

Because of the great variability in the surface visibility of archaeological record in the landscape, each survey area should be stratified into zones according to visibility (Banning et al. 2006: 740; Banning et al. 2017: 469). That is according to visibility as determined by the first and fourth level, as well as second level in case its properties are already known to a certain degree. For each zone, method, technique and strategy of a survey that will perform best in the given conditions should be determined, while most surveys use a standardized procedure throughout the survey area despite the differences in visibility. This difference is related to two different approaches to survey which are connected with the problem of data comparability, views on which differ among survey archaeologists. The core of this problem is in the question whether survey results of a certain area are comparable if the same standardized procedure was used throughout, or are they comparable if the same chance of discovery was assured by using different procedures according to different visibility conditions? Here, the latter approach is being emphasised.

Furthermore, the capability of addressing the question “What did we miss?” is generally absent in most surveys and therefore their effectiveness cannot be realistically evaluated (Burger et al. 2004: 411). For tackling with this problem to a certain degree, the proposition of using control seeding experiments has already been mentioned, while intensive resurveys of certain control areas through property-based investigations would be yet another or additional option which would allow for comparison of recovery rates with the standard procedure used in the survey project. Regarding the definition of a property-based approach, we may follow Burger et al. (2004) who differentiate architectural surface survey procedures according to their numerous goals. Discovery-based surveys identify geographical aspects of the surface record by locating and describing clusters of artefacts, while property-based approach focuses on evaluating the accuracy of technique and strategy used as well as on formational aspects of the regional record. A property-based approach will, therefore, be especially valuable in cases of geomorphologically active and topographically diverse landscapes (Ibid.: 410; also see Banning et al. 2017: 474-476).

An example of such a property-based investigation is provided by Burger et al. (2004) with the experiments performed in the Oglala National Grassland (Nebraska, U.S.A.), where the multi-scale Modified-Whittaker sampling plot was used and surveyed at different intensities. The main technique used was a distributional survey in 70 cm intervals, followed by a resurvey of smaller control areas with crawl survey (the fieldworkers inspected the surface by crawling on their knees shoulder by shoulder). In 14 experiments done in this way it was discovered that the crawling survey recovered from 170% to 1000%, or on average 350%, more artefacts than the walking survey in 70 cm intervals, which was in itself already absurdly intensive if compared to more standard survey procedures done in 10–15 m intervals. These

---

11 This is a point also emphasised in the Best Practices Prospectie project (see f.n. 10).
results have drastic implications about the recovery rate of such surveys and also prove the already mentioned point that without such additional property-based techniques or experiments incorporated into survey design the fraction of the sample acquired cannot be determined and interpretations of survey results based on quantitative analysis may be invalid. Furthermore, in these experiments, the results of both survey intensities were compared with test excavations of the upper 10 cm of the taphonomically active topsoil, results of which have worrying implications about the relationship of surface and subsurface archaeological record. The sample
of the crawling survey was capable of predicting 72% of artefact variance in the topsoil, while the very intensive distributional survey was able to predict only 24%. This means that the surface visibility of the properties of the subsurface record or in other words the capability to predict its properties on the basis of surface distributions is very poor even at great intensities let alone when using standard intensity levels of 10–15 m intervals. But nonetheless, such surveys are usually expected to detect the presence and predict the properties of much more deeply buried subsurface record, which is obviously very problematic, especially in case of development projects (Burger et al. 2004: 414-420; Burger and Todd 2006: 242-243; Burger et al. 2008: 221).

As demonstrated in the case study above, the expectation that surface distributions can be used to predict the properties of subsurface archaeological record can be problematic or unrealistic. Similarly, the expectations that patterns observed in the surface record reflect past behaviour also seem unwarranted because surface distributions primarily reflect post-depositional formation
processes or taphonomy of the landscape (see Burger et al. 2008: 203-211). But setting these interpretative problems aside an important methodological question is whether standard survey procedures even allow for documentation and recognition of realistic patterns in surface distributions? Because standard procedures use aggregate units smearing effect occurs, disguising spatial patterns and associations in the distributions we record (Fig. 7). Also by using aggregate units, which can be of different shapes and sizes, we are faced with the Modifiable Areal Unit Problem or MAUP, which is connected with the question to what degree our choice of areal units conditions the results of the analysis. This problem arises because we use arbitrary areal units for documenting a continuous space and thus obtain arbitrary spatial patterns (Harris 2006: 48). Data are ascribed to areal units, which are arbitrary and modifiable and have no natural meaning in the continuous space. However, if the units are arbitrary and modifiable than so are the results of the spatial analysis, which are heavily dependent on the shape and size of units used. Different ways of spatial data aggregation lead to an almost infinite number of possible spatial units, and patterns that result from this may vary widely and are thus an artefact of modifiable areal units and not of the archaeological phenomena themselves (Ibid.: 49). Modifiable areal units are the main cause of variability in spatial data interpretations because the choices of units and data aggregation process condition the patterns we may recognise. Different ways of data aggregation give different results but without any systematic trends. With the changing of areal units we arrive to different statistical results and generally the bigger the unit the greater the correlation between two variables. Thus, bigger areal units cause greater stability in the results and mask important spatial variations, which could be discerned if smaller units were used (Fig. 8) (Ibid.: 46, 49-50).

Incapability of discerning between spatial associations of data aggregated into units and real associations of unmodifiable individual data is endemic to all kinds if analyses based on spatial data aggregated into units. Thus, techniques of data collection which are not dependent on the frame and which allow joining and disjoining of data in different ways are called for (Harris 2006: 50-51). In the case of the surface survey, the only solution to this problem seems to be the use of point provenience instead of data aggregation. With the rapid development of GPS technology, it is now possible to do this in an efficient way and many survey archaeologists have
recently called for the need of point provenience or artefact accurate survey (see Wessel and Wohlfarth 2008: 15-18, 42-43; García-Sánchez 2013; García-Sánchez and Cineros 2013: 297-299; García-Sánchez and Ezquerra 2014; Trachet et al. 2017; de Neef et al. 2017: 285, 296; Gruškovnjak 2017b), which allows for recognition of real distributional patterns and spatial associations. However, such high resolution surveys may bring forth the problem of surface coverage because more work hours are needed to accomplish it, and despite the high data resolution and their representativeness within a small area, such an area may be too small to be representative and useful from the regional point of view (e.g. Bintliff 2000: 205; Fentress 2000: 44, 50-51). By lowering the intensity of the survey, the speed of discovering new data is increased, but their resolution is decreased (Burger and Todd 2006: 247-248, fig. 15-5). This is caused by the interaction of two main aspects of visibility, determined by survey strategy or two reasons why different survey strategies do not discover cultural material: (1) sacrifice of space or coverage – smaller the coverage less material will be discovered; (2) sacrifice of intensity – lower the intensity, less material will be discovered. This is an insurmountable problem because both sacrifices are inevitable and unacceptable at the same time. Regarding this, the difficult question is: “What is a better way not to discover artefacts, by not looking in enough places or by not looking closely enough?” The need for archaeological resource management and protection on the regional level, the fact that archaeologists will never know where all cultural material in the landscape is located as well as restrictions of time and resources probably call for continued use of conventional survey procedures. However, at least one phase of survey design should include control experimental surveys at different scales and intensities, which would then allow a quantitative understanding of these methodological sacrifices and a better understanding of the regional surface record (Burger et al. 2004: 420). As an ideal frame for such property-based investigations, the already mentioned Modified-Whittaker sampling strategy (Fig. 5) has been proposed, while other options are also worth exploring.

4th level determining visibility: surface and other environmental conditions during the survey

The fourth level that determines visibility is connected with surface conditions, accessibility and other environmental conditions during the time of the survey. Some surfaces are not accessible due to strong vegetation, difficult terrain, buildings or due to owners who prevent access and such areas cannot be surveyed. Because of such factors, total coverage of survey area is almost never possible and consequently, we are always dealing with surface samples or incomplete distributions (Schiffer et al. 1978: 8-10; Terrenato 1996: 223-224). The effectiveness of the survey is affected by different environmental conditions such as lighting conditions, weather conditions, flora and fauna etc. (e.g. Chapman 1989: 57; Barker 1996: 167). Among generally measured aspects of surface visibility is the assessment of surface exposure in relation to groundcover, usually measured with 1 to 10, 10 being 100% of the surface is exposed (see Bintliff 1985: 210; Gallant 1986: 406; Cherry et al. 1991: 27–28; Gaffney et al. 1991: 61; Terrenato 1996: 223; Terrenato 2000: 60, 66). However, a variety of other factors also has been proposed by Jitte Wagner at the Finds in the Landscape. New Perspectives and Results from Archaeological Surveys. / Funde in der Landschaft. Neue Perspektiven und Ergebnisse archäologischer Prospektion international conference, held on June 12th–13th at the Fritz Thyssen Stiftung in Cologne, Germany. There he presented point sampling as a subsampling technique performed in 10 m intervals as it was used in the case of Tappino Valley Survey (2013–2017), results of which have not jet been published (see Gruškovnjak 2017b; also see the comment on point sampling technique by Burger et al. 2004: 420-421).

13 Another strategy for property-based investigation that could also be borrowed from landscape ecology surveys, is the North Carolina Vegetation Survey (NCVS) nested plot (Peet et al. 1998). At least one comparative study (Goslee 2006) shows that it performs just as well as the Modified-Whittaker plot, but in comparison, it might be easier to set up and modify according to field circumstances. Jet another option for property-based investigations could be the use of point-sampling (Van de Velde 2001), combined with a standard survey procedure using transects or quadrats. With point sampling areas of around 2 square meters spaced in regular intervals are thoroughly cleaned and inspected which insures the discovery and collection of all artefacts present on the surface and enables comparisons with the results of standard procedure. This possibility for using point-sampling...
affect visibility which is why simple correction formulae (see Bankoff and Winter 1982: 152; Bintliff 1985: 210; Bintliff 2000: 204; Bankoff et al. 1989: 65, tab. 1; Gaffney et al. 1991: 64; Terrenato 2000: 66-69) cannot rectify all the biases that differential visibility conditions incorporate into survey results (see e.g. Banning et al. 2006: 739-740). These other factors for example include phase in the cultivation cycle and plant rotation, type of soil, colour and composition of the soil matrix, type of vegetation, relief, type of surface treatment (ploughing, disking), rain before survey etc. (see Hirth 1978: 126, 130; Jermann 1981: 79-82, 88; Gallant 1986: 406; Bankoff et al. 1989: 65, 69, tab. 1; Barton et al. 2002: 164; Banning et al. 2017: 473). The main problem in the assessment of visibility, however, is that it is very hard or impossible to formalize and realistically quantify the effects of the interaction of all such factors on surface visibility. Furthermore, gathering data on these factors is usually done by using aggregate units, and consequently subjected to the Modifiable Aerial Unit Problem (see above).

An example of the effects of surface treatment (ploughing and disking) as well as differences in soil types and micro-relief, is provided by a survey performed on a site 45-SA-17b in SW Washington, U.S.A., presented by V. Jermann (1981). There one is able to see a comparison between three consecutive phases of surface survey, first on an unmodified surface, then on a ploughed surface and last on a disked surface (Fig. 9). The differences in recovery rate between these phases were drastic: unmodified surface yielded 80 artefacts, ploughed surface yielded 600 artefacts, and disked surface yielded 750 artefacts. In addition, each consecutive phase recovered more artefacts of smaller sizes and lower weights than the previous phase. Because of the effects that ploughing and disking have on surface visibility, it has been suggested on several occasions that every surface should be prepared by ploughing and disking before the survey, though this is only acceptable in case of already cultivated or otherwise disturbed surfaces. Besides the differences in the number, size and weight of artefacts, there

FIGURE 9. Comparison of consecutive collection phases (a) on an unmodified surface, (b) on a ploughed surface, (c) on a disked surface and (d) a representation of joined results of all three phases of the surface collection. Each of collection phases was done after rain (Source: Jermann 1981: Fig. 3.5-8).
were also differences in spatial distributions between the three collection phases on the undulating surface of the surveyed field. Artefacts recovered in the first phase were generally confined to the central and lower slopes of major hillocks, which also correspond to parts of the survey area exhibiting sandier soils. This spatial pattern is probably connected with the effects of long-term exposure and eolian action on these dune-like hillocks, while in the parts exhibiting more silty or clayey soils the precipitation caused “puddling” which subsequently obscured the surface, and colluvial sediments accumulating between hillocks buried any artefactual remains. The difference between sandy and silty soils was particularly evident in the post-ploughing and post-disking collections. The northeast portion of the area is characterised by much sandier soils, and post-ploughing recovery rates were considerably higher in these sandier soils. Silty soils require much more intensive mechanical preparation to render optimal exposures as evident in the post-disking collection, which yielded a much higher number of artefacts on siltier surfaces than previous phases (Ibid.: 73-79, 83-88).

5th level determining visibility: the human factor

The fifth level determining visibility is the human factor or fieldworkers themselves for in the end it depends on them what will actually be noticed and collected in the field. Survey is done with “sentient instruments” capable of learning, boredom, exhaustion and so on, so the accuracy of measurements or effectiveness of recovery may vary according to their experiences, skills, motivation, interest, talent, visual focus, mood, health, fitness etc. (e.g. Schiffer et al. 1978: 14; Wandsnider and Camilli 1992: 185; Barker 1996: 167; Banning 2002: 65; Hawkins et al. 2003: 1504; Banning et al. 2017: 472,484).

One of the greatest problems concerning human factor here is that in the presence of a mixture of different types and/or colours of artefacts our attention and visual perception becomes unequally distributed and biased towards and/or against some of the types and/or colours of artefacts (Banning et al. 2011: 3454). The most problematic, even in the case of experienced workers, is the visual focus on ceramic artefacts, while recovery of stone artefacts is extremely low. To ensure the discovery of stone artefacts besides the ordinary field team a specialist for stone artefacts would be needed to inspect the surface only for this type of finds. In Boeotia Survey, for example, experimentations with a lithic specialist collecting only stone artefacts resulted in 1 tool per hectare discovered by the specialists, while the average for the rest of the fieldwalking team was zero (Bintliff 2000: 207).

Other ways to control and evaluate inter-observer bias have also been proposed. Control seeding experiments and resurveys, especially with property-based investigations, mentioned above in the context of second and third levels determining visibility, are already two procedures which can prove invaluable in evaluating inter-observer biases (see Banning 2002: 214; Hawkins et al. 2003: 1504; Banning et al. 2017: 474-476). Also, some basic steps which allow for the evaluation of inter-observer bias must be followed in any survey design. An essential first step is to track the work of different collectors by recording specific units they survey. If fieldworkers follow a consistent pattern of examining alternating transects, the evaluation of inter-observer biases is more effective, because the principle of autocorrelation can be used to advantage during the analysis. Evaluation is also made easier when transects are spaced at close intervals. It is also important for crews to mix fieldworkers of different experience and skill levels. In addition, crews and fieldworkers should be allocated in such a way to enable easy distinguishing of their respective abilities from spatial variation. For example, allocating each crew and/or fieldworker to equal numbers of survey units in each stratum of the survey area stratified according to different visibility conditions (according to 1st, 4th and possibly 2nd level) simplifies evaluation. Monitoring and documenting health, mood etc. of crew members can also prove valuable when comparing results of different survey areas. Training the crew members is obviously also essential. All fieldworkers should be familiar with the kinds of materials, site characteristics, subsurface feature characteristics, visibility conditions etc. that might be expected in the survey area, as well as trained to scan the surface in the same and consistent way. Motivating crew members is also an important factor as most highly motivated people will report most material. Motivation could possibly be stimulated by increasing continuity between fieldwork and analysis as well as involving staff in the publication of results (Banning 2002: 66; Hawkins et al. 2003: 1506, 1507-1508, 1510-1511).

Determining detection functions of each surveyor has also been proposed by Hawkins et al. (2003) and Banning et al. (2006; 2011; 2017) as a way to allow for the analysis of inter-observer biases in survey results. This is done by testing surveyors’ abilities to detect different types of artefacts under different controlled but realistic field conditions, in both cases such as are anticipated in the specific survey area (Fig. 10). For this purposes, 20 m wide and 100 m long experimental plots are set up on
calibration fields characterized by environmental conditions present in the survey area. The plots are overlain with a 2 m grid and longitudinally divided into two halves with a clearly visible rope. A known number of artefacts, simulating characteristics of archaeological materials anticipated in the survey area, is seeded throughout the plot in such a way that each long column in the grid would have the same number of artefacts of each type in a randomized position along the column. Crew members are then asked to walk along the central line without leaving it and record the positions of the finds they spot on a sheet with a corresponding grid. Such an experimental design might seem similar to the already mentioned control seeding experiments incorporated into survey design, however, in this instance, the experiment is performed differently and not as part of the field collection and also serves somewhat different purposes. It enables the determination of detection functions for each crew member in relation to different characteristics of artefacts and environmental conditions. Simply put, it shows to what distance laterally from the transect line crew members are able to spot a satisfying fraction of different types of artefacts in different field conditions. Such data can be invaluable when deciding on survey design, specifically optimal transect intervals which will ensure the same level of detectability in different visibility conditions of the stratified survey area. Furthermore, through the use of exhaustion maps in the analysis of survey results this data will allow for a realistic evaluation of inter-observer biases in the recorded artefact types and densities, which might otherwise be severely limited or impossible. Thus, with the use of such experiments one is able to adjust the survey strategy (intensity or sweep width) and evaluate inter-observer biases according to capabilities of any specific field crew in any specific survey area. Preferably such experiments would be performed repeatedly over the course of a survey project in order to control for maturation and history and produce “average” results (Hawkins et al. 2003: 1509; Banning et al. 2006: 728-730, 737-741; Banning et al. 2011; Banning et al. 2017: 478-481).

Conclusions

All the levels determining the visibility of archaeological record on the surface are in complex interactions and it is very hard to formalize their effects and realistically account for them let alone correct all the biases they incorporate into survey results. Nonetheless, we should strive towards accounting for them as precisely as possible by incorporating additional methodological procedures into survey designs. These should include initial geomorphological and pedological mapping followed by stratification of the survey area according to properties of the 1st, 4th and possibly 2nd level determining visibility and modifying the survey design according to these properties and experimentally determined crew member detection functions. Very precise descriptions of the soil matrix, micro-topography and other environmental variables need to be documented during fieldwork. Also, phases with control seeding experiments and/or property-based investigations incorporated into survey design may be the only way to allow for a realistic evaluation of precision, reliability and accuracy of survey results. And as the surface survey is an inherently biased discovery method multiple survey methods detecting different types of constituents of archaeological record are called for.

We should also realize what it is we are primarily discovering with the surface survey. The surface survey is not a discovery method, which would show the presence or absence of archaeological resources in the survey area. Instead, this method is geared towards discovering only disturbed and exposed archaeological record. From this point of view, this method primarily allows us to study post-depositional disturbance processes in the landscape or landscape taphonomy (see Burger et al. 2008) and its effects on the archaeological record. Thus, the surface survey can never be expected to reveal a complete distribution of archaeological remains still preserved in the landscape, let alone a complete distribution that once existed in that landscape.

In addition, we should be aware of the difference between the totality of archaeological record and the archaeological record as it is realised through our investigation methods during which loss of information or imperfect realisation of the archaeological record is inevitable. This is because the accuracy, reliability and precision of our methods are conditioned by a multitude of factors, archaeological record itself being only one of them.

14 Exhaustion maps of surveyed areas allow us to determine whether density variations in the distribution maps might be due to thoroughness of survey rather than real variations in the evidence present in the field. On such a map, each survey unit or collection unit shows the average detection rate of the crew members who worked there as an estimate of unit's exhaustion or thoroughness of survey. Units with high detection rate have a lower probability of overlooked evidence, while units with low detection rate have a higher probability of overlooked evidence and might be far from exhausted. Plotting artefact or site distributions on exhaustion maps shows how much the distributions may depend on the degree of exhaustion. It can also provide direction for decision making about which areas need to be rechecked in the subsequent phases of the survey or in future survey projects (Hawkins et al. 2003: 1509, fig. 6; Banning et al. 2017: 482-483, fig. 6; also see Banning 2002: 220-223).
FIGURE 10. An example of assessing crew detection functions on calibration fields in Jordan and Cyprus. In the upper part are views of the plots: a) pasture, b) olive grove, c) guava orchard, d) mixed field, all in Jordan, e) stubble field, and f) ploughed a field in Cyprus. Below are detection functions for all lithics (solid curve) and all pottery (dashed curve) along with half the corresponding sweep widths (vertical lines) at the corresponding calibration sites. p(r) is detection probability at range r and r is a range in meters (Source: Banning et al. 2017: fig. 3-4).
References


