

Ptolemy's Error: Truths and falsehoods in heterogeneous spatial data

(Presented at the Methods Network Workshop '*Space & Time: Methods in Geospatial Computing for Mapping the Past*')
(Presented at the Methods Network Workshop '*Space & Time: Methods in Geospatial Computing for Mapping the Past*')

Leif Isaksen

Electronics & Computer Science/Archaeology

University of Southampton

l.isaksen@soton.ac.uk

Please see leifuss.wordpress.com for accompanying slide presentation

[Slide 1: Intro]

This paper falls into two sections. In part one I shall present some observations on Ptolemy's *Geography* which seem apposite to this session on the reuse of varying representations of space, deliberately playing upon the ambiguity of my title. In particular they will focus on how he dealt with the problematic nature, or 'error' in his data. The solution he adopts leads to many of the problems which not only Ptolemy scholars, but many other map users, face today, namely that the information *appears* to be more robust than was (presumably) the author's intent. This 'error' in judgment will lead to a further reflection on the way that we too, through the necessary process of simplifying our spatial representations, create dangerously ambiguous datasets.

The second part of the paper is an all-too-brief attempt to catalogue the 'invisible' parameters of spatial data in a way that can be simply, and perhaps computationally, defined. Based on the premise that spatial representation, as simplified metaphors for their subject matter, cannot be considered the 'truth', they are defined by stating what deviations from the properties of their real-world correlates would render them 'false'. The intention is thereby to make it easier for data compilers to describe, and data users to understand, what conclusions can be meaningfully drawn from those representations (and combinations thereof). The catalogue and its uses are only intended as an initial proposal to be the basis for further development and discussion.

A final quick word about the terminology employed here. I shall use the words 'map', and 'spatial representation' more or less interchangeably to designate a relatively homogenous set of data that may form part of a more heterogeneous assemblage. 'Objects', 'things' and 'features' refer to things-in-the-world, (i.e. real stuff), whereas 'entities' and 'elements' refer to their expression within a map, either as points, lines, polygons or cells. I will also note that the history of Ptolemy and his sources is long, complicated, and frequently uncertain. For the purposes of keeping this talk within the prescribed time and avoiding digressions which are not pertinent to this session, I will generally use

the name 'Ptolemy' as shorthand for the body of Ancient Greek scholarship which ultimately led to the 'World Map' that we have come to inherit.

Slide 2: Ptolemy's World Map

The work of Ptolemy and his predecessors was more of a revelation than a revolution. The repercussions of their work for the history of cartography would only come into effect over a millenium later and there are good reasons for this. The prevailing method of mapping in the ancient world was by means of topological itinerary maps and gazetteers that provided their users with useful travel guides. Of primary concern to most travelers was knowledge of definite and relatively unhazardous routes from A to B. The idea of a world map that placed locations relative to an independent spatial framework, whilst certainly a fascinating scientific curiosity, was both too inaccurate and too uninformative (of terrain, winds, sea currents, etc.) to be of any practical use. Both kinds of maps share a common concept however, similar to the GIS systems of today – namely that they can be represented either spatially or textually. The Ravenna Cosmography itineraries, for example, appear to be derived from the graphical Peutinger Table, which in turn appears to have been developed from a series of written itineraries. Ptolemy was fully aware that copying a visual map was guaranteed to introduce a great quantity of error. In order to reduce these problems of transmission, his *Geography* is separated into two parts. The first, along with his methodology, describes how to draw a map according to two different projections. The second is a catalogue of locations, listing both towns and notable geographical features with their latitude and longitude.

Latitude presents a series of ever decreasing circles, known as parallels, set at regular intervals between the equator and the poles and measures absolute distance North-South. Longitude is marked by a series of Great Circles called meridians that run through the poles and subdivide the parallels into equidistant degrees of arc. Both are divided into units called degrees, which are in turn divided into 60 subdivisions called minutes, but whereas latitudinal degrees and minutes are of equal length, longitudinal degrees and minutes decrease as one approaches the the poles. Together they create a framework that is entirely independent of content, in contrast to itineraries in which all locations are described relative to others.

Whilst a quantity of erroneous material has crept into his work in the intervening period, a glance at the *Geography's* coverage of Europe and North Africa immediately shows the general stability of such an approach. On the other hand, we can also see that much of his data was inaccurate, and hence the reluctance of his contemporaries to use his work as an everyday tool. And, just as his contemporaries knew that it was problematic, we can also be certain that Ptolemy did too. Not only

does he refer to a variety of major changes to his predecessor's work (which imply an acceptance that world geographic knowledge was still in its relative infancy) but he limits the level of precision to which he assigns his locations. The most immediate symptom of this is that the lowest level of precision he gives is 1/12 of a degree (or 5 minutes) of latitude or longitude. In other words, within an area of about 16 square miles at the equator.

Slide 3: Ptolemy's Error

Now, assuming the data were entirely accurate, this level of accuracy might be just about bearable in terms of actual travel distance, but further examination begins to show that Ptolemy knew his information to be much less reliable than even this. We can demonstrate this fact by generating a histogram of his sites by their longitudinal location depending on which fraction of a degree they are assigned to, in this case those from the provinces of Hispania. Now if Ptolemy were equally confident of all his data we would expect to see a more or less even distribution of sites as they would fall randomly into this or the other location. But it immediately becomes clear that this is not the case – instead we see that there is an entirely uneven, but strongly patterned, distribution of sites which focuses principally upon degrees, half degrees, and then, apparently to a lesser extent, 1/3s, 1/4s, and 1/6ths. Only a very few seem to fall into those categories which can *only* have an precision of a 1/12th of a degree. In other words, when Ptolemy and his predecessors were assigning locations, they were doing so with a much lesser degree of precision than a cursory observation would suggest.

In fact, with some simple mathematics we can gain an approximate idea of what proportion of the data was assigned with what level of precision. If we assume that the number of sites falling into locations which must have been assigned to the nearest 1/12th (i.e. at 5, 25, 35, and 55 minutes) are broadly representative, we can postulate that a similar number of sites will have fallen into the other locations as well. By factoring them out (or an average of them), we can then use the same process to deduce approximate numbers, and hence proportions of data, assigned at coarser degrees of granularity - ie. at 1/6, 1/4, 1/3, 1/2 and a whole degree. The results are rough but no less clear for that. The following chart shows that, longitudinally, the majority of data was assigned to the nearest 1/3 or 1/2 of a degree (although none was assigned to the nearest degree). Now, At 1/3 of a degree the distance east-west becomes 16 miles at the equator and if a similar level of imprecision were to be assumed north south, a site allocated to that location, even assuming it was accurate, could come from anywhere in a potential area of 256 square miles. And at 1/2 a degree - which amounts to another third of the data, it is 576 square miles! We find similar results looking at latitude (although not as extreme), and across the entire World Map (where in places they are considerably worse).

Clearly navigating with such a map would be impossible. It also makes vividly clear why attempts to correlate Ptolemy's map with known locations are rendered more or less unviable.

What makes Ptolemy's data so frustrating is that we *know* that some of his information was intended to be relatively precise, even if only in relation to other, imprecisely located sites. But because we can only determine the level of precision statistically, in the majority of cases there is no way to ascertain the 'better' data from the worse. The use of a homogenising 'point' methodology, as opposed to fields of precision, mask the differences. Now I do not wish to accuse Ptolemy of disservices to Geography – rather, the point I wish to make is that the representation of Ptolemy's data potentially leads us to make assumptions about it which the underlying processes that led to its generation simply do not support.

Slide 4: Cautionary Tales

But surely this is a case from history, and ancient history at that. In the modern age of GIS and the OS National Grid, such problems can't arise, can they? Well, yes they can and do, despite the considerable efforts of geographers to make such issues more transparent. We need not look further than an early (unpublished) draft of a well known specification for Monument Inventories. Within its instructions for the use of point data recorded as an OS National Grid reference, it suggests that imprecision can be dealt with by buffering the points. At first glance, this may appear to be logical – if we know the degree of precision, this can be catered for by extending a bounding circle around the point by the equivalent amount. But this entirely misses the point (quite literally), because a grid reference is not a point location at all – it is a reference to a grid square. The Northing and Easting that make up the reference relate only to its most southerly and westerly location whereas the degree of imprecision (indicated by the number of digits it comprises) dictate its length and breadth to the North and East. A buffer around such a point will contain only a fraction of the actual area and great deal more besides in which the desired feature *could not* be found.

I could give examples of further misconceptions about mapping which abound, but that is not the focus of this paper. Rather, I wish to contend that much of the confusion which arises, and will continue to arise as scholars and the public at large increasingly engage with spatial information – a development which has been dubbed 'neogeography' - is due to a reliance on spatial data categories that are too crude to provide sufficient guidance for use. Points, lines, polygons, vectors and rasters are frequently referred to as though they came with a one-size-fits-all set of guidelines, and the importance of metadata, whilst acknowledged, is seldom adhered to. For the remaining duration of this paper I would like to attempt a separate categorization of spatial data by means of identifying a

number of orthogonal 'facets' which may help clarify both to compilers and users what conclusions it is capable of sustaining, and to what extent it may be meaningfully combined with other data. None of these categorizations are new, but I hope that in the process of naming and clearly defining them, it may be simpler for a compiler to indicate a lot of meta-information by means of a very simple schema. It is the simplicity of the schema which is key because it potentially opens the door to a number of other advantages which I shall address briefly at the end of this paper.

Slide 5: The Nature of Historic Mapping

Before we look at the categories themselves, I should like to identify some basic assumptions about the maps with which we are dealing. Firstly they are composed of two core components – a spatial framework, and a set of entities that are assigned to it. The framework may, indeed almost certainly will, contain a degree of distortion, but it must be topologically cohesive. i.e. it must be describable (and hence reproducible) according some function, f , rather than simply an ad hoc jumble of locations. Entities must be given a spatial location (although not necessarily extension) by assignment to that framework and must represent things-in-the-world, thus artistic endeavours are not the subject of this discussion. By way of representing things, they must also be assigned at least one semantic property that indicates their ontological nature – they cannot *merely* be points, lines, polygons or cells. The data must furthermore be descriptive, by which I mean that it attempts to describe the world as it is, was or will be – either in terms of its physical reality, or in the more metaphysical sense of algorithmic output. I am thereby also excluding prescriptive data: i.e. that intended to cause the instantiation of its subject by virtue of its own existence. An engineer's blueprint, for example, may be ill-conceived, or the construction may be built in a different manner, but it cannot in any meaningful way be deemed correct or incorrect.

As a side note to these distinctions, we should also note the difference between maps of the past, present and future. Maps of the past are necessarily descriptive and may well describe things which no longer exist, thus we may have no epistemic root by which to determine falsity, but this does not preclude us from talking about their subject matter as 'real things'. Maps of the present, are, *de facto*, descriptive of the past. They differ from past maps in that the frame of reference by which their correctness can be judged is an ever-rolling present. One minute the map may be true, the next false. That's why Ordnance Survey have to keep charging us for data we've already paid for. Maps of the future may be either predictive (and hence descriptive) or prescriptive. This is not in the sense of a 'predictive model' however, which describes an algorithm applied to a dataset. Predictive maps represent a spatial state-of-the-world that may or may not come into existence, and its state of

'correctness' must be suspended until that time, and perhaps indefinitely.

The fundamental problem that the following categories are intended to address is that the contents of a map *underdetermines* the basis upon which it was compiled. In other words, it is, by definition, impossible for a user to ascertain the principles by which a map was created merely by looking at it. In order to address this, a number of universal attributes of the data need to be declared, each of which can take a strong or a weak form. The weak form identifies a minimum criterion for a map to have meaning in that sense whilst the strong form identifies further criteria. It is not necessarily to be expected that the map will actually comply with those criteria, in fact it may seldom be the case that they do in their entirety (and hence most maps will contain a degree of falsity). Their importance lies in the fact that it is specifically *against* these parameters that the map has been compiled. Hence, the focus is not whether a spatial representation can be said to be 'true', but instead, what potential disparities would render it 'false'. To put it another way, the dataset itself should define the criteria by which it should be deemed 'correct' or 'incorrect', and all *other* correlations between representation and reality are merely happy coincidence.

Slide 6: The Categories

The list of categories here presented is most likely incomplete, and certainly in need of modification, but will serve as the starting point for discussion. Listed in each case with the weak form preceding the strong form, they are:

Incomplete vs. Complete

Fuzzy vs. Hard

Durée vs. Snapshot

Marker vs. Depiction(?)

Observed vs. Derived

As I look at each briefly in turn, I will use the examples of Ptolemy's map and an OS map, such as the Landranger series, in order to help make the distinction between each, and it should be borne in mind that all maps fall into either the weaker or stronger of *each* of these categories.

Slide 7: Incomplete vs. Complete

Incomplete maps are aggregations of arbitrarily obtained data. They are only false if they present entities which do not have a real correlate. Ptolemy's world map falls into this category – it is not

rendered false for not representing the location of Roman Edinburgh, whereas the presence of, say, Atlantis would make it incorrect. If a map is Complete, it must express as an entity *all* members of an abstractly defined set of features (which includes a spatial boundary) that it intends to represent. It is false if *either* any of those members is missing, or any are present which do not exist. The OS map of a town presents a good example – if fictional houses are included it is incorrect, and once new houses are built it becomes obsolete. Complete maps can be used for density analyses and statistical correlations with other data whereas Incomplete maps cannot. A special case of Completeness and Incompleteness might be termed 'evenness' and 'unevenness'. A plan of fieldwalking results presents a Complete set of *observations*, because they are evenly distributed, even though it is not a Complete set of all archaeological material in the study area. Plotting of randomly encountered finds, however, is uneven, and hence Incomplete.

Slide 8: Fuzzy vs. Hard

Fuzzy data is, by its very nature, hard to pin down. For it to have any meaning at all however, there must be, at the very least, a continuum between one category and another (else all the data would be uniform by definition and hence meaningless). Spatially, there are two main kinds of fuzziness – Vagueness and Imprecision. Vagueness is inherent in certain concepts, such as 'romanized' and 'non-romanized', to which although no absolute boundary condition can be set, *some* entities could be assigned to *either*, but no entities can be assigned to both. We may represent this either as a relatively arbitrary spatial division (e.g. along the Antonine Wall) *or* by giving the vague entities their own category. The data is false if elements which can only be assigned to one category, are assigned to the other. Imprecision is caused by having (or imparting) imperfect knowledge of the state of the subject. The data is false if it is inaccurate - i.e. it falls outwith the boundaries of that imprecision. Hard data is that which intends to divide subject into distinct categories. It is false if either the real objects are incorrectly assigned in the representation, *or* if that condition boundary cannot meaningfully be made to hold. As we have seen, despite its appearance, Ptolemy's map is fuzzy, due to Imprecision, rather than Vagueness. OS maps are generally Hard, with clearly defined spatial limits, although some elements, such as beach boundaries are, by their very nature, Vague.

Slide 9: Durée vs. Snapshot

Because spatial representations tend to be static, they frequently imply the temporal coexistence of their subjects when this may not be the case. 'Durée' maps are a composite of elements from within a specified, contiguous time period, generally with a conceptual boundary (e.g. Iron Age Britain). They are only rendered false if they contain elements which were not present (or never took the represented form) in the given period. As it is not possible to represent all manifestations of the subject matter throughout the duration of the period, absence of an element (or of a form) does not

render the map false, so they are, by definition, also Incomplete. A 'Snapshot' map is correct if and only if all of their contents exist together, and in that form, at a specified moment in time. Maps of the present (which is a 'rolling' moment) are a special instance of this, although they will rarely, if ever, be correct for long. Ptolemy's map is often perceived as a snapshot of the Roman empire during his lifetime, but is in fact principally composed of data that may stretch back as far as Eratosthenes in the 3rd century BC, or even earlier, and Ptolemy was well aware of this fact. Hence, it must be considered a 'Durée' map. OS maps, naturally, are intended to be maps of the present and hence Snapshots.

Slide 10: Marker vs. Depiction

The elements of a map may be locational markers or they may be Depictions. A Marker is only rendered false if it deviates from the location with respect to the spatial framework. It is important to note that Markers need not be points – in fact, as we have seen from the preceding discussion, it may be better to render them as imprecise fields which designate the area in which the feature is to be found. A Depiction, as well as locating the entity correctly, must represent the prescient *spatial* features of the feature - falsity is caused by deviation of either form or location of the entity from the subject matter. Ptolemy's map, as for once is obvious, uses markers rather than representations – it is designed to present the locations of its subject, rather than their spatial attributes. OS maps show both the form and location of most of their elements, and are thus Depictive.

Slide 11: Observed vs. Derived

Observed data is composed of elements describing things that exist in the world. It may include physical objects such as trees or houses, or social or political objects like borders or traffic zones. Falsity is caused when the presence, absence or properties of these elements do not correspond to their realworld counterparts. Derived data is the output of combining such elements with a defined function. Predictive models, node networks and viewsheds all fall into this category. It is important to note that the *map's* correctness is not based on whether such a function actually determines a concept (such as visibility), but instead, whether that function is properly rendered. It follows from this that, for the map to be meaningful, a user needs to have access to both the function and the underlying data from which the output was generated. Whether the function accurately models some abstract concept is a separate issue from the map's correctness, as is the truth-value of the underlying data. Ptolemy's map is in fact a form of Derived data, as his latitudes and longitudes are generated for the main part algorithmically from source data based on his reappraisal of the Earth's size. If we had the original data set and function to compare it to, we could legitimately critique it where it deviates from that algorithm. OS maps are Observed - they seek to represent things themselves.

Slide 12: Summary

So what are the advantages of categorizing maps along these axes? The first is that it provides us with a very simple schema (in fact just a handful of Boolean values) with which we can tag up datasets for dissemination. Because these can be handled computationally, not only can we automate the production of generic caveats for use, but we can also use it to provide a host of other services as well. For instance, Complete datasets must by definition have a spatial boundary, and our software could autogenerate one. Or it could autoconvert between point and precision field representations of Marker data. This kind of information markup could greatly enhance the capabilities of an 'intelligent' GIS. The second is to note that we are swimming in an ever-growing sea of spatial data which is constantly being repurposed - this is the blessing and curse of Neogeography. If we are to make the most of it, we need to find clear and simple means by which individuals can produce, understand and utilise it. A cartographical map (either paper-based or digital) may comprise of numerous different layers of data, and the 'weak' and 'strong' nature of the axes means that the entire cartographical presentation can also be categorised in this way by defaulting to the weaker of the two types where these conflict between layers, whilst not losing sight of its component parts. Most importantly of all, they encourage the compiler to vocalise the epistemic limits of the map.

The foregoing discussion is intended primarily as food for thought - it is unlikely that I have considered all aspects of truth and falsity in mapping, or got the balance entirely right. We have seen, for example, that Durée maps must also be Incomplete – perhaps they ought to be merged into a 'Composite' category. However, I do feel that the approach of defining specific orthogonal facets of well-defined 'correctness' has great utility, and it would be an interesting exercise to see whether certain combinations of these attributes form a natural 'periodic table' of mapping types. Most of all, I hope that the preceding observations have provided some further ideas for discussion in our attempt to more transparently deal with heterogeneous data.