[FC-07-024] Conceptual Models of Error and Uncertainty

Abstract

Uncertainty and error are integral parts of science and technology, including GIS&T, as they are of most human endeavors. They are important characteristics of knowledge, which is very seldom perfect. Error and uncertainty both affect our understanding of the present and the past, and our expectations from the future. 'Uncertainty' is sometimes used as the umbrella term for a number of related concepts, of which 'error' is the most important in GIS and in most other data-intensive fields. Very often, uncertainty is the result of error (or suspected error). As concepts, both uncertainty and error are complex, each having several different versions, interpretations, and kinds of impacts on the quality of GIS products, and on the uses and decisions that users may make on their basis. This section provides an overview of the kinds of uncertainty and common sources of error in GIS&T, the role of a number of additional related concepts in refining our understanding of different forms of imperfect knowledge, the problems of uncertainty and error in the context of decision-making, especially regarding actions with important future consequences, and some standard as well as more exploratory approaches to handling uncertainties about the future. While uncertainty and error are in general undesirable, they may also point to unsuspected aspects of an issue and thus help generate new insights.

Keywords: assumptions, error, forecasts, fuzziness, modeling, scenarios, time, Uncertainty, unknowns, vagueness

Author & citation

Couclelis, H. (2020). Conceptual Models of Error and Uncertainty. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2021 Edition), John P. Wilson (Ed.). DOI: 10.22224/gistbok/2021.1.3.

This topic is also available in an earlier edition:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Definitions within a conceptual model of uncertainty. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Explanation

1. Introduction: Knowledge, Error, and Uncertainty in GIS&T
2. Kinds of Error and Uncertainty in GIS&T
3. Uncertainty and Error in Spatial Forecasting and Decision-Making

1. Introduction: Knowledge, Error, and Uncertainty in GIS&T

Though closely related, uncertainty and error are very different concepts. One way to think about the difference is by means of the familiar sequence: data, information, knowledge,

wisdom. On that scale, error would be more closely associated with data, and uncertainty with knowledge. Information is the step in between, whereby some specific knowledge is acquired from specific data, while the notion of knowledge is broader, integrating information from many different fields and sources. But these distinctions can be subtle, which is why a large number of additional terms are also used to express more specialized forms of, or additional perspectives on error and uncertainty. Terms such as inaccuracy, imprecision, fuzziness, vagueness, ambiguity, bias, misinformation, ignorance, and others are used to populate the transition between data and information, or indeed between data and knowledge.

Another distinction between uncertainty and error may be made along the objective-subjective axis. Error is the more objective notion of the two. Roughly speaking, something is right or wrong, correct or incorrect. If there is doubt, potential errors may be expressed in probability terms and other methods. In addition to the common uses of the term, data science includes a specialized statistical concept of error as “the (unknown) difference between the retained value and the true value” that is the converse of ‘accuracy’. Uncertainty is more subjective and relative, and may not even be quantifiable. Uncertainty depends on someone’s knowledge, interests, and motivations, and on the context within which an uncertainty may arise. One user may consider a particular database good enough, while for another its errors and related uncertainties may make it practically useless. By contrast, a GPS navigation system with a location error will equally affect every driver who follows it.

Error and uncertainty have been major concerns for GIS&T researchers since its earliest days in the 1970s, first with more emphasis on spatial data handling and data quality and thus on data error, and on the uncertainties resulting from data error (Goodchild and Gopal 1994). As the field matured and broadened technologically and conceptually, new research areas were added that brought with them their own aspects of uncertainty and error. Data visualization became much more than a good map, a growing emphasis on societal issues led to the introduction of qualitative data, spatial was extended to spatiotemporal and to the notions of change, motion, and event, data synthesis in the form of process modeling increasingly complemented data analysis, and the recent growing interest in the role of artificial intelligence (AI) in GIS&T is also accompanied by its own novel forms of uncertainty. There is now hardly an area in GIS&T that is not engaged in active research on error and uncertainty, or in the systematic probing of proposed solutions.

1.1 The Role of Time in Error and Uncertainty

There is also an important temporal dimension to error and uncertainty, as there is to knowledge. Typically, the most reliable information is from the present and the relatively recent past. In principle, the further back we go in time, the less secure our knowledge is. What is more, we cannot go back and check our data, unless sufficient physical traces or reliable documents remain. Scientists who study the past such as archeologists, anthropologists, historians, paleobotanists, climatologists, geologists, and others have to rely to a large extent on indirect clues or unreliable written sources. Even though GIS is now used in all these disciplines that study the past, their approach to error and uncertainty is based as much on expert hunches and informed speculation (the ‘wisdom’ that is ranked above mere ‘knowledge’), as on the kind of rigorous analyses possible with relatively current data.

On the other side of the present is the future. Humans have always been more interested in the future than in the past, if only because that's where the threats and the consequences of our actions lie. Yet whether we are trying to prepare or to decide, we are faced with an unsurmountable handicap: we cannot have any data from the future, and never will. Thus, uncertainty about the future increases much faster than uncertainty about the past. Since times immemorial people have tried to guess what to expect, first with oracles, prophecies, and divination, recently with the help of highly sophisticated technologies and methods. As we will see in section 3 below, for many years now GIS&T has been contributing to society's search for approaches that may help mitigate certain more tractable kinds of future uncertainties.

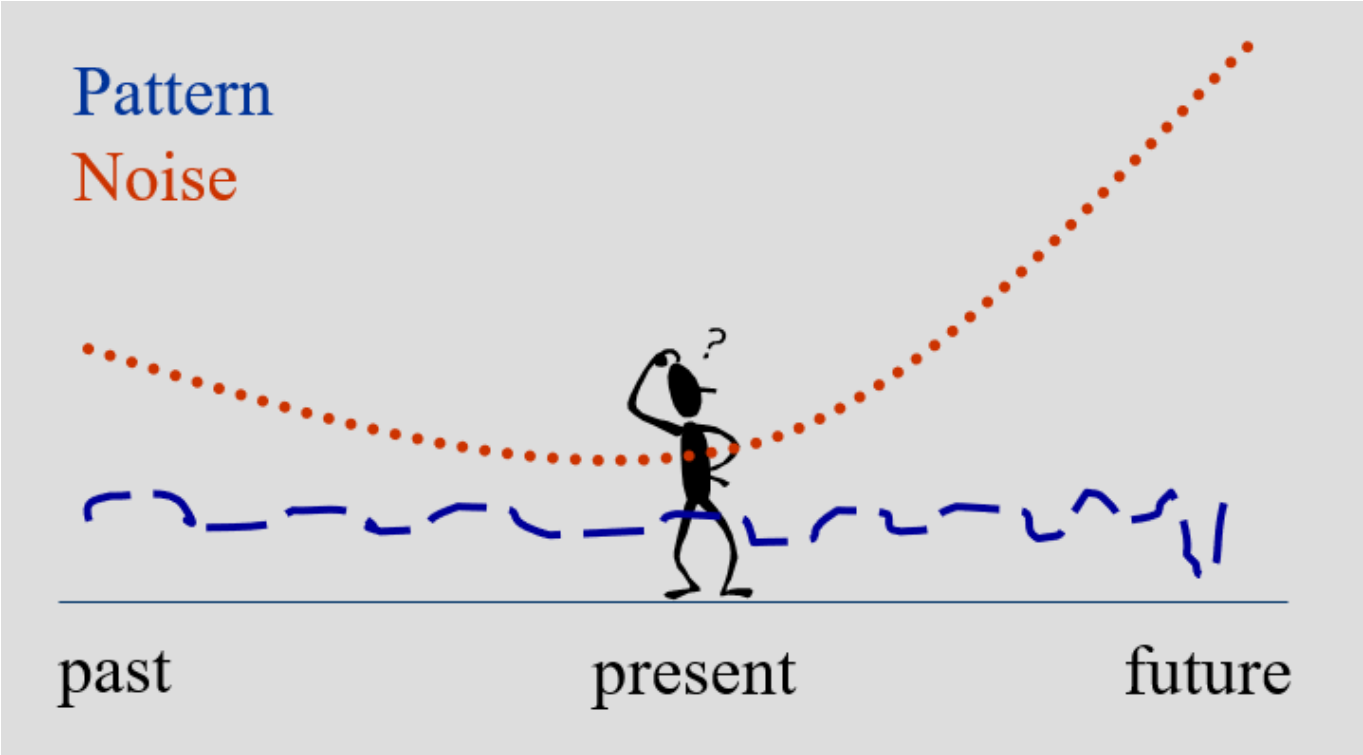


Figure 1. While uncertainty increases the farther we go towards the future or the past, we can still count on many things being largely the same as today. Source: author.

There is, however, one factor thanks to which no matter how much uncertainty the past, present, or future may hold, not everything is totally unpredictable. This is because most change in the world is gradual, and abrupt changes are rare. We can thus confidently assume that ancient people were very much like us though living under very different circumstances, that there will be fewer commuters on the roads on Sundays than on weekdays, that cities will continue to exist pretty much in their current forms, that technology will continue to develop, that there will be more hurricanes in the late summer than in other times of the year, and so on, depending on the time horizon considered. We can think of these expected regularities as the 'pattern in the noise', illustrated in Figure 1 as the broken horizontal blue line that gives some basic structure to the 'noise' of the surprising, the unexpected, the unexplainable, and the unknown represented by the area below the red dotted line. In section 3 below we will see how GIS&T capitalizes on that element of predictability in reality to contribute to the reduction of error and uncertainty in areas where spatiotemporal data are used.

2. Kinds of Error and Uncertainty in GIS&T

2.1 Error

In our era, when an abundance of good data is taken for granted, Big Data is more than just having access to lots of data. The concept involves additional aspects that are expressed by the notion of the '5Vs'. These stand for volume (more data is better), variety (more comprehensive is better), velocity (fast updates, ideally in real time, are better), veracity (truth, reliability), and value (meeting users' needs). But ideal situations are rare, and there can be problems with any one or even several of the V's at once: sparse data and incomplete databases, important data categories missing, data that are obsolete, presence of errors, available databases not quite suitable for desired applications. Each of these imperfections gives rise to uncertainties that often combine to create serious problems.

Problems with data errors and uncertainties have recently become even more prominent. This is because the flipside of today's data abundance is that uncertainty about every aspect of data quality has greatly increased in the past couple of decades. Until late in the 20thcentury spatial data were collected, standardized and published mainly by authoritative sources such as USGS, using their own networks of sensors and other sources of known reliability. Since then, however, there has been a flood of data posted on the web by unofficial contributors ranging from experienced researchers and respected organizations to young school-children. Their sources may be second-hand, as when data-mining information on socio-spatial networks from platforms like Twitter, geo-coded references to places from Wikipedia, or images of places from Flickr are used. Alternatively, the data may be collected and contributed directly (crowd-sourced, volunteered) by individuals or groups using GPS-enabled smartphones, drones, thermometers, cameras, or just basic observation skills. These newer, unofficial kinds of databases can be very valuable in principle because they address aspects of human life and the environment that may not be covered in other ways. But they may also be plagued by problems of inaccuracy, imprecision, plain mistakes, missing data, bad metadata, or they may be obsolete or too small to work with. Research is underway on several fronts to mitigate or circumvent these problems so as to help valorize the promise of informal databases (Sui et al. 2013, Zuefle et al. 2020).

Dealing with error (lack of 'veracity') even in regular databases is significantly complicated by the fact that error can result from many different sources, and can take many different forms. Guptill and Morrison (1995) distinguish seven elements of spatial data quality that are subject to error and may thus also be major sources of uncertainty: Lineage, positional accuracy, attribute accuracy, completeness, logical consistency, semantic accuracy, and temporal information accuracy. Most errors in these elements can affect any kind of data. (In)accuracy in particular and the related notion of (im)precision are fundamental kinds of error in the sciences. Accuracy is in fact the converse of error, i.e. it is defined as the extent to which an estimated value approaches the true value, while precision is the degree of dispersion of observations around a mean. Errors in positional (or locational) accuracy, also reflected in fundamental geographical applications such as distance measurements, the drawing of physical boundaries and other linear elements, or the spatiotemporal analysis of motion, uniquely characterize the spatial domain. Finally, common human errors are always possible in GIS as elsewhere. These may be errors of judgement, errors due to ignorance, to

inattention, to misunderstandings, to insufficient preparation or effort, to reasoning, or to any other cause.

Beyond straight data errors, many application contexts can lead to error-like problems and uncertainties due to the special nature of space as an object of study (O'Sullivan and Unwin 2003). Spatial data require distinct analytic techniques to deal with fundamental geographical concepts such as distance, direction, adjacency, interaction, and neighborhood. These in turn are reflected in well-established analytical properties or characteristics of geographic space such as spatial autocorrelation, spatial heterogeneity, the modifiable areal unit problem (MAUP), the ecological fallacy, the boundary or edge problem, the ecological fallacy, and scale (the last two also found in the non-spatial sciences). These properties of space that defy standard statistical analysis present pitfalls for GIS users who are not careful or knowledgeable enough, and can lead to major blunders if not treated appropriately.

2.2 Vagueness and Ambiguity

Additional sources of imperfect knowledge are not specific to geographic space but have very specific spatial manifestations. Vagueness is the general term for a class of common situations whereby objects with sharp boundaries need to be identified on surfaces and for phenomena that are in reality continuous, that have uncertain boundaries, or whose shifting boundaries cannot be precisely determined (Bennet 2010). There are many special cases of vagueness. The Sorites paradox was first described by the ancient Greeks ('soros' is Greek for 'heap'), and goes as follows: 1 grain of sand does not make a heap, 1+1 grains don't make a heap, 2+1 grains don't make a heap, 3+1 grains... a zillion +1 grains don't make a heap. This is the problem of the transition from the individual to the collective. Substitute 'tree' or 'house' or 'island' for grain, and you will never get to the 'forest' or the 'neighborhood' or the 'archipelago'. Another case of vagueness is to problem of Individuation, whereby discrete objects ('individuals') with a specific extent, shape and boundary must be pulled out of a continuum. Examples include finding where the mountain ends and the valley begins, separating out the lakes from the river that runs through them, telling the end of an ecosystem or microclimate from the beginning of another, or the city's edge from the suburbs. The best-studied case of vagueness is probably that of indeterminate or ill-defined boundaries (Burrough and Frank 1996). Here we are no longer carving objects out of, or drawing dividing lines along a continuum. We are dealing with recognizable individual objects but which have no certain extent, shape and boundaries. These may be static objects or deformable continuous phenomena such as storms and floods, active fires, rivers and lakes, estuaries, or migrating animal herds.

The most general approaches to modelling vagueness are based on fuzzy logic and the theory of fuzzy sets. Unlike classical set theory where set membership is binary (0 or 1, something is or isn't a member of a set), fuzzy set theory works with degrees of membership between 0 and 1 both inclusive. Fuzzy logic allows the manipulation of fuzzy relationships, and thus of imprecise information. A core concept is defined (say, 'mountain peak'), and the value of 1 is assigned to parts of mountains that are clearly recognizable as mountain peaks. But where does the peak really start or end? How far up the mountain must one be before they can claim that they are 'at the peak'? 'Not quite' at the peak may be assigned a 0.80 fuzzy value, whereas 'just starting to climb the mountain' may be a 0.0002 for 'peak'. (Fisher et al. 2003). While to the untrained eye fuzzy logic may appear to be some inferior form of statistics, it is in fact a very versatile approach to uncertain

information that can be extended in different ways, in particular with formal concept analysis, rough set theory, and possibility theory (Wu et al. 2019).

Cohn and Gotts (1996) note that in many cases it is possible to reason about objects with indeterminate boundaries without the use of fuzzy sets and fuzzy logic, by relying on topological relations. For example, we know that a lower-elevation contour of a mountain's DEM will contain the peak (as well as the summit), that southern England (a large region) will almost certainly contain London (a city in the south of England), that London and Paris do not overlap or touch because they are contained in non-adjacent regions, and so on. Another possibility, from the perspective of spatial cognition, is to just ask people where they think a boundary lies. This method has been used for many years to identify the extent of neighborhoods and other kinds of regions. In a series of increasingly sophisticated papers, Gao et al. (2017) have used this approach to find 'where northern California ends and southern California begins', with recent work supporting the cognitive results with more formal methods. An additional way to approach the ill-defined boundary problem is to be more specific about why and for whom that information is needed. (For a farmer? For a housing developer?) Couclelis (1996) develops a taxonomy of ill-defined boundary cases based on the following three dimensions: (a) the empirical nature of the object itself (as determined by basic material, temporal, and topological properties); (b) the mode of observation (including scale, space-time resolution and error); and (c) user purpose. Each of these allows multiple options that may be individually combined across dimensions. In the example used, after the impossible combinations are eliminated, over one-hundred answers to the boundary-specification problem are found.

Related to vagueness but simpler as a concept is ambiguity, which can also lead to errors. Ambiguity has to do with meanings that can be interpreted in more than one way. It is thus about unclear semantics and language. A map where the colors on the legend don't quite correspond to those used for the phenomenon represented on that map is ambiguous, and so is a search for the 'distance' between two places when it not clear whether time distance or distance in miles is meant.

A further class of problems involves mapped coastlines and some other lines projected on natural landscapes, such as borders between countries. These have the interesting property of being fractal, that is, they have no true length: because they are very irregular, with many nooks and crannies, their measured length depends on the length of the 'yardstick' with which they are measured. The longer the measuring rod, the shorter the boundary. A well-known example of this phenomenon is the border between Spain and Portugal, which is apparently 20% longer for the Portuguese than for the Spanish. Some may see this discrepancy as a major data error. As with vagueness and ambiguity, there are no inherent errors with fractal phenomena, but these can cause confusion and uncertainty because of the subjectivity involved in their measurement. Interestingly, a phenomenon similar to that of the linear fractal feature whose length depends on resolution also appears in in the 3rd dimension, where multi-scale, multi-resolution representations of a DEM may reveal different classes of features in different locations. Fuzzy logic can help provide an answer to the resulting uncertainty (Fisher et al. 2003).

2.3 Uncertainty

"… as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not

know. But there are also unknown unknowns—the ones we don't know we don't know". People thought this famous quote by former Defense Secretary Donald Rumsfeld was very funny, and it went viral (Rumsfeld 2002). Yet it is part of military-speak, and as we will see, for good reason. We could even add a fourth one: unknown knowns. That's when information relevant to our problem exists out-there but we are not even aware that we should look for it – or where and when to look, or whom to ask.

One of the great mysteries of the past decade, a tragic (and very geographical) example of these kinds of uncertainties and their implications, is the disappearance of Malaysia Airlines flight MH 370 in March of 2014. Ironically, later that same year, Europe's Rosetta spacecraft put a lander on comet Churyumov-Gerasimenko very close to the originally planned site – whereas here on Earth, where the position of aircraft is supposed to be known exactly at all times, all the technology and expertise of several advanced countries have been unable to locate the lost Boeing 777 or to explain its disappearance, after six years of efforts. There are the known knowns: the tracked sections of the plane's trajectory, including an abnormal sharp turn and the sudden disappearance of the aircraft from primary radar, while it continued to be tracked in flight by the military. There are the known unknowns, which include the reasons for the above two abnormalities, the remaining untracked trajectory, the cause of the crash, the possible role of the pilots or others in the crash, the location and extent of the optimal search area, the actual location of the wreck, and so on. (Note that until summer of 2015, even the fate of MH 370 – whether it had actually crashed – was among the known unknowns.) Then there are the unknown unknowns, several of them in this case related to human factors: if the cause of the known anomalies and/or the crash itself was technical, what was it? If someone deliberately caused the crash – who, how, and why? Were the contradictions, delays, half-truths, and inaccuracies in the information provided by three neighboring countries at the time of the incident deliberate, and if so, in which cases, and why? Was critical information suppressed? (Couclelis 2016). And an unknown known would be: Major X and a couple of his colleagues know exactly what happened but they are not talking as long as nobody asks. (This is of course hypothetical, since 'unknown knowns' can only be known in retrospect, if at all).

2.4 External Dimensions of Uncertainty

The above example is a reminder that there is a surprising number of things that we cannot know, and questions that we cannot answer about a spatiotemporal problem, which are not directly the result of imperfect data. Another way to look at this is by the sequence of questions: 'where, when, what, how, who, why' that could be used to tell the story of Flight MH 370 discussed above, but which can also apply in whole or in part to most any empirical problem addressed with GIS. There can be errors and uncertainties at any of these steps. 'Where' and 'When' call for quantitative answers (and errors and uncertainties), but beyond that, quantification is at best partial, and uncertainty increases the further we go along the sequence.

Uncertainties plague empirical studies in practically any field. This is in part because the specific problem area or system of interest described, studied or modeled can only be a very minor part of an environment of facts, events, actions and possibilities that are extraneous to the specific study or model. That is, there is a lot of 'noise'. Change is continuous, and the world does not stand still once the data are analyzed, the conclusions drawn, and the study submitted. The longer the study's time horizon, the more likely it is to be eventually invalidated by the unexpected. Figure 2 provides a rough illustration of the

different layers that contribute to this state of affairs. First, there are nearly always stakeholders and actors that are part of the system of interest: land owners, environmental groups, farmers, fishers, urban residents, and so on. These we can include in the study using questionnaires, public participation, and other crowdsourcing methods, all with their own uncertainties and errors. But there is no guarantee that even these known actors (or their descendants) are not going to think differently down the road. Second, there is a policy system that sets goals, develops policies, and implements decisions reflecting current societal and political values. But governments change, laws change, technologies open up new options, economic conditions change, and social goals and values change over time. More housing downtown or in the suburbs? Bullet trains or new airports? The economy or the environment? More immigration or less? More foreign aid or domestic infrastructure? Third, we may develop formal or computational models such as agent-based or cellular-automata models to help predict how the network of study-internal and external interactions may play out. But models bring their own errors and uncertainties and can at best capture some of the current 'pattern' in the noise. As a famous Cambridge economist once said, "A model is the unrolling of a carpet that exists now". And fourth, there is always the wider world of domains beyond the above three, such as Mother Nature and other countries, that keep throwing major surprises at us.

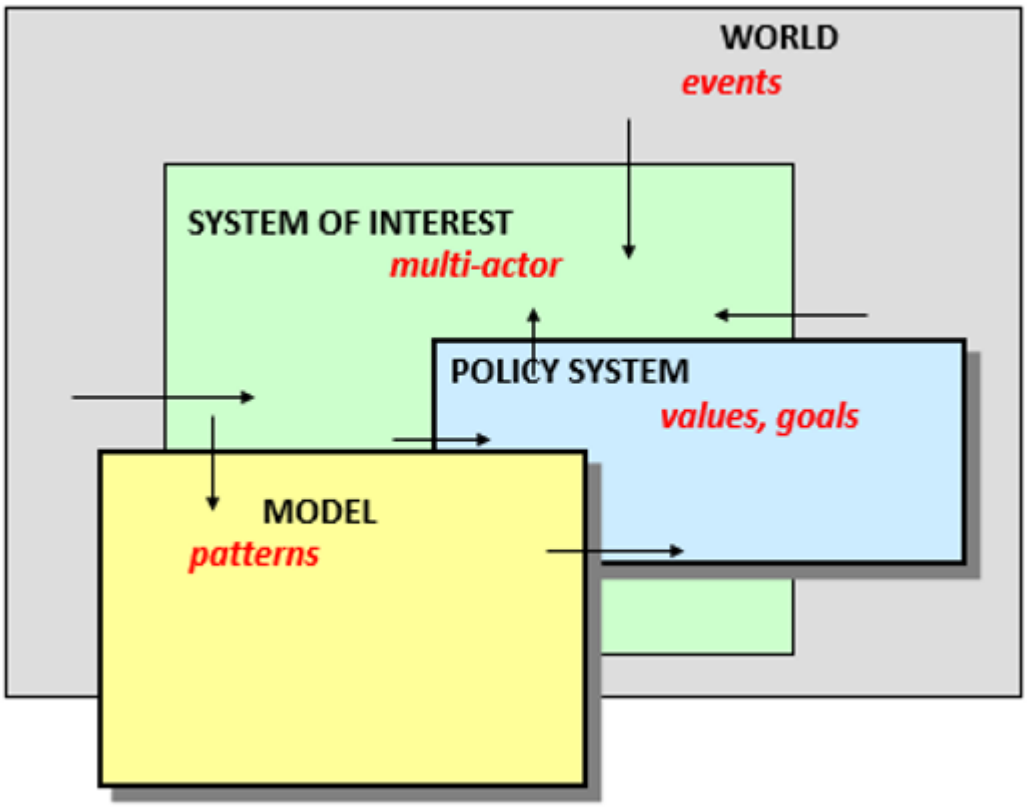


Figure 2. Many important sources of uncertainty are beyond the reach of researchers. Source: author.

There are additional external dimensions of uncertainty beyond the source areas. There are also degrees of uncertainty, ranging from tractable statistical uncertainties to scenario uncertainties (see next section), to ignorance (the known unknowns and unknown unknowns). There are also at least two qualitatively different kinds: epistemic uncertainty

that could be ‘cured’ with more information, and aleatory uncertainty, that of random, truly unpredictable events. Finally, there may be problems of logic. Couclelis (2003) argues that the three classic modes of reasoning: deduction, induction and abduction, as used in GIS applications and elsewhere, make it logically impossible to obtain results free of uncertainty. Basically, uncertainties related to these dimensions external to the data are the flipside of the assumptions we must make in order to develop stories, models, and forecasts, and to make decisions that go beyond the data.

3. Uncertainty and Error in Spatial Forecasting and Decision Making

3.1 Predictions, Projections, and Forecasts

For some time now, GIS&T has been a widely used platform for helping scientists, planners, administrators, industry, the private sector, the health sciences, the army and many other users to anticipate future developments of interest to them. This is usually done by building statistical or process models in areas where critical actions should be taken now to prevent undesirable outcomes in the future, to accelerate desirable outcomes, or to prepare for the inevitable. Areas of interest may include urban growth, rural and urban land use change, migrations and future population distributions, the spread of epidemics, the growth and change of transportation networks and traffic flows, ecological changes under climate change pressures, deforestation and land erosion, the evolution of river basins, and any number of other topics that have important spatial dimensions.

Even though we cannot have knowledge of the future and in many cases also of the past, we can reduce our ignorance through inferring unknown facts from known facts. The known facts in this case are largely based on the ‘pattern’ mentioned earlier, the part of reality that can be assumed to remain more or less constant over the relevant time horizon, and which usually also include trends that today appear solid. For environmental and other earth-science related topics physical laws are of course a major component of the pattern. When trying to figure out the future, data quality is critical in all cases because even small errors can get amplified during modeling, without the possibility of model validation. Moreover, all spatial forecasts, whether in physical or social domains, are also subject to the properties of space mentioned in section 2.1, with the attending possibilities for uncertainty and error if handled incorrectly. Whether modeling the present or the future, space is indeed at the core of the pattern, just as physical laws are.

The inference process from known to unknown involves techniques such as projection, forecasting, prediction (retrodiction if it is about the past), or scenario-building (see Section 3.2 below) and it offers some degree of confidence that today’s decisions and actions are not shots in the dark. Projection or trend extrapolation is the most basic method of the three and works best for short time horizons and simple problems, e.g. as in estimating student enrollments and related space needs for the next five years. Forecasting is more general and complex, and can include projections as well as more sophisticated techniques. Forecasts and projections always have some greater or lesser degree of uncertainty, since they can at best be rational attempts to guess – not to know – what the future holds in the particular area of interest. Prediction on the other hand implies much higher levels of accuracy and precision (barring human or technical errors). Predictions in the spatiotemporal domain are rarely possible outside some areas of physics such as classical

and relativistic mechanics, which enable the extraordinary engineering feats of outer-space exploration. But forecasts rather than predictions address the evolution of most physical processes on Earth, such as those studied in hydrology, climatology, oceanography, geology, biogeography, and so on. Phenomena in these domains are complicated by varying local conditions and a host of cross-impacts, and in many cases, by human interventions, making predictions in the strict sense practically impossible. (The term 'predictions' can however be used for the outputs of models, even when the latter are part of studies that generate forecasts).

Technically, most standard approaches to forecasting involve statistical, mathematical or computational process models that range from simple trend extrapolation to elaborate models of natural or socioeconomic phenomena based on complex systems science. These can range from weather forecasts for the next several days, to climate or population forecasts for the next century. Forecasts are strengthened by agreeing with several other forecasts of the same phenomena using different data and models, though dissenting outliers are always possible (and occasionally may end up being correct). Current climate models provide a good example of the strength in numbers.

An interesting case of a forecast that turned out to be totally wrong, even though it was based on good data and a good model, is the one that as recently as the late 1970s forecasted global cooling (Figure 3). There was indeed a "temperature anomaly" in the two decades that provided the temperature data. The problem was the erroneous assumption that the extremely short-time horizons used in the study to describe global temperature change (1965-1975 vs. 1937-1946) was good enough for a very long-range phenomenon.

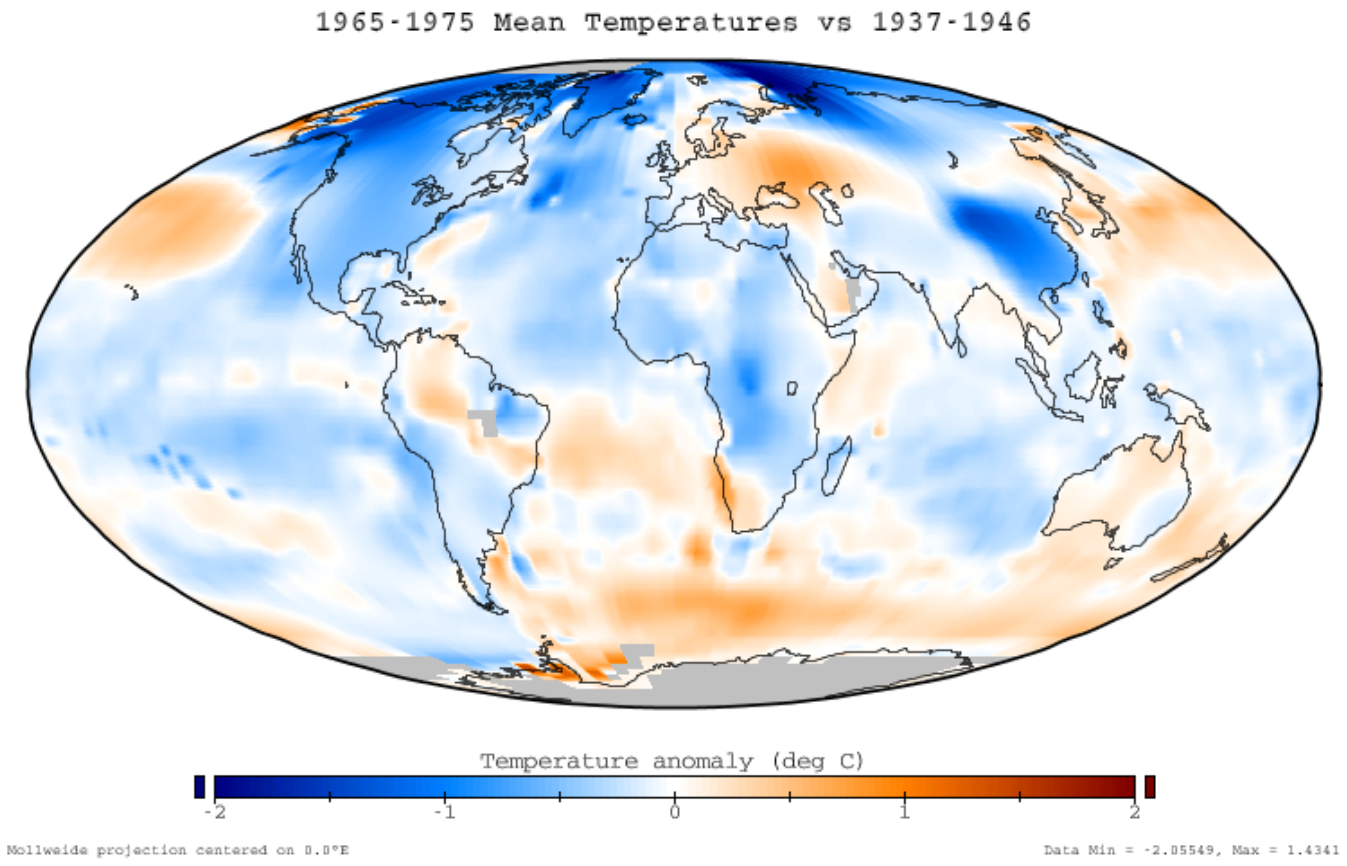


Figure 3. In the 1970s, global cooling was modeled. Source: Wikimedia Commons.

A similar case of an improper assumption about the time horizon led to incorrect flow volume figures being used in the Colorado River Compact signed in 1922. The Compact allocated 7.5 million acre-feet per year to the five states comprising the Upper Colorado Basin (Arizona, Colorado, New Mexico, Utah, and Wyoming) and another 7.5 million acre-feet to the three states in the Lower Colorado Basins (Arizona, California, and Nevada). An additional 1.5 million acre-feet was allocated to Mexico under an international treaty signed in 1944. Many studies have concluded that the period used as the basis for calculating the "average" flow of the river (1905-1922) when the compact was negotiated included periods of abnormally high precipitation, and that the lower flows experienced in many of the years since the Compact and international agreement were signed represent the more realistic long-term patterns (NRC 2007).

3.2 Scenarios

A very common way to mitigate the uncertainties of forecasts is to use scenarios. GIS&T is an excellent platform for generating and comparing scenarios because along with numerical model forecasts it generates informative visualizations that can also be understood and discussed by non-specialists. Errors in crowd-sourced data can also be easier to spot visually as anomalies on the maps displayed.

Scenarios are alternative hypotheses of what might happen, using the same model with different assumptions, usually constructed around a central trend extrapolation or ‘business as usual’ scenario. The oil company Dutch Shell is believed to have been the first to develop and use modern scenarios, starting back in the 1970s, though the military has a much longer history with scenario-based ‘strategy games’. The issue examined may be something easily quantifiable such as ‘growth’ (e.g. of demand for oil or of urban size measured in additional developed acres), which can be examined with forecasts for high, medium, or low growth under different assumptions. The urban growth model SLEUTH that was first proposed in the late 1990s (Clarke and Gaydos 1998) is based on spatial and land-use policy scenarios and has been continuously refined and expanded over the years. ArcSLEUTH is now a custom extension of ArcGIS. The model is still being used internationally to forecast growth and change in very diverse types of cities around the world.

In other cases, the issue studied may be at least in part qualitative, as in trying to figure out and compare the effects of different policies on – say – quality of life or environmental sustainability, which are very broad concepts. The scenarios in this case will involve data, maps and models where possible, but they may also be expressed as qualitative descriptions (‘stories’) based on informed ‘if…then…’ speculations. Fully qualitative scenarios are also being developed, especially in areas such as urban or environmental planning where public participation in selecting among development alternatives is desirable, and in many cases required by law. Such scenarios are sometimes called ‘story-telling’. You might think that the role of GIS&T ends at this point, but that’s wrong: Around 2010 Esri began developing a new product called ArcGIS StoryMaps (Esri 2020), which in its latest iteration allows users to ‘tell their story’ by seamlessly blending text, images, custom maps, sounds, and data as needed.

Today, qualitative scenario exercises are used by companies, administrations, public utilities, other large organizations, and the military to stimulate strategic thinking especially about the long term, facilitating brainstorming, sharpening intuition, and allowing participants to imagine, discuss and evaluate plausible futures. Next we'll see how this most low-tech among approaches to forecasting can be combined with advanced technical sophistication to support society's efforts in getting a handle on the future's big unknown unknowns.

3.3 Emerging Approaches to Future Uncertainty Reduction

'Deep' or 'radical' uncertainty are the terms used for the unknown unknowns, especially in cases where decisions must be taken now that will have important consequences in a future beyond the reach of current forecasting methods. The longer the time horizon, the more likely it is that there will be major discontinuities in the predictable patterns, that today's best models will no longer be reliable, and that any probabilities assigned to forecasts made today will be useless. A small number of novel approaches developed in the past couple of decades seek to address this problem by working with uncertainty rather than trying (in vain) to eliminate it. Their goal is to increase the resilience of our approach to the future and the question they pose is not "what will happen?" but: "given that one cannot predict, what actions available today are likely to serve us best in the future?". Plausible scenarios are still being derived (sometimes in massive numbers). But instead of picking a most probable or desirable one to recommend or act upon, scenarios here are used as part of a very large space of time-varying possibilities that is analyzed to guide the next set of recommendations and actions.

The names of the original approaches in this group, Robust Adaptive Planning (RAP), and Assumption Based Planning (ABP), reflect two key ideas of this philosophy: dynamic adaptation, and a major emphasis on assumptions. Using very large ensembles (sets) of plausible scenarios, RAP strives to identify paths of action that can serve as many desirable alternative futures as possible, while keeping one's options open as long as possible. (Example: keeping an umbrella in your car serves both alternative futures of rain and no-rain). As time goes by, these paths of action are regularly tested against reality, scenarios keep being added to and dropped dynamically from the set, and the paths of action are adjusted as needed.

ABP's method is similar, also using very large numbers of scenarios based on combinations of assumptions, but the emphasis here is on the assumptions themselves as gathered from a wide range of relevant areas (see Figure 2). The idea here is to start with formulating policies based on a number of key assumptions that currently appear valid, to keep checking their validity as time goes by, and to be able to re-orient one's strategy on the basis of new assumptions when any of the previous ones fail.

These alternative approaches to future radical uncertainty were originally applied to climate change, but they are also suitable for mega-infrastructure projects such as the development of new airports, bullet trains, new towns, energy production and distribution networks, multi-objective, long-term major river delta management programs, and many other areas. (Marchau et al. 2019).

With increasing emphasis on evidence-based policies and the continuing development of accessible artificial intelligence (AI) applications, GIS&T and other modeling platforms

should see their role strengthened in the treatment of future uncertainty as in other areas. It is already evident that practically no area of geographic information science, whether theoretical or applied, is immune to the problems of error and uncertainty, and that increasing amounts of effort are being devoted to devising effective ways of countering these problems. Among other things, a novel role for GIS&T in the search for a more transparent future could be the systematic discovery of the 'pattern in the noise' – the knowable within the unknown - on which specific forecasts are based, using machine learning methods. As mentioned in section 3.1, a fundamental part of the pattern in any spatial forecast or other approach to the future should be the properties of geographic space itself as revealed in spatial analysis and related work. Based on that foundation, GIS may also play a special role in the generation of scenarios in large numbers to feed new as well as more traditional methods of probing the future both near and distant. The newer methods in particular are very computation-intensive and require high-level analysts to develop and interpret the scenario ensembles at the core of Robust Adaptive Planning (RAP) and related approaches. A final special advantage of GIS is the capability to allow combinations of strictly quantitative and mixed quantitative and qualitative scenarios, along with strategic games and storylines, thus covering the range of possible ways of describing and analyzing error and uncertainty.

References

Bennett B. (2010) Spatial Vagueness. In: Jeansoulin R., Papini O., Prade H., Schockaert S. (eds). Methods for Handling Imperfect Spatial Information. Studies in Fuzziness and Soft Computing, vol 256. Springer, Berlin, Heidelberg.

Burrough, P. A. and Frank, A. U. (eds.) (1996) Geographic objects with indeterminate boundaries. Proceedings, GISDATA Specialist Meeting #2 on Geographical Objects with Undetermined Boundaries. London: Taylor & Francis.

Clarke, K. C., and L. Gaydos (1998) Loose Coupling A Cellular Automaton Model and GIS: Long-Term Growth Prediction for San Francisco and Washington/Baltimore. International Journal of Geographical Information Science, vol. 12, no. 7, pp. 699-714.

Cohn, A. G., and Gotts, N. M. (1996) The 'egg-yolk' representation of regions with indeterminate boundaries. In: Burrough, P., Frank, A. M. (eds.) Geographic objects with indeterminate boundaries. Proceedings, GISDATA Specialist Meeting on Geographical Objects with Undetermined Boundaries. London: Tayor & Francis, pp. 171–187,

Couclelis, H. (1996). Towards an Operational Typology of Geographic Entities with Ill-defined Boundaries. In Burrough, P.A. and Frank, A.U. (eds). Geographic objects with indeterminate boundaries. London: Taylor & Francis, pp. 45-55.

Couclelis, H. (2003). The Certainty of Uncertainty: GIS and the Limits of Geographic Knowledge. Transactions in GIS 7(2): 165-175.

Couclelis, H. (2016). The Earth as Sensor and as Enigma: Geographical Science in the Global Data Age. Cybergeo: European Journal of Geography, 20th Anniversary Issue.

Esri. (n.d.). ArcGIS StoryMaps: Storytelling that resonates. Last retrieved 2020.

Fisher, P. F., Wood, J., and Cheng, T. (2004). Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. Transactions of the Institute of British Geographers NS, 29, 106-128.

Gao, S., Janowicz, K., Montello, D.R., Hu, Y., Yang, J-A., McKenzie, G., Ju, Y., Gong, L., Adams, B., and Yan, B. (2017). A data-synthesis-driven method for detecting and extracting vague cognitive regions. International Journal of Geographical Information Science, 31:6, 1245-1271.

Goodchild, M. F. & Gopal, S. (Eds.). (1989). The Accuracy of Spatial Databases. CRC Press.

Guptill, S.C and Morrison, J.L. eds. (1995) Elements of Spatial Data Quality, International Cartographic Association. 1st Edition. Pergamon-Elsevier.

Marchau, V.A., Walker, W.E., Bloemen, P.J., Popper, S.W. eds. (2019). Decision Making under Deep Uncertainty: From Theory to Practice. Cham, Switzerland: Springer Nature.

National Research Council (NRC). (2007). Colorado River Basin Water Management: Evaluating and Adjusting to Hydroclimatic Variability. Washington, DC: The National Academies Press.

O'Sullivan, D. and Unwin, D. (2003). Geographic Information Analysis, 1st Edition. Hoboken, NJ: John Wiley & Sons.

Rumsfeld, D. (2002) 'Known knowns'. Department of Defense news briefing, 12 February.

Sui, D., Elwood, S., and Goodchild, M. eds. (2013). Crowdsourcing Geographic Knowledge: olunteered Geographic Information (VGI) in Theory and Practice. Springer Netherlands.

Wu, X., Wang, J., Shi, L., Gao, L., and Liu, Y. (2019) A fuzzy formal concept analysis-based approach to uncovering spatial hierarchies among vague places extracted from user-generated data. International Journal of Geographical Information Science, 33:5, 991-1016,

Zuefle, A., Trajcevski, G, Pfoser, D., and Kim, J.-S. (2020). Managing Uncertainty in Evolving Geo-Spatial Data. In 2020 21st IEEE International Conference on Mobile Data Management (MDM), Versailles, France. pp. 5-8.