Why Bayesian Marine Spatial Planning?

C. Goldfinger, D. Ullman, K, Halsey, C. Romsos, D. Lockett, Ann Radil, Paul Manson, Flaxen Conway

With so many spatial planning and decision support tools available today, all of which claim to solve a multitude of problems, why does there remain a need for a Bayesian Marine Spatial Planning Tool? The reasons lie at the root of the methods currently in use. Virtually all spatial planning and analysis today are built around the power of existing GIS applications. This has a lot of appeal because these systems exist, and adapting them for planning tasks seems to be an attractive and straightforward path. They handle spatial datasets of varying resolutions, handle image (raster) and point and line (vector) data at the same time, two basic requirements for such a system. They are widely available and there is a large user base with familiarity with GIS systems, more so with time as they become more user friendly. Yet despite these attractive features, GIS systems fundamentally fall short in several critical ways: They do not handle multidimensional data well; they do not account for or propagate uncertainties in the data; they do not handle temporal data well or at all, and they do not really help the user make a decision, they simply present the result of a simple "canned" analysis. These issues may not initially seem serious; however, a more detailed look at each of them is needed.

Incompatible data types: Understanding the limitations of GIS becomes more apparent when the GIS architecture is considered. GIS systems can display rasters and vectors; they are placed in the same spatial context, but in GIS systems, they do not interact very much. So all data ingested in a GIS must be reduced to either a raster image, or a vector shape consisting of lines, points or polygons. In any natural system, reducing the data to these forms is difficult at best, and impossible in many cases. For example, marine GIS layers such as bathymetry, aerial photography, satellite imagery and derivatives naturally lend themselves to GIS layering. Likewise, point sample data, areas that can be well – described by polygons and lines, such as shipping routes, jurisdictional and regulatory boundaries, and simple natural layers such as surficial geology, can be reduced to vector polygons. So far so good. But can all data be effectively fit into simple 2D rasters and polygons? Unfortunately, no. While most marine systems still have many gaps and patchy data, many available datasets today are not amenable to reduction to 2D polygons and rasters. Many biologic and oceanographic datasets reside in multidimensional databases, some of them relational databases, and many include the element of time. For example, on the US West Coast, a biological database known as the Habitat Use Database (HUD) contains information about 323 species of bottom or near bottom dwelling fish in a Relational Database Management System (RDBMS). For each species, the HUD contains information about the preferences that fish has for substrate, water depth, temperature, and other attributes where known. It also contains similar information for the life stages of the species, separating juveniles from adults. Each preference includes a "strength of affinity" measure of how strong its preferences are thought to be, and gives maximum and preferred ranges for them. With oceanographic data, a typical example is a 4 dimensional database of current velocity in 3D volume space, and time (now it's 4D), combined with other attributes such as dissolved oxygen, salinity etc. Neither of these two examples, typical in natural systems, can be reduced to 2D polygons nor analyzed in a GIS. Yet this is the basis for most CMSP systems today, the data must be "dumbed down" in order to fit into the GIS software architecture, severely compromising much of the power of the data in order to use the convenience of a GIS. Once the "dumbing down" and simplifying is done, areas can be scored for their positive attributes by simply counting up a value for overlapping polygons with positive attributes, and comparing one score to another score at another location. In this way, a map of areas that are more positive for a given analysis goal, and less positive can by constructed. This "analysis" has the appearance using the data and good scientific method, but the fatal errors were committed before the analysis began. This would be akin to equipping a Boeing airliner with a passenger car engine because they were easily available, even though it would never fly.

How good are the data? All types of data come with uncertainties, and in the marine world of patchy data, they are worse than average. Uncertainties come in many forms, and can range from insignificant to insurmountably large. In any analysis, you need to have some grasp of these errors in order to know if the analysis is valid, over what range is it valid, and is it a good basis for decision making? GIS systems inherently do not handle uncertainties at all, a significant problem when regulatory decisions are to be made, or when a rigorous analysis is needed. These two concepts go hand in hand, a good analysis is a *requirement* of a good decision if the decision is to be science based, and even more so if the decision has legal and societal implications that may last for many years.

What about time? Everything changes, the ocean environment changes constantly on many timescales, with the biological, oceanographic and geological elements of the system constantly in flux in ways we are only beginning to grasp. Understanding such systems at even the most basic level requires consideration of time. From seasonal to decadal change, planning involves some projection of what things will be like in the future. GIS systems, in the process of reducing the data into polygons and rasters, usually lose the element of time along with other information in order to fit the required mold. How can temporal changes and trends be considered in such systems? Most often, they are not.

What to do? The solution for these issues is not as insurmountable as it might seem. What is needed is to add the spatial power of GIS, to a processing engine that can handle the three missing components: complex data, uncertainties, and time. In addition, we need to add a fourth element, that is, the capability to help the users actually reach a consensus decision once the scientific analysis is complete.

One solution that has been applied effectively is to build an analysis engine based on Bayesian analysis methods. Bayesian analysis is a simple and straightforward way to incorporate uncertainties, time, complex and patchy data, or even missing data, into a robust analysis system that can also report the robustness of the outcome and how sensitive the outcome to any particular piece of data. This allows the user to know which data are important, which can be ignored, and which would help the most in making a decision more robust. The Bayesian system fundamentally combines probabilities in conceptually the same way a GIS combines rasters and vectors, but with a rigorous method replacing the ad-hoc additive method used in GIS analysis. Lastly, a Bayesian system can also be used in the final stage of decision-making, allowing users to engage in "what if" scenarios, and input their subjective values into the decision. By having all cards on the table, with a robust science based foundation, consensus building is greatly simplified. Lastly, the output can be visualized in a GIS system, so that the users can view the outcomes, the underlying data, and the analysis results in an intuitive way.