International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Modelling Ecosystem Service Flows under Uncertainty with Stochastic SPAN

Gary W. Johnson [a], **Robert R. Snapp** [b], **Ferdinando Villa** [c], **& Kenneth J. Bagstad** [d]

[a] *Gund Institute for Ecological Economics and Department of Computer Science, University of Vermont, Burlington, VT, USA (gwjohnso@uvm.edu)*

[b] *Department of Computer Science, University of Vermont, Burlington, VT, USA*

[c] *Basque Centre for Climate Change, IKERBASQUE, Basque Foundation for Science, Bilbao, SPAIN*

[d] *Rocky Mountain Geographic Science Center, U.S. Geological Survey, Denver, CO, USA*

Abstract: Ecosystem service models are increasingly in demand for decision making. However, the data required to run these models are often patchy, missing, outdated, or untrustworthy. Further, communication of data and model uncertainty to decision makers is often either absent or unintuitive. In this work, we introduce a systematic approach to addressing both the data gap and the difficulty in communicating uncertainty through a stochastic adaptation of the Service Path Attribution Networks (SPAN) framework. The SPAN formalism assesses ecosystem services through a set of up to 16 maps, which characterize the services in a study area in terms of flow pathways between ecosystems and human beneficiaries. Although the SPAN algorithms were originally defined deterministically, we present them here in a stochastic framework which combines probabilistic input data with a stochastic transport model in order to generate probabilistic spatial outputs. This enables a novel feature among ecosystem service models: the ability to spatially visualize uncertainty in the model results. The stochastic SPAN model can analyze areas where data limitations are prohibitive for deterministic models. Greater uncertainty in the model inputs (including missing data) should lead to greater uncertainty expressed in the model's output distributions. By using Bayesian belief networks to fill data gaps and expert-provided trust assignments to augment untrustworthy or outdated information, we can account for uncertainty in input data, producing a model that is still able to run and provide information where strictly deterministic models could not. Taken together, these attributes enable more robust and intuitive modelling of ecosystem services under uncertainty.

Keywords: ecosystem service flows; probabilistic modelling; SPAN; service path attribution networks; uncertainty propagation; Bayesian networks

1 INTRODUCTION

The lens of ecosystem services (ES) provides researchers and decision makers with a framework for assessing the contribution of ecosystems to human well-being [Millennium Ecosystem Assessment (MA), 2005]. With growing interest in incorporating ecosystem services into decision making, there is a greater need for quantitative methods and tools to model the complex relationships between ecosystems and human needs and values.

To date, numerous researchers have used spatial data to map and value ecosystem services for particular case studies [Troy and Wilson, 2006; Raudsepp-Hearne et al., 2010]. In recent years, specific computing tools have emerged to automate the mapping and valuation of ecosystem services [Kareiva et al., 2011]. Many researchers have noted that the provision and use of ecosystem services often incorporate different temporal and spatial scales [Ruhl et al., 2007; Tallis and Polasky, 2009]. Yet aside from some hydrological service models, researchers have often ignored that ecosystem services can only be realized when there exists a direct measurable connection between ecological resources and human beneficiaries. We address the problem of identifying these connections using the Service Path Attribution Networks (SPAN) model [Johnson et al., 2012].

Although quantitative ecosystem service models are increasingly in demand for decision making, the data required to run many of these models, particularly those which are spatially explicit, are often *patchy*, *missing*, *outdated*, or *untrustworthy* (as defined in section 2.2). Without a mechanism for addressing this fundamental problem, many modelling efforts will face difficulty from the start. Furthermore, many ecosystem service models are intended to be used in a policy or decision-making context rather than for scientific research. In order to strike a balance between high-level management questions and low-level biophysical and social modelling, a variety of simplifications may be used, which can increase the uncertainty in model results. Communicating this uncertainty intuitively to decision makers thus becomes of paramount importance in order to promote better informed decisions.

In this work, we introduce a systematic approach to addressing both the data gap and uncertainty communication problems presented above. To demonstrate its applicability to existing models, we apply these techniques to create a stochastic adaptation of the Service Path Attribution Networks (SPAN) framework for spatial ecosystem service flow assessment [Johnson et al., 2012]. Results from a scenic viewshed assessment in southern Arizona's San Pedro River watershed suggest that much of the potential visual blight on the landscape has little to no impact on residential viewsheds.

2 METHODS

To overcome the problems of data gaps and uncertainty communication, we apply the following probabilistic approach. First, we use Bayesian belief networks to address issues with patchy or missing data. Next, we attach uncertainty information to outdated or untrustworthy data sets in a way that is compatible with the Bayesian networks we used in the previous step. Finally, we alter the previously published SPAN simulation model to accept probabilistic inputs and to propagate uncertainties throughout its calculations. Taken together, these tools enable us to generate probabilistic results that account for the combined effects of data and model uncertainty.

2.1 Service Path Attribution Networks

The SPAN formalism assesses ecosystem services through a set of up to 16 maps, which characterize service delivery within a study area in terms of flow pathways between ecosystems, their human beneficiaries, and landscape features or human activities which may obstruct these flows. To create a SPAN model for a particular service, we first determine the form of matter, energy, or information – called the *service medium* – which will travel from an ecosystem to a human beneficiary and, upon arrival, transmit either a benefit or detriment. We include both benefits and detriments because although some flows may be positive to users (e.g., food, fuel, fiber), others may be either beneficial or hazardous, depending on the context (e.g., flood, wildfire). The strength of focusing on the service medium, its flow, and its effect on people is that it provides a unifying framework for quantitative simulation that is applicable both to services with a biophysical transport component and to information-based services (e.g., aesthetic landmarks and cultural resources).

Next, we estimate the amount of this service medium that each location in the study area can produce or absorb during a given time period. Locations that exhibit a positive production value are called *sources*. Those which possess a positive absorption capacity are labeled *sinks*. Potential human targets of the service medium are also identified along with their associated demand or vulnerability, called the service *use*. Finally, we represent the service medium as a collection of agents, each with a weight value indicating the quantity of the medium it carries. These agents are directed from their origins at ecosystem sources through the study region by means of a user-specified *movement* function. If an agent encounters sinks along its path, its weight may be reduced by the absorption capacity remaining in the sink. When an agent encounters a use location, the SPAN algorithm records its remaining weight, the route followed from its origin, and all sink effects accrued along the way. If the service use by the human beneficiary at that location is *rival* (e.g., destructive consumption of the resource), then the agent's weight will be further reduced by the demand associated with the use location (Figure 1).

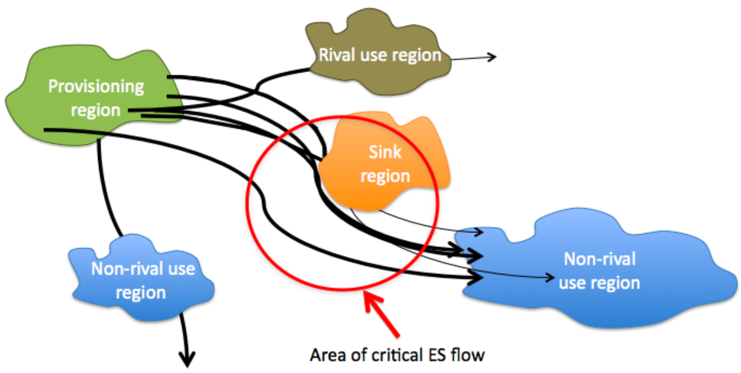


Figure 1: Source, sink, use, and flow relationships in the SPAN formalism

In all cases, agents continue to follow their movement functions until either their remaining weight reaches zero, they leave the study area, or they become trapped in a local attractor of the specified movement function. Should any of these three conditions be met, the agent is removed from the simulation. Once all agents have been terminated, the information recorded on the use locations is passed to 16 post-simulation analysis functions, each of which produces a map showing the spatial distribution of one aspect of the service delivery. For a detailed description of these 16 output maps and the rationale behind using an agent-based approach for ecosystem service modelling, see Johnson et al., 2012.

The SPAN framework, as originally described in Johnson et al., 2012, uses deterministic source, sink, use, and movement functions for each service. Therefore, the 16 post-simulation analysis functions are all deterministic. Thus, a SPAN model cannot be run in conditions of patchy or missing data and cannot capture or express in its outputs the uncertainty associated with outdated or untrustworthy inputs. In the following sections, we address these shortcomings by proposing a SPAN revision that replaces these deterministic functions with probabilistic methodologies.

2.2 Bayesian Belief Networks

In our first adaptation, we replace the deterministic source, sink, and use functions run at each location in the study area with conditional joint probability distributions. That is, rather than computing the source value at each location as $s = f(x_1, x_2, \ldots, x_n) \in \mathbb{R}_+$, we would represent it as $p(s|x_1, x_2, \ldots, x_n)$, where x_1 through x_n are the source-specific features measured at this location. The same holds true for both the sink and use functions. However, as n – the number of input features – increases, calculating these joint probability distributions quickly becomes intractable. This problem can be remedied by introducing Bayesian belief networks.

A Bayesian network is a graphical representation of a joint probability distribution that has been factored into a directed acyclic graph, in which nodes are conditionally-dependent only on those nodes with which they share edges [Pearl, 1988]. This means that if we can describe our source, sink, and use functions as a hierarchical tree of sub-functions, then a Bayesian network can be used to represent this factored distribution and efficiently compute the marginal distribution of the toplevel source, sink, or use node in each network.

By focusing on the conditional probabilities between dependent nodes in a Bayesian network, we can capture uncertainty in the relationships between a location's input features. Although this is a concern in many modelling efforts, the specific intent of this paper is to address the issue of model input data being patchy, missing, outdated, or untrustworthy. We examine each of these four cases and show how we can use Bayesian networks to overcome these hurdles.

Patchy data. Patchy data occur when information exists for part of a study area but not for the entire region. A common example in ecosystem service assessment might be that hydrological services need to be modelled over a large watershed, but only some of its sub-watersheds have been studied in sufficient detail. The remaining ones only have data for a subset of the input features to the model. In this instance, Bayesian networks can be trained to recognize the conditional relationships between input features in the highly instrumented sub-watersheds. Then because of its ability to propagate evidence bi-directionally along edges (a consequence of Bayes' Theorem), the trained network can be given the partial information available in the remaining sub-watersheds and asked to provide probability distributions for the values of the missing features. In this way, we can probabilistically fill gaps in our data sets.

Missing data. High-quaity training data are not always available for every study region. If necessary input data are entirely missing and a previously trained Bayesian network is unavailable, we can assign a prior distribution to this missing feature's node in the network based on accepted knowledge or information from local experts. For example, a location may lack the greenhouse gas emissions measurements needed to model regional carbon balances. However, the local energy bureau can provide a probability distribution for per capita emissions within the region. Using a population density map along with this information, we can construct a prior distribution for greenhouse gas emissions, which can then be assigned to a node in the Bayesian network model. If the spatial variability of this distribution is partially known (e.g., emissions are 20% greater in the north-eastern quarter of the study area than in the remaining three quarters), then a different prior distribution may be applied at each location in the region.

Outdated data. Because of the expense associated with collecting and processing spatial data, in many cases some of the model input features will have readily available data that were collected reasonably recently, but the remaining features will only have measurements that are dated. We have already described two ways to treat this case. We could train a Bayesian network on a more recent data set for the outdated features and then use it to estimate their distributions in the study region. Alternatively, we could discard the outdated maps and request prior probability distributions from local experts.

A third approach that combines the outdated, mapped information with expert knowledge is to ask experts to construct transition probabilities from the states in the map to their current states. For example, for a feature with states in A, B, C and mapped value A, replace A with a conditional probability table, such that $p(A|A) + p(B|A) + p(C|A) = 1$. If the number of possible states for a feature is infinite (e.g., a real-valued continuous variable), then we can replace the mapped value with a continuous probability distribution. For example, for a feature with states in $\mathbb{R}$ and mapped value r, ask experts to select a σ and replace r in the map with $\mathcal{N}(r, \sigma)$. If a normal distribution is not a good fit for a given feature, select a different distribution function as appropriate.

Untrustworthy data. Untrustworthy data can be treated similarly to outdated data as described above. Trained networks, expert-provided priors, or expert-provided transition functions are all possible solutions to attaching explicit uncertainty to the data set in question. However, one important difference is that expert-provided transition functions will estimate measurement or recording error by the untrustworthy data source in this case. In the situation of outdated information, these transition probabilities would represent the likely change over time since the point of measurement.

2.3 Uncertainty Propagation

Now that we have a formal methodology for overcoming various issues with input data limitations, we must address the problem of propagating the uncertainty in our inputs through our SPAN simulation. As described in the previous section, our source, sink, and use functions are each replaced with a Bayesian network which captures the uncertainty in each location within our region. The final piece needed for our SPAN ecosystem service model is a movement function to propel the service medium from its source location across the study area.

As described in section 2.1, an agent's initial weight is the source value at its origin. Because we are using Bayesian networks for this value, the weight will be represented as a probability distribution. Whenever an agent encounters a sink, its weight w will be depleted by the absorption capacity at that location s as $w - s$. Although this was a simple arithmetic calculation in the deterministic SPAN model, this becomes a potentially complex probabilistic calculation in our new stochastic model. The same operation will occur when an agent encounters a rival use location, in which its weight w is reduced by the local destructive consumption value u as $w - u$.

To enable these calculations on probabilistic values, we assume that each source, sink, and use value is represented as a mutually independent, normally distributed random variable with mean μ and variance σ^2 inferred from the Bayesian networks. Standard error propagation is applied when evaluating the arithmetic and comparison operators needed by the movement function and the post-simulation analysis functions. For example, given two such random variables X and Y, represented as mean/variance pairs (μ_x, σ_x^2) and (μ_y, σ_y^2) respectively, their difference $X - Y$ returns $(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$. The means of these random variables will look like the values in a strictly deterministic simulation. However, their variances will grow and shrink over the course of sequential operations, thereby propagating a rather simple but highly efficient metric of the uncertainty associated with each intermediate output.

2.4 Case Study Example

To illustrate the methods outlined above, we constructed Bayesian networks for the source, sink, and use functions for a scenic viewshed assessment in southern Arizona's San Pedro River watershed. The uncertainty propagation approach was compared with a strictly deterministic run on the means from the Bayesian networks to validate the probabilistic arithmetic described in section 2.3. All 16 output results as well as the probabilistic inputs used in this simulation were displayed as map pairs, showing means and standard deviations per pixel respectively. For details on how to interpret these results, see Bagstad et al., 2011.

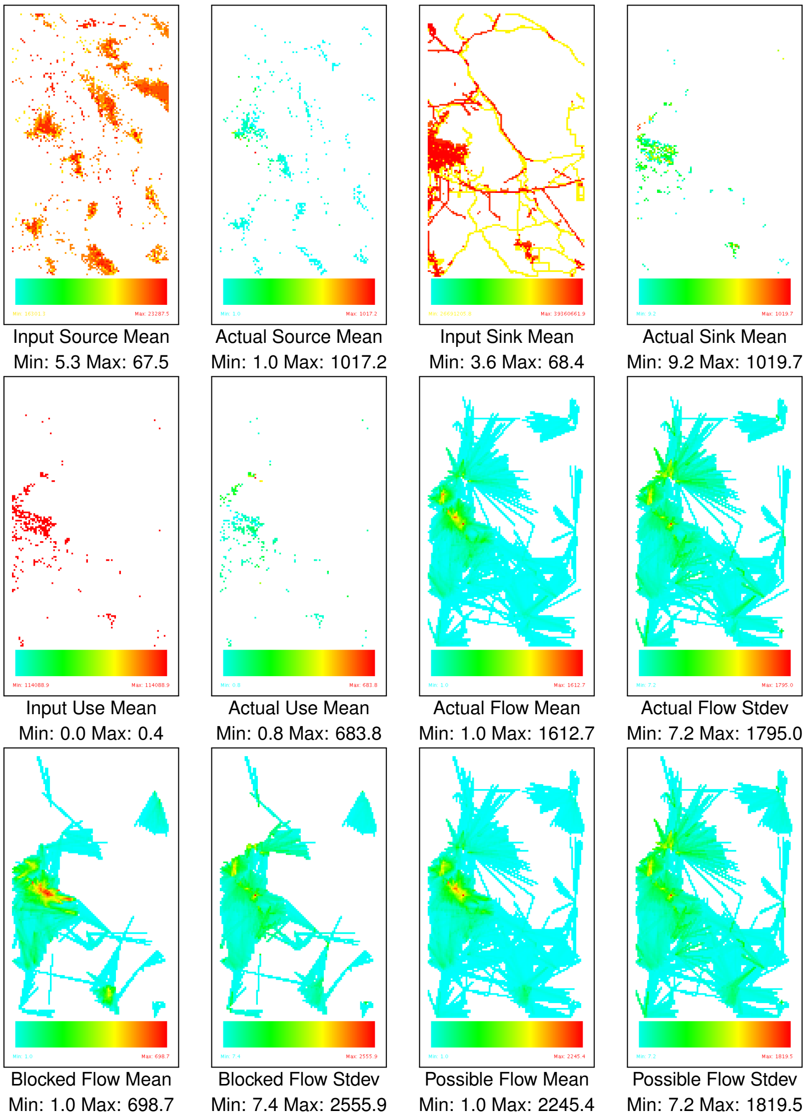
This scenic viewshed was analyzed on a raster grid with 160 rows and 96 columns (15360 cells total). Each cell measured 1161.31 meters in width by 1356.77 meters in height for a total study area size of roughly 217 km x 111 km. Means for source (scenic beauty) and sink (visual blight) were ranked from 0 to 100 in each cell. Means for use were in the range from 0 to 1, indicating the presence or absence of potential view beneficiaries. To reduce background noise and simplify analysis of the results, all source and sink values with means below 25 were excluded from the simulation. Because of the non-rival and non-material nature of view transmission along lines of sight, both sources and sinks were considered inexhaustible for the purposes of this model. That is, sources could provide their full scenic beauty value to any number of users, and sinks could impair any number of lines of sight.

3 Results

We ran the SPAN simulation on a machine with 4GB of DDR SDRAM and an Intel(R) Core(TM)2 Duo CPU clocked at 2.13GHz. The first run used Bayesian network inputs approximated with mean/variance pairs, and the second pass used only the means and a strictly deterministic SPAN implementation. The output results for the mean value maps were identical in both runs, and the runtime performance was only slightly slower in the probabilistic simulation – roughly 15 minutes 3 seconds versus 12 minutes 42 seconds on average over 10 runs apiece. The real advantage of the explicitly probabilistic run was that each of the 16 outputs contained maps not only of the mean value per pixel but also of the standard deviation associated with each such value. We show here only a subset of the 32 output maps (16 maps of mean cell values plus 16 corresponding uncertainty maps, Figure 2) returned by the stochastic SPAN simulation. Refer to Johnson et al., 2012 for a detailed discussion of each SPAN output.

These results show that although most use locations receive scenic view benefits and a large percentage of source points contribute to the quality of these views, only a relatively small fraction of all the sinks on the landscape actively reduce scenic view quality. This is noteworthy given the prevalence of sinks in the region (compare the input and actual sink maps above). Whereas the actual flow maps depict the line-of-sight connections between source and use points, the blocked flow maps show which view paths are impaired by sinks. Finally, the possible flow maps show the expected view quality in the absence of sinks (i.e., if the blocked flow were to be made available to users). The value of the flowpath-based approach to ecosystem service assessment is that for any service we model, we can make these distinctions between those source, sink, and use locations which actually produce, receive, or interrupt the flow of benefits or detriments from ecosystems to people and those which have no effect on ecosystem services in their region. This can have profound impacts on any management decision undertaken on the basis of these results.

Figure 2: Selected Inputs and Outputs from San Pedro Aesthetic View SPAN simulation



4 Conclusion

The probabilistic approaches outlined in this paper provide a generalizable set of tools for addressing issues with patchy, missing, outdated, or untrustworthy data. Their application using the SPAN model illustrates how these tools can be used to enable ecosystem service models to run in data-limited conditions that were previously beyond the operational scope for quantitative simulation.. By propagating uncertainty in the input data through to the results, we were able to visualize a spatial distribution of the relative uncertainties associated with each pixel in our 32 output maps. We hope that by showing mean and uncertainty values side by side, we can help decision makers to see which parts of our study areas favor strong assertions and which ones are best slated for more precautionary decision making or further research and data collection.

Only one ecosystem service application was shown and its result set was truncated. We also limited our discussion to strictly deterministic movement functions rather than the much more interesting class of stochastic movement functions. The SPAN extensions described work for all classes of ecosystem services handled by the original algorithm.

Future research directions include addressing spatio-temporal interdependence between Bayesian network outputs, incorporating feedback loops and service delivery thresholds into the models, and seeking more efficient ways to represent and analyze flow path information so that SPAN models may be run over larger scales and higher resolutions without requiring expensive computing resources.

References

Bagstad, K. J., F. Villa, G. W. Johnson, and B. Voigt. *ARIES Artificial Intelligence for Ecosystem Services: A guide to models and data, v1.0 beta*. The ARIES Consortium, 2011.

Johnson, G. W., K. J. Bagstad, R. R. Snapp, and F. Villa. Service path attribution networks (SPANs): A network flow approach to ecosystem service assessment. *Special Issue on "Analysing, Modelling and Visualizing Spatial Environmental Data" International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, 3(2), 2012.

Kareiva, P., H. Tallis, T. H. Ricketts, G. C. Daily, and S. Polasky, editors. *Natural Capital: Theory and Practice of Mapping Ecosystem Services*. Oxford Biology. Oxford University Press, New York, first edition, June 2011.

Millennium Ecosystem Assessment (MA). Millennium ecosystem assessment: Living beyond our means - natural assets and human well-being. Technical report, Washington, D.C.: World Resources Institute, 2005.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1 edition, September 1988.

Raudsepp-Hearne, C., G. D. Peterson, and E. M. Bennett. Ecosystem service bundles for analyzing tradeoffs in diverse landscapes. *Proceedings of the National Academy of Sciences*, 107(11):5242–5247, March 2010.

Ruhl, J., S. Kraft, and C. Lant. *The law and policy of ecosystem services.* Island Press: Washington, DC, 2007.

Tallis, H. and S. Polasky. Mapping and valuing ecosystem services as an approach for conservation and natural-resource management. *Annals of the New York Academy of Sciences*, 1162:265–283, 2009.

Troy, A. and M. Wilson. Mapping ecosystem services: Practical challenges and opportunities in linking GIS and value transfer. *Ecological Economics*, 60(2):435–449, December 2006.