Catherine R. Emond

Guiding Scenario Exploration in the PECAS Model Output Interface: Summary of Research in Progress

Background:

There is concern that California's complex infrastructure is being fragmented by state agencies applying data, assumptions and processes without regard for compatibility with one another. Secondly, the traditional separate models for urban land use, transportation and urban economic systems do not support comprehensive understanding of the interactions of these systems contributing to California's infrastructure. This has prompted development of an integrated modeling platform in an attempt to support a cohesive effort by all state agencies to maintain and improve California's infrastructure. It is believed that an integrated land use/transportation model will help California's infrastructure in two ways: first to coordinate the collaborative efforts amongst the agencies responsible for California's infrastructure and second to provide useful regional data to state planners and in turn state policy makers. (McCoy, 2007)

The PECAS model, which stands for "Production Exchange Consumption Allocation System" was chosen for a number of reasons. As an input-output microsimulation model based on economic allocation theory, it will forecast land use and material movement changes over time, which is important when studying the effects of policy implementation. The implementation of PECAS will encourage data sharing and standardization between agencies and MPO's, further encouraging cohesive infrastructure development. PECAS when used in conjunction with the California State Transportation model will allow the analysis of the effect of interregional transportation policies on California's economy and land use that cannot be forecast with the traditional 4-step transportation model alone. (Johnston & McCoy, 2006)

PECAS is a joint project of the University of California, Davis, and the University of Calgary. During 2006-2009, three stages of the model will be developed; the Set-Up model, which is being completed; the Demonstration model, and finally the Working model in 2009 (Johnston, 2007). I am the human computer interaction (HCI) specialist of the PECAS output interface development team. This paper details the process I am using to define PECAS user needs and the output data presentation format.

Output Interface Design Motivation:

The extensive collection of indicators that the PECAS model uses will result in extremely large output sets. We envision the output to be used "… by agencies to evaluate land use policies, test transportation investment scenarios, and to evaluate compliance with various legal mandates." (Johnston & McCoy, 2006) In order to optimally support these uses, the output interface will emphasize visual presentation of the data in the form of maps and graphs, with data in tabular form available for more in-depth exploration of the outputs. The interface, to be accessed through the internet, will allow agencies to efficiently and effectively explore, analyze, synthesize and present PECAS output data without having to invest in standalone geovisualization software.

Before the PECAS output interface can be realized, the following steps must be undertaken:

1. Derive guidelines from existing visualization design theory to direct PECAS interface design,
2. Identify potential users of the interface,
3. Define the user tasks the interface must support,
4. Select the output data needed to support these user tasks,
5. Define the visualizations required to optimally present this data.

Implementing these five interface design steps in conjunction with the model database development and model assembly and calibration will help the PECAS design team determine what file formats and software will be required to create an output interface that will effectively support scenario analysis.

Interface Design – Theory and Methodology:

Since the output data consists of both statistical and cartographic components, I am drawing on heuristics and design theory from both information visualization and geovisualization research to develop PECAS output visualization design guidelines. Computing advances have allowed spatial-geographical models such as PECAS to become so complex that their parameters can never be completely validated against existing data (Batty, Steadman & Xie, 2006). However, visualization of model data allows the viewer to more easily holistically understand and explore these complex systems being modeled; this allows problem domain experts to be able to determine if the model output – even if it cannot be completely tested – follows trends that match their predictions and expectations (Batty, Steadman, & Xie, 2006; Keim, Panse, & al., 2005)

The geovisualization research process commonly used to define visualization design (DiBiase, 1990; MacEachren, 1995) consists of four steps: data exploration, analysis, synthesis and presentation. Designing the PECAS interface to support all four steps is a challenge since much of the current geovisualization research that I am basing the interface design guidelines on only focuses on the first two steps: data exploration and analysis (Robinson, Submitted). Since synthesis and presentation of model output is crucial to the policy analysis process, designing the output interface to support these last two steps is of main concern in the PECAS interface design.

PECAS Users:

Three levels of users are expected to access the PECAS output data. They are defined relative to their modeling experience: expert users (modelers); intermediate users (non-model domain experts such as policy analysts,); and novices (stakeholders such as policy makers and general public). Model proficiency also identifies the amount of interaction each user level will have with the model software, with the highest proficiency level having the most interaction.

User Tasks:

At this stage of PECAS development, visualizations are being used in two different areas: to help create the model, and to develop tools for scenario analysis. Since PECAS is being created by a diverse team with members and stakeholders having different levels of modeling expertise, visualizations of the output data have become critical in communicating, assembling and calibrating the model. Modelers (or expert users) in the PECAS development team can explore visualizations and tables of both the input and output data in order to calibrate and assemble the model, and then present domain experts involved in PECAS development with synthesized visualizations of the output to help these team members understand the model.

For the output interface design, user interaction with the data is also being determined by model proficiency. In terms of DiBiase's (1990) geovisualization research process steps listed earlier, expert users will interact with the model to generate visualizations required by other users for scenario exploration. Intermediate users will have access to this set of data visualizations for scenario analysis. They will in turn summarize, or synthesize their analysis for presentation to novice users. Consequently, the output presentation needs of policy analysts are very important since their interpretation and presentation of the output will often be the only scenario data viewed by policy makers and the general public.

Selecting the Output Data:

Although geovisualization is often associated with exploratory spatial data analysis (ESDA), where the user is expected to sift through all the model output in the hope that the model pattern will emerge naturally, the PECAS output data will undergo preliminary analysis and manipulation in order to facilitate user data exploration. As models such as PECAS become more complex, it has been recognized the amount of data available for exploration is too immense to facilitate effective ESDA and that selecting the output needed to study higher-level analytic questions such as encountered in policy analysis is a more effective use of the model. (Amar & Stasko, 2005)

The PECAS interface design is a complex project due to a number of factors: the diversity of users, complexity of data, multi spatial scales, and temporal aspects. The lengthy computation times and vast amounts of storage space needed to generate and store output for all possible combinations of PECAS input prevent permanent or dynamic generation of every feasible graph and map file in the initial model output.

Visualizations of the PECAS output data can be separated by computation time into three categories: visualizations with short enough computation times to be generated on demand; visualizations crucial for scenario analysis with lengthy computation time; and visualizations not crucial for initial scenario analysis with lengthy computation time.

To reduce the amount of output, but still allow preliminary scenario analysis that satisfies all involved state agencies, a limited set of questions is being developed to help determine which outputs are essential for successful scenario analysis and which are nonessential. We will use this question set to decide which visualizations will be offered for immediate access, and which ones due to computation time and storage requirements are only available by user request.

Three sources are being used to determine the questions answered by the model output: Feedback collected from participants in a series of non-technical workshops conducted by Michael McCoy (2007) on the kinds of information state planners and policy analysts hope PECAS will provide; Robert Johnson's (2007) research on indicators for sustainable transportation planning; and Caltrans California Regional Blueprint Planning (Blueprint) progress indicator definitions (California Center for Regional Leadership in Partnership with

Caltrans, 2007). Once we have defined this set of questions, the workshops' participants and other representatives from state agencies that have expressed interest in PECAS will then be invited to review the questions to help determine whether we have correctly anticipated potential model user needs.

Defining the Visualizations:
The finalized question set will then be used to build statistical and cartographic visualizations for display and user interaction in the output interface. Since the model forecasts both spatial and temporal processes, visual techniques where a number of frames, each portraying a different temporal or spatial snapshot are used, have been found to be effective in conveying a sense of change in time and space. This can be done statically by presenting a series of separate maps (Tuft's small multiple technique), or by animating the frames. (Tufte, 1990; Batty, Steadman & Xie, 2006; MacEachren, 1995) Another important visualization approach is multi linked windows, where each window presents a different visualization of the variables, for example a scatter plot and a chloropleth map of the same phenomenon, and selecting a specific object in one window also links and highlights the same data object in the other window (Batty, Steadman & Xie, 2006).
These visualization strategies are only effective if they are presenting output variable relationships needed by the interface users to explore, analyze, synthesize and present the results of the model analysis. In order to determine whether the output variable relationships chosen for visualization are suitable, they will be evaluated by the same respondents that reviewed the original questions as to the effectiveness of the geovisualizations in answering the question set. Modifications to the visualizations will be guided by respondent feedback.

Future Design Strategy:
At this step in the design, software tool options for developing the output interface are still being explored. We feel confident that once the question set and corresponding geovisualizations have been developed, we will have a much clearer understanding of user needs, making the software tool choice easier.

References

Amar, R. A., & Stasko, J. T. (2005). Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics, 11*(4).

Batty, M., Steadman, P., & Xie, Y. (2006). Visualization in spatial modelling. In J. Portugali (Ed.), *Complex artificial environments*. Berlin, Germany: Springer.

California Center for Regional Leadership in Partnership with Caltrans, C. U. C. D. (2007). *California Regional Progress Report*

DiBiase, D. (1990). Visualization in the Earth Sciences. *Earth and Mineral Sciences, 59*(2).

Johnston, R. A. (2007). *Indicators for sustainable transportation planning*. Paper presented at the TRB Annual Conference 2008.

Johnston, R. A., & McCoy, M. C. (2006). *Assessment of integrated transportation/ land use models, final report*: Information Center of the Environment. Department of Environmental Science and Policy. University of California, Davis.

Keim, D. A., Panse, C., & al., (2005). Information visualization: Scope, techniques and opportunities for geovisualization. In J. Dykes, A. M. MacEachren & M. J. Kraak (Eds.), *Exploring geovisualization*. Amsterdam: Elseviers.

MacEachren, A. (1995). *How maps work: Representation, visualization and design*. New York: Guilford Press.

Robinson, A. C. (Submitted). Synthesis in geovisualization to support epidemiology. *Information Visualization*.

Tufte, E. R. (1990). *Envisioning information*. Cheshire, CN: Graphics Press.