Hierarchies of sustainability in a catchment

N. Dunstan
School of Science and Technology, University of New England, Australia

Abstract

This paper investigates notions of sustainability in a catchment area represented by a multidimensional spatial data set. Predictive catchment Land Use, Cover and Change (LUCC) models simulate the environmental and economic effects of land use over time. These predictions can be used to form a data cube of aggregated data tables, where the aggregation measures indicate issues of sustainability across different data dimensions and at different levels of summarization. As well as a global (or catchment-wide) concept of sustainability, measures of sustainability can be provided by region, farmer, land use, and other data dimensions. A hierarchy of sustainability is apparent when data dimensions are hierarchical.
Keywords: Land Use Cover and Change models, sustainability indicators, data cubes, decision support, discovery-driven exploration, catchment management.

1 Introduction

Catchments are complex systems of inter-related elements. Park and Alexander [1] refer to economic, environmental and social sustainability in a catchment. Brown [6] describes knowledge resources for decision-making in the context of complex sustainability problems. Scott [12] lists over 30 sustainability indicators for whole-farm sustainability. Li *et al.* [8] used 12 variables to assess region sustainability in a catchment. Land use, Cover and Change (LUCC) models have been widely used to predict the affects of land use (e.g. Beverly *et al.* [2]) and to provide support for catchment management and planning (e.g. Choi and Deal [7], McColl and Aggett [10]). Catchment models are typically based on principles of hydrology and land use economics. These principles are applied to a grid of blocks representing the catchment at a certain spatial resolution. Land use, rainfall, ground water salinity, depth of ground water table and other data are stored for each block of land.

Simulations representing periods of years generate large amounts of data that could indicate sustainable or unsustainable land use. The catchment model used for this project permits hypothesis testing by setting land use patterns and generating periodic snapshots of key data. Aggregation measures over this data set can be used as sustainability indicators, that is, parameters to user-specific sustainability functions. These aggregations include minimum, maximum, average, sum and count over block values for depth of water table, profit, runoff and salinity.

This paper describes a framework for exploring patterns of sustainability in a catchment based on a data cube of the catchment model's output data. The data cube contains aggregated measures relevant to sustainability in the catchment. A data cube is a data structure at the core of On-line Analytical Process, a data mining technique for exploring high-dimensional data sets (Gray *et al.* [3]). The goal is to provide answers to questions like:

Is the current pattern of land use over the catchment sustainable?;
Is cropping sustainable in the North-eastern region above 300 meters?;
Is the current land use on the farm "Roxbury Downs" sustainable?.

The data cube permits summarizations across different combinations of dimensions and at different levels of detail. Hence it is possible to investigate sustainability from many perspectives.

The next section describes the catchment model, its data dimensions and operation. This is followed by a description of the data cube. Its use in searching for sustainable patterns of land use is discussed.

2 Representing a catchment

2.1 Data dimensions

Consider the catchment data structure to be a two dimensional grid of points. Each point represents a block of land and has associated values for each of a number of data dimensions, such as:

- Land Use
- Locality
- Land Owner
- Rainfall
- Elevation
- Ground Water Salinity Level
- Depth of Ground Water

Some dimensions have natural conceptual hierarchies. For example, Locality is divided into regions. Within regions there are sub-regions. Numerical dimensions, such as Rainfall, Elevation, Salinity level and Depth of Ground Water table may form a hierarchy by successively dividing the range into bins. It is desirable from an analytical point of view that there be sufficient levels in hierarchies to be able to select areas of interest.

2.2 Catchment model

Catchment models are based on principles of hydrology, soil science and economics. For example:

- different land uses have different hydrological characteristics:
    - cropping has a relatively low evapotransporation rate;
    - forest has a relatively low runoff per block (Liu *et al.* [9]);
- recharge increases the depth of the water table;
- runoff effects down-hill neighbors;
- high ground water salinity and water table diminish yield (Griener and Cacho [4]).

In this work, a model was used to generate snapshots of a catchment including the depth of the water table, profit and runoff for each block, for each simulated year. It is a large and multidimensional data set. This data set can be used to provide sustainability indicators. For instance, a measure of environmental sustainability might focus on the minimum depth of the water table. Likewise, average annual profit could be used in a measure of economic sustainability.

3 Data cube

The areas of interest for sustainability are well expressed as combinations of data dimensions at all levels in their hierarchies. For example: All blocks in the North West sub region belonging to Farm "F1". A data cube is a way of storing sustainability indicators that permits searching through dimensions and hierarchies. There are $\prod(L_i + 1)$ nodes in a data cube, where L_i is the number of levels in the hierarchy of dimension i (Han and Kamber [5]). For example, consider this data set of 3 dimensions.

- Region
- Farm and
- Land Use.

The levels of partitioning of blocks of the catchment using these data dimensions are given in Figure 1.

The dimensions Farm and Land Use have only one level, whereas Region divides the blocks of the catchment into North and South and further into sub regions NE (North-East), NW (North-West), SE (South-East) and
SW (South-West). That is, Region has two levels. The data cube is shown in Figure 2.

Each node of the data cube is in fact a table. The notation $A \rightarrow B$ indicates that table B can be derived from table A by summarization. This is the Rollup feature of Online Analytical Processing. An example of a table is given in Figure 3.

A vector notation is a way to refer to the entities represented by each row of the tables in the data cube. A table is referred to as a *cuboid*. Rows are refereed to as *cells*. A cell can be written
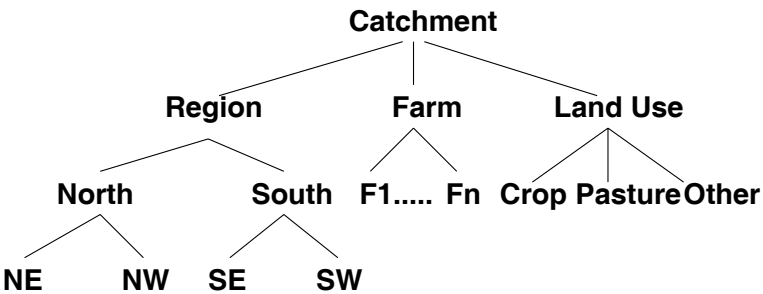
$$(D_1, D_2, \ldots, D_n)$$
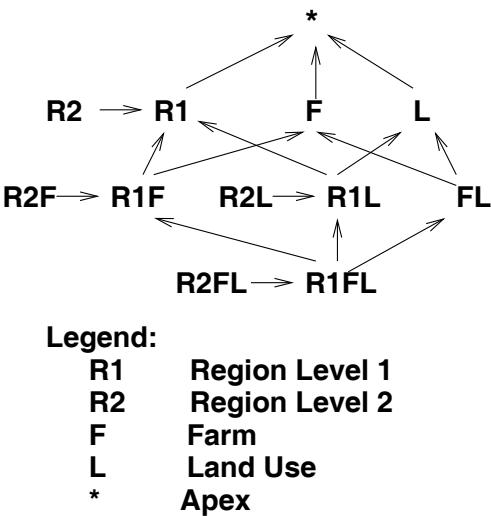
Figure 1: Hierarchical data dimensions.

Figure 2: Data cube.

where the elements are divisions of the *n* data dimensions. For example,

$$(North, *, Crop)$$

is a cell of the cuboid $R1L$ shown in Figure 3. The $*$ symbol means that this dimension (Farm) is not detailed. In general, a cell from an *m* dimensional cuboid has *m* non-$*$ elements. The most detailed cuboid is known as the *base cuboid*. In the example of Figure 2 it is $R2FL$. A cell of the base cuboid has no $*$ elements. An example is

$$(NW, F1, Crop)$$

The derivation of cells between levels in the data cube can be expressed $v_1, \ldots, v_w \rightarrow v_x$, meaning that aggregations at cell v_x can be derived from those at cells $v_1, \ldots, v_w$.

R1L : Region Level 1, Land Use

Region	Land Use	Aggregation Measures....
North	Crop	
	Pasture	
	Other	
South	Crop	
	Pasture	
	Other	

Figure 3: Aggregation table.

Call $v1, \ldots, v_w$ antecedent and v_x descendent cells. For example:

$$(NW, F1, Crop), (NE, F1, Crop) \rightarrow (N, F1, Crop)$$

4 Sustainability

When the number of dimensions and levels are high the data cube of sustainability indicators is a large space in which each cell is be labeled *sustainable/ unsustainable*, depending on some user-defined function f. It can be used to find answers to the questions posed in Section 1. It might also be used to find rules and generalizations by searching for patterns using techniques like Discovery-driven exploration (Sarawagi *et al.* [11]). A further useful feature is an anomaly indicator at each cell to flag that an antecedent cell is *unsustainable/sustainable*.

In some case it may be sufficient to generate only the sub cube of the whole data cube, where f indicates that the cell is sustainable. For such sub cubes, there is special case when f has the property

$$\text{if } f(v_x) \text{ is sustainable then so is } f(v_i), \ \forall v_i \rightarrow v_x \tag{1}$$

An example of such a function is:
   $f(v)$ *is sustainable if the minimum depth of the water table*
   for all blocks of v are greater than some threshold.

Then the sub cube can be generated efficiently from top (the Apex) down. At each cuboid in the data cube if a cell v_x indicates sustainability then each cell v_i for which $v_i -> v_x$ is also sustainable.

However, many sustainability functions will not have the property (1). A counter example is:
   $f(S)$ *is sustainable if the average annual profit*
   of the blocks of S exceed some threshold.

5 Conclusions

A data cube of sustainability indicators for a catchment can be generated for a variety of sustainability functions reflecting influences from individual or combined environmental, social, ecological or economic points of view. Predictive LUCC models can be used to generate the raw data set. The data cube structure permits selections of areas of interest according to the hierarchies of the data dimensions describing the catchment. The set of cells of the data cube that are deemed sustainable form a space from which it may be possible to search for rules and generalizations describing sustainability in the catchment.

References

[1] Park, G. and Alexander, J., Integrate or perish - lessons in integrated NRM from North Central Victoria, *Australasian Journal of Environmental Management*, **12**, pp. 47–56, 2005.

[2] Beverly, C., Bari, M., Christy, B., Hocking, M., and Smettem, K., Predicting salinity impacts from land use change: comparison between rapid assessment approaches and a detailed modeling framework, *Australian Journal of Experimental Agriculture*, **45**(11), pp. 1453–1469, 2005.

[3] Gray, J., Chaudrhuri, A., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F. and Pirahesh, H., Data cube: a relational aggregation operator generalizing group-by, cross-tab and sub-totals, *Data Mining and Knowledge Discovery*, **1**, pp. 29–54, 1997.

[4] Griener, R. and Cacho, O., On the efficient use of a catchment's land and water resources: dryland salinization in Australia, *Ecological Economics*, **38**, pp. 441–458, 2001.

[5] Han, J. and Kamber, M., *Data mining: concepts and techniques*, Morgan Kaufmann, 2nd edition, page 139, 2006.

[6] Brown, V. A., Knowing: linking the knowledge cultures of sustainability and health, in Brown, V.A., Grootjans, J., Ritchie, J., Townsend, M. and Verrinder, G. (eds.), *Sustainability and Health: Working Together*, Allen and Unwin, Sydney, 2004.

[7] Choi, W. and Deal, B. M., Assessing hydrological impact of potential land use change through hydrological and land use change modeling for the Kishwaukee River basin, *Journal of Environmental Management*, **88**, pp. 1119–1130, 2008.

[8] Li, X., Peterson, J., Liu, G., and Qian, L., Assessing regional sustainability: the case of land use and land cover change in the middle Yiluo catchment of the Yellow River basin, China, *Applied Geography*, **21**, pp. 87–106, 2001.

[9] Liu, J., Liu, S., Tieszen, L. and Chen, M., Estimating soil erosion using the USPED model and consecutive remotely sensed land cover observations, *Proceedings of the 2007 summer computer simulation conference*, San Diego, 2007.

[10] McColl, C. and Aggett, G., Land-use forecasting and hydrologic model integration for improved land-use decision support, *Journal of Environmental Management*, **84**, pp. 494–512, 2007.

[11] Sarawagi, S., Agrawal, R., and Megiddo, N., Discovery-driven exploration of OLAP data cubes, *Proceedings of the International Conference of Extended Database Technology*, Valencia, March, 1998.

[12] Scott, J., How can farmers and researchers measure sustainability?, *Proceedings of the 11th Australian Agronomy Conference*, Geolong, February, 2003.