# Application of Growing Self-Organising Maps to the Classification of Catchment *Form* and *Forcing*

**Michael Allchin[1] and Axel Anderson[2,3]**

1 University of Northern British Columbia; allchin@unbc.ca
2 Foothills Research Institute, Hinton, AB; aanderson@footshillsri.ca
3 Alberta Environment and Sustainable Resource Development, Edmonton, AB

## Abstract

Hydrologists are often required to transfer information from gauged to ungauged catchments: in order to do so, they must be confident that the landscape attributes and climatic conditions of the target and its analogue are sufficiently similar that the flow regimes to which they give rise will also be comparable. This is challenging, because of the wide variety of influences involved in determining hydrological response, and the degree of heterogeneity among and within catchments. Pattern recognition – or classification – can help with this. This paper describes the application of the Growing Self-Organising Map, a data-mining algorithm which classifies by unsupervised machine learning, to the task of identifying similarity among catchments in terms of their landscape and climate attributes. The technique identified distinct spatial associations, reasonable internal consistency among underlying metrics, and strong correspondence with an existing independent classification.

## Background and Relevance

The catchment, drainage basin or watershed is the fundamental spatial unit of hydrological research. These are conceptualised as contiguous topographic funnels which gather water from the atmosphere, and focus the portion not lost to evaporation or aquifer recharge into outflow at a single point (Wagener *et al.*, 2007). The state and transmission of water is controlled primarily by thermodynamic, gravitational and chemical influences within the physical settings encountered: catchments thus function as complex spatio-temporal filters, transforming climate inputs to outlet streamflow (Woods, 2002).

Hydrologists are often tasked with assessing likely hydrological behaviour in ungauged watersheds by transferring information from gauged catchments with similar landscape attributes and climatic influences. However, it may be challenging to identify the degree of similarity between two basins (Beven, 2000). One method for doing so is to detect distinctive patterns relating to specific associations of properties, processes and responses (Sivapalan, 2005). 'Pattern recognition' is synonymous with 'classification': there have been repeated calls for the development of objective approaches to catchment classification (McDonnell and Woods, 2004; Wagener *et al.*, 2007; Sivakumar *et al.*, 2013). Such schemata should integrate characteristics of physical landscape, climatic influences and flow regime: new approaches are thus required through which to associate these components of *Form*, *Forcing* and *Function* (Wagener *et al.*, 2007).

The diversity of influences to be acknowledged make this a daunting challenge. However, this is the type of task for which data-driven techniques have been developed: the Self-Organising Map (SOM: Kohonen, 1982, 1990), a type of Artificial Neural Network, is particularly appropriate. Originally developed for high-performance systems, SOMs have been applied only rarely for this purpose (Kelteh *et al.*, 2008), but increases in computational power now permit their operation on mainstream platforms. Moreover, past difficulties in obtaining data of appropriate quality, resolution and extent have been addressed by more widespread open publication.

This paper describes how a SOM variant was used to generate informative characterisations and classifications of hydrological catchments, based on descriptions of their intrinsic physiographies and prevalent meteorological conditions.

## Methods and Data

The investigation focused on the Province of Alberta, where the Water Program at the Foothills Research Institute seeks improvements in catchment characterisation and classification to guide the investigation of hydrological processes in different environmental contexts, support the transfer of information and knowledge between basins, and inform planning for future research activities.

Data describing physical and climatic characteristics were selected against criteria of ready and free availability, quality, consistency, credibility of provenance, and spatio-temporal extent and resolution, and summarised at kilometric resolution. Physiographic descriptors included topography, land-cover, solid and surficial geology, soil drainage and permafrost content: mean 1989-2009 climate data were generated using ClimateWNA (Wang *et al.*, 2012), which downscales the PRISM dataset (Daly *et al.*, 2002) for Western Canada. The boundaries of 213 gauged catchments across Alberta were kindly provided by Environment Canada.

Detailed description of the SOM training process is outside current scope, but is available from Kohonen (1982, 1990). On its completion, the SOM partitions the data-space into Voronoi Regions, and projects them onto its grid: spatial relationships between nodes reflect those within the data-space. One problem with a fixed array is that some idea of the dataset's internal variability is required in order to size it appropriately: this may not be available for large datasets. Dynamic SOMs therefore begin with a few nodes, and add more as additional variability is encountered in the dataset. The Growing SOM (GSOM: Alahakoon *et al.*, 2000) is one such design: this study used software understood to implement the enhanced algorithm described by Amarasiri *et al.* (2004).

The general approach of the study was as follows, repeated for data describing physiographic *Form* and climatic *Forcing*;

- Stage 1;

    - Prepare a dataset of multi-dimensional points which comprehensively represents the conditions throughout the study-area (or period);
    - Train a GSOM to identify distinct classes within this dataset;
- Stage 2;
    - Describe every gauged catchment (or gauge) in terms of these classes;
    - Train another GSOM to identify clusters among these descriptions;

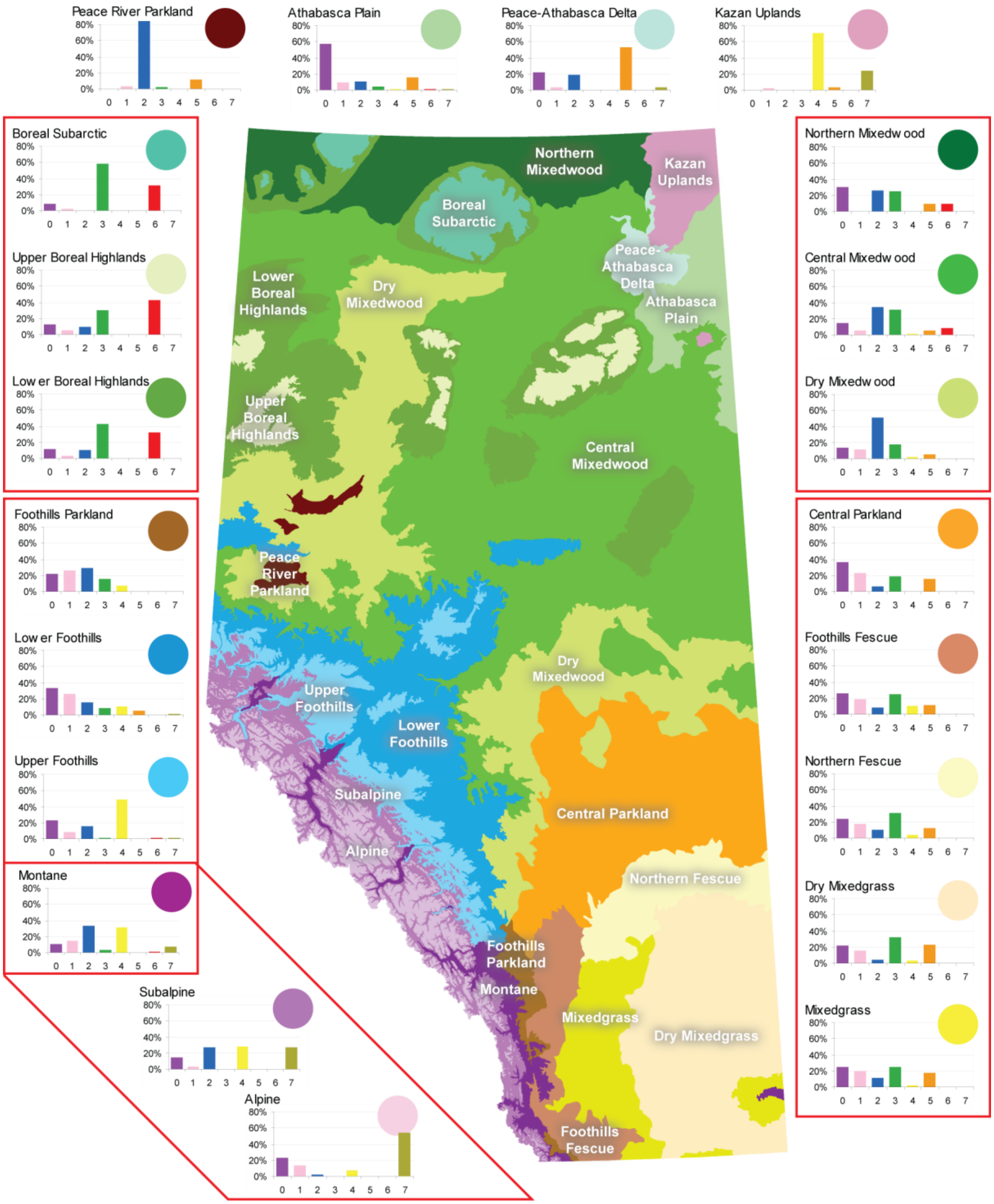
Training progressed by generating prototype GSOMs from different training durations. 'Success' at each stage was judged against reported quantisation errors, apparent organisation of nodes and clusters within the GSOM, spatial segmentation of catchments (gauges) when mapped by the cluster with which they were associated, internal consistency of the underlying metrics, and comparison with independent classifications – particularly the Alberta Natural Sub-Regions (NSRs: Government of Alberta, 2006). In this way, GSOM-based classifications were developed for the stages described above.

## Results

The best of the Stage 1 *Form* GSOM prototypes (judged on quantisation error and apparent internal organisation) identified eight landscape classes, but the disparate metrics involved made it difficult to judge how meaningful these were. However, on computing their proportional cover in the NSR polygons, it was found that these formed distinctive signatures for each (Figure 1), indicating that they provide a useful representation of landscape type. The Stage 2 *Form* GSOM was thus developed using a dataset which characterised every catchment in terms of these Stage 1 classes. The best prototype identified nine clusters, which showed reasonable spatial segmentation across the study-area, and within the NSRs: however, the distributions of some of the underlying metrics suggested that the classification was not fully mature.

It is acknowledged that using a climate dataset generated by spatial interpolation to seek Stage 1 *Forcing* clusters is rather self-confirming, but unfortunately this was the only option available. The resultant GSOM identified fifteen clusters, which were – as anticipated – closely associated with latitude and elevation.

Given the relatively low frequency of spatial variation among these climate classes, the Stage 2 *Forcing* process was adapted: instead of simply classifying the catchments in terms of their cover by these Stage 1 classes, a new dataset was generated which described the proportional representation of juxtapositions between the Stage 1 *Form* and Stage 1 *Forcing* classes. This was intended to reflect the importance of incidence of climatic conditions on different landscapes. The best resultant GSOM identified twelve clusters, which displayed a clear spatial segmentation, improved internal consistency in terms of the underlying metrics, and close correspondence with the NSRs (Figure 2). Intriguingly, the mean proportional cover of the Stage 1 *Form* classes among these Stage 2 *Form-Forcing* clusters was broadly comparable to those of the Stage 2 *Form* clusters, suggesting that the latter were not as unreliable as initially suspected.




**Alberta Natural Sub-Regions Data:**
Natural Regions Committee 2006. Natural Regions and Subregions of Alberta.
Compiled by D.J. Downing and W.W. Pettapiece. Government of Alberta. Pub. No. T/852.
Available from
http://tpr.alberta.ca/parks/heritageinfocentre/naturalregions/


Figure 1: Stage 1 *Form* cluster representation in the Alberta Natural Sub-Regions

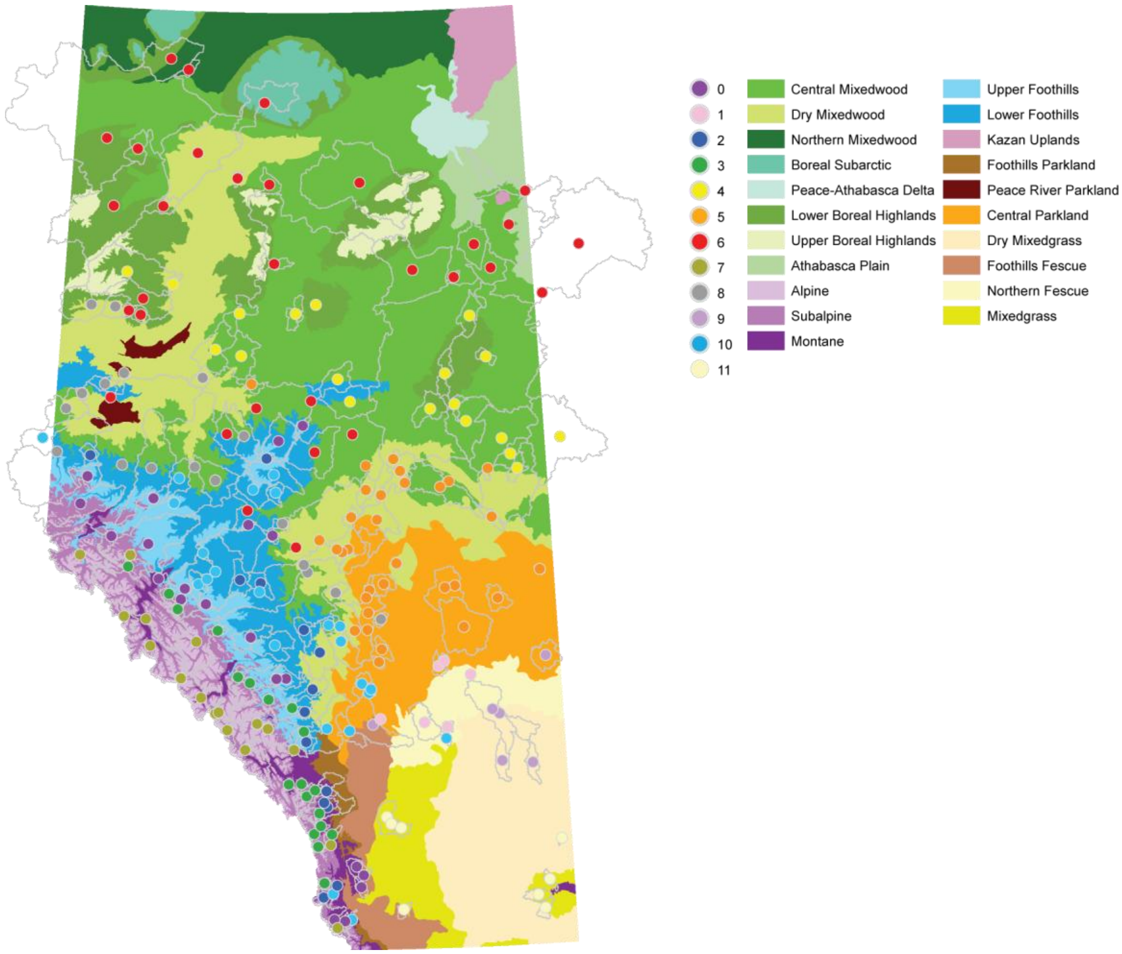



Figure 2: Distribution of basins in Stage 2 *Form-Forcing* clusters, contextualised by NSRs

## Conclusions

This study thus demonstrated that the GSOM algorithm, implemented on non-specialist platforms using readily- and freely-available datasets, was able to identify clusters of catchments and their associated gauges on the basis of their *Form* and *Forcing*, reflecting linkages between hydrometeorological inputs and landscape attributes. Subsequent stages in this project developed a further grouping of catchments based on hydrometrically-manifested *Function*, and then a unifying classification spanning all three characterisation components. Whilst these results are far from perfect, and are likely to be improved considerably through application of less heuristic approaches to parameter optimisation, it seems clear that this technique has much to offer in this role.

## References


Alahakoon D., Halgamuge S.K. and Srinivasan B. (2000) Dynamic self-organising maps with controlled growth for Knowledge Discovery *IEEE Transactions on Neural Networks* 11(3): pp. 601-614

Amarasiri R., Alahakoon D. and Smith K.A. (2004) HDGSOM: a modified growing self-organising map for high dimensional data clustering *Proceedings of the Fourth IEEE International Conference on Hybrid Intelligent Systems, December 2004*: pp. 216-221

Beven K.J. (2000) Uniqueness of place and process representations in hydrological modelling *Hydrology and Earth System Sciences* 4(2): pp. 203-213

Daly C., Gibson W.P., Taylor G.H., Johnson G.L. and Pasteris P. (2002) A knowledge-based approach to the statistical mapping of climate *Climate Research* 22: pp. 99-113

Government of Alberta Natural Regions Committee (2006) *Natural Regions and Sub-Regions of Alberta*. Compiled by D.J. Downing and W.W. Pettapiece. Government of Alberta Pub. No. T/852. URL:
http://tpr.alberta.ca/parks/heritageinfocentre/docs/NRSRcomplete%20May_06.pdf
(Retrieved 15th June 2013)

Kalteh A.M., Hjorth P. and Berndtsson R., (2008) Review of the Self-Organising Map (SOM) Approach in Water Resources: Analysis, Modelling and Application *Environmental Modelling and Software* 23: pp. 835-845

Kohonen T., (1982) Self-Organised Formation of Topologically Correct Feature Maps *Biological Cybernetics* 43: pp. 59-69

Kohonen T. (1990) The Self-Organising Map *Proceedings of the IEEE* 78(9): pp. 1464-1480

McDonnell J.J. and Woods R. (2004) Editorial: On the need for catchment classification *Journal of Hydrology* 299: pp. 2–3

Sivakumar B.M., Singh F.V.P., Berndtsson R. and Khan S.K. (2013) Catchment Classification Framework in Hydrology: Challenges and Directions *Journal of Hydrologic Engineering* In press doi:10.1061/(ASCE)HE.1943-5584.0000837

Sivapalan M. (2005) Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale *Chapter 13 in Encyclopedia of Hydrological Sciences* (Anderson M.G. – ed.) 2008 John Wiley & Sons, Ltd

Wagener T., Sivapalan M., Troch P. and Woods R. (2007) Catchment Classification and Hydrologic Similarity *Geography Compass* 1(4): pp. 901–931

Wang T., Hamann A., Spittlehouse D. and Murdock T.N. (2012) ClimateWNA - High-Resolution Spatial Climate Data for Western North America *Journal of Applied Meteorology and Climatology* 61: pp. 16-29

Woods R. (2002) Seeing catchments with new eyes *Hydrological Processes* 16: pp. 1111–1113