Characterisation and Classification of Hydrological Catchments in Alberta, Canada Using Growing Self-Organising Maps

Michael Allchin*[1]

[1]Natural Resources and Environmental Studies,
University of Northern British Columbia, Canada

January 8, 2015

Summary

Operational hydrologists are often required to transfer information from well-understood instrumented research basins to 'wild' catchments for which few details are available. To do so successfully, the climatological inputs and physiographic processing in both must be sufficiently similar that their resultant flow regimes will also be comparable. This is challenging to determine, because of the wide variety of influences on hydrological response, and the degree of heterogeneity among and within catchments. Pattern recognition – or classification – can help with this. This study explores the application of Growing Self-Organising Maps, a data-mining technique based on unsupervised machine learning, for this purpose.

KEYWORDS: Catchment Hydrology; Classification; Data-Mining; Self-Organising Maps

1. Introduction: The Need for Classification

Most hydrological research is driven by the dependence of human and ecological systems on freshwater resources, and the potential impacts of their surfeit or deficit. The fundamental spatial unit adopted for many studies is the catchment or drainage basin, conceptualised as a topographic funnel which converts spatially-distributed precipitation into streamflow at a single outlet (Wagener *et al.*, 2007). Processes hosted by the basin thus modulate meteorological inputs, acting as complex spatio-temporal filters which control the pathways and rates of water transmission (Woods, 2002).

While landscape attributes evolve through mutual interaction, so catchments possess a degree of self-organisation (Sivapalan, 2005; Ehret *et al.*, 2008), their physiographies are also highly heterogeneous (Troch *et al.*, 2008): every catchment is essentially unique (Beven, 2000). It follows that the accuracy with which a catchment's transformation of climatic inputs to streamflow outputs – its 'hydrological response' – may be modelled, depends largely on the resolution at which these properties are represented. However, operational hydrologists require generalised, practical representations of links between climate, landscape and flow regimes for predictive purposes. One approach is *regionalisation*, which seeks to transfer information describing the behaviour of instrumented basins to ungauged 'wild' catchments. This in turn depends on the recognition of signature spatio-temporal patterns of climate and landscape, and their association with different hydrological responses (Sivapalan, 2005; Beven, 2000). Pattern recognition – *the association of an infinite, continuous set of inputs with a finite variety of outputs* (András, 2008) – implies the identification and labelling of components based on their distinguishing characteristics, or 'classification'.

There have been repeated calls for the development of objective and rigorous methods for catchment classification (Wagener *et al.*, 2007; McDonnell and Woods, 2004; Sivakumar *et al.*, 2013). Such a framework should integrate physiographic *Form* and climatological *Forcing* with hydrological

* allchin@unbc.ca

response or *Function*, as manifested in synoptic hydrometric records (Wagener *et al.*, 2007). Whilst the diversity of influencing factors makes this a daunting challenge, this is the type of task for which data-driven techniques such as the Self-Organising Map (SOM: Kohonen, 1982, 1990) have been developed. SOMs have so far been applied only rarely for this purpose (Kalteh *et al.*, 2008), but increases in computational power now permit their operation on mainstream platforms, and more suitable datasets have been made available from credible sources.

This study applied a SOM variant to generate classifications of *Form*, *Forcing* and *Function* for approximately 200 catchments across the Province of Alberta, Canada (Figure 1), supporting the identification of associative patterns between climatological inputs, physiographic processing, and hydrometric outputs.

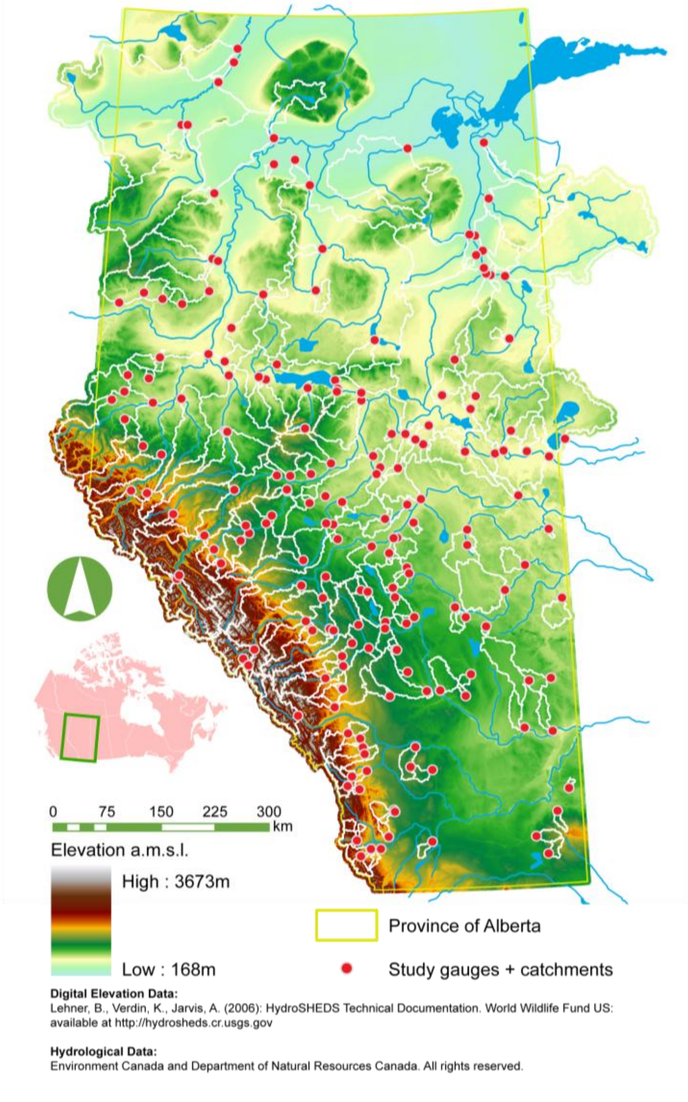


Figure 1: The Province of Alberta, showing catchments and gauges included in this study

2. Data and Methods

The conventional SOM comprises a fixed grid of neurons, each owning an ordered set of numerical weights. Training is achieved through unsupervised machine learning (Kohonen, 1982, 1990): on its completion, a SOM partitions a dataset into Voronoi Regions, projecting these onto its grid so that spatial relationships between neurons reflect those within the data-space. Individual neurons thus represent a fine-resolution classification, but may also form distinct contiguous clusters. One problem with a fixed SOM is that some idea of the dataset's internal variability is required to size the grid appropriately, but this may not be available. Dynamic SOMs, such as the Growing SOM algorithm adopted for this study (GSOM: Alahakoon *et al.*, 2000; Amarasiri *et al.*, 2004), therefore begin with a few neurons, and add more as additional variability is encountered in the dataset.

Physiographic and climatological datasets (Tables 1, 2) were sourced for the Province of Alberta (~700,000 km²), using criteria of ready and free availability, quality, consistency, credibility of provenance, and spatio-temporal extent and resolution, and summarised at kilometric resolution. Hydrometric data measured at 213 gauges between 1989 and 2009, and corresponding catchment boundaries, were provided by the Water Survey of Canada.

Table 1: *Form* descriptors

<table>
<tr><th>Group</th><th colspan="2">Metrics</th></tr>
<tr><td>Slope</td><td colspan="2">Mean
Coefficient of variation</td></tr>
<tr><td>Slope / Aspect
(%age cover)</td><td colspan="2">NW-NE / NE-SE / SE-SW / SW-NW + shallow (<= 10°) / moderate (10° <= 30°) / steep (30° <= 60°) / very steep (> 60°)</td></tr>
<tr><td>Surface Complexity</td><td colspan="2">Represented by coefficient of variation of
Beven and Kirkby (1979) Wetness Index</td></tr>
<tr><td rowspan="5">Solid Geology
(%age cover)</td><td>Cenozoic</td><td>coarse siliciclastic
coarse-medium siliciclastic
medium siliciclastic
medium-fine siliciclastic
volcanic</td></tr>
<tr><td>Mesozoic</td><td>coarse siliciclastic
coarse-medium siliciclastic
medium siliciclastic
medium-fine siliciclastic
fine siliciclastic
carbonates</td></tr>
<tr><td>Palaeozoic</td><td>coarse siliciclastic
coarse-medium siliciclastic
medium siliciclastic
medium-fine siliciclastic
fine siliciclastic
carbonates
evaporates</td></tr>
<tr><td>Proterozoic</td><td>coarse siliciclastic
coarse-medium siliciclastic
medium siliciclastic
fine siliciclastic
carbonates
plutonic / high-grade metamorphic</td></tr>
<tr><td>Archaean</td><td>plutonic / high-grade metamorphic</td></tr>
<tr><td>Drift Geology
(%age cover)</td><td colspan="2">Alluvial deposits
Coarse grain
Colluvial blocks
Colluvial fines
Colluvial rubble
Colluvial sand
Complex
Aeolian deposits
Fine grain</td></tr>
</table>

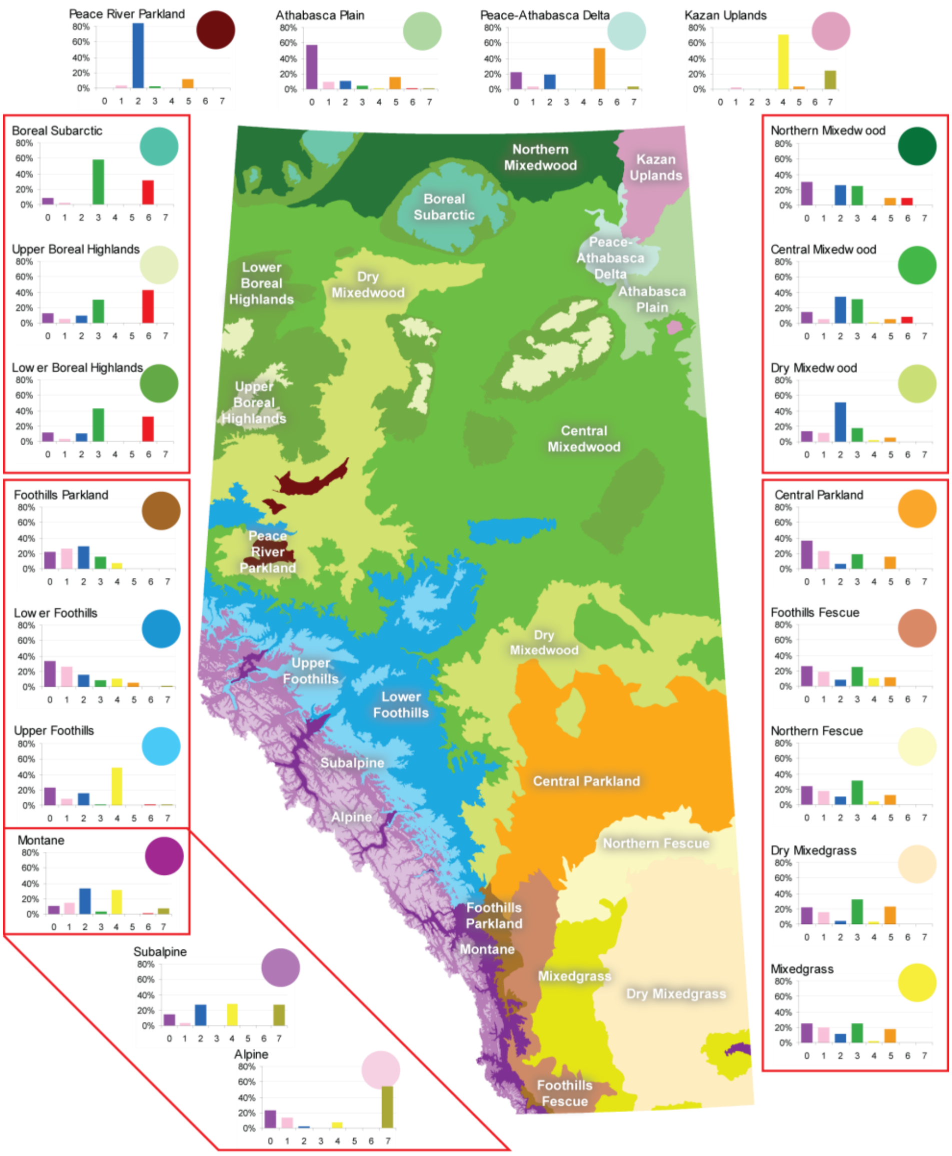
Group	Metrics
	Glaciers Organic deposits Plain sands and gravels Till blanket Till veneer Undivided (implies negligible or absent)
Land Cover (%age cover)	Temperate / sub-polar needleleaf forest Sub-polar taiga needleleaf forest Temperate / sub-polar broadleaf deciduous Mixed Forest Temperate / sub-polar grassland Sub-polar / polar grassland-lichen-moss Sub-polar / polar barren-lichen-moss Wetland Arable Barren Lands Urban / Built-Up Water
Soil Drainage (%age cover)	n/a (no soil) Very poor Poor Imperfect Moderate Good Rapid Very rapid
Permafrost (%age cover)	Isolated patches (0-10%), low (<10%) ground Ice Isolated patches (0-10%), low-nil (0-10%) ground ice Sporadic discontinuous (<10%), low (<10%) ground ice
Total 81 values per data-point	

Table 2: *Forcing* descriptors

Description
Mean daily maximum temperature (°C)
Mean daily mean temperature (°C)
Total monthly precipitation (mm)
Degree-days below 0°C
Degree-days above 5°C
Degree-days below 18°C
Degree-days above 18°C
Number of frost-free days
Precipitation as snow (mm)
Hargreaves reference evaporation (mm)
12 monthly values per metric: ***120 values per data-point***

GSOMs were developed to classify representative samples of these descriptions. The physiographic training data comprised 50% of the cells in a chequerboard pattern: given the lower spatial frequency of variation in the climatological dataset, 25% of these were used in this dataset. Prototype GSOMs were generated from increasing training durations, with outcomes judged by the contiguity of neuron-clusters identified by the GSOM software; metrics of central tendency and dispersion of neuron quantisation errors within clusters; their spatial segmentation when mapped; internal consistency of the underlying descriptors; and comparison with an independent classification, the Alberta Natural Sub-Regions (NSRs: Government of Alberta, 2006).

The second stage identified from these GSOMs the *Form* and *Forcing* class of every 1 km² cell in each catchment, and characterised every basin in terms of the fractional cover of each combination. These profiles were used to develop a further GSOM, to identify clusters of catchments with comparable climatological inputs and physiographic processing.



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 2: Distributions of fractional cover of the eight *Form* classes in the NSR polygons

The approach developed to classify the catchments' gauged hydrometric profiles lies outside the core scope of GIS Research: however, a brief description is required. If two catchments exhibit comparable hydrological response (driven primarily by their physical attributes), then with similar climatic inputs in a given year, their annual hydrographs should develop similar shapes. A GSOM was first developed to classify the set of annual hydrographs measured at the available gauges from 1989 to 2009 into clusters with broadly comparable timings and intensities of peak flow. A second GSOM was then developed to cluster gauges based on consistent inter-annual similarities in annual hydrograph shape.

Having thus identified the membership of each catchment in classes of physiographic / climatological *Form-Forcing* and hydrometric *Function*, overlaps could then be explored between these associations, thereby supporting inferences about the probable hydrological behaviour of catchments based purely on readily-available spatial data.

3. Results

The *Form* GSOM identified eight physiographic classes across the study area. The disparate metrics involved initially made it challenging to determine whether or not these were meaningful. However, computing their fractional coverages in each of the twenty-one NSR class polygons revealed clear signature distributions, and evident associations with the established NSR categories (*e.g.* Prairie, Mountain / Foothill, Boreal Forest) (Figure 2), indicating that these classes provided useful representations of landscape type.

The *Forcing* GSOM identified fifteen climate clusters, which were closely associated with latitude and elevation. It is acknowledged that this was somewhat self-confirming, given that the training dataset had itself originally been generated using the ClimateWNA software (Wang *et al.*, 2012), which downscales the PRISM regional re-analysis for Western Canada (Daly *et al.*, 2002) by spatial interpolation. However, this was the only practical option available.

The percentage spatial cover of each of the 120 theoretically possible combinations of juxtaposed *Form* and *Forcing* classes was computed for every catchment, and this dataset was used to train a further GSOM. Twelve clusters were identified, which were evidently associated with distinct spatial domains across the study area (Figure 3). To confirm that these provided a meaningful representation of physiographic and climatological characteristics, the fractional cover of the twenty-one Alberta NSRs was computed for every catchment, and plotted for the basins in each identified cluster. The rationale here was that the NSRs had been derived largely through qualitative expert judgement (Government of Alberta, 2006) to combine aspects of physiography, climate and ecology, and therefore provided a 'benchmark' of landscape against which to assess this unsupervised, and arguably more objective, classification. The resultant distributions showed clear associations between the two schemes.

The *Function* classification yielded seven clusters of gauges exhibiting broadly consistent similarities in their annual hydrograph shapes from year to year. The long-term mean hydrographs of the members of each cluster also showed distinct characteristics, which may be related to the principal causative influences on their flow distributions, such as snow-melt, rainfall or glacial melt (Pardé, 1933; Lvovich, 1938).

When the membership of the catchments within the *Form-Forcing* and *Function* classes were compared, very clear associations were immediately evident between the two (Figure 3). The similarity of flow-regime distributions in the Boreal Forest, Parkland / Mixedwood and Prairie landscape groups, and the gradual transition in the fractional representation of each *Function* class with diminishing elevation from the high alpine to the lower foothills, are particularly interesting. Note also that although *Function* classes 3 and 5 have quite similar shapes, with a peak relatively late in the spring and slow decline through the summer, they occur in distinct settings: analysis of the *Form* and *Forcing* attributes of the catchments with which they are each associated strongly imply that the former results from higher summer precipitation in the northern foothills, while the latter is driven by wetland fill-and-spill in the Boreal Forest. The broad summer peaks in the alpine basins of the Rocky Mountains and their higher foothills are inferred to result from a combination of an extended late snow-melt season at these elevations, large amounts of summer precipitation, lacustrine spill-and-fill, and glacial melt.

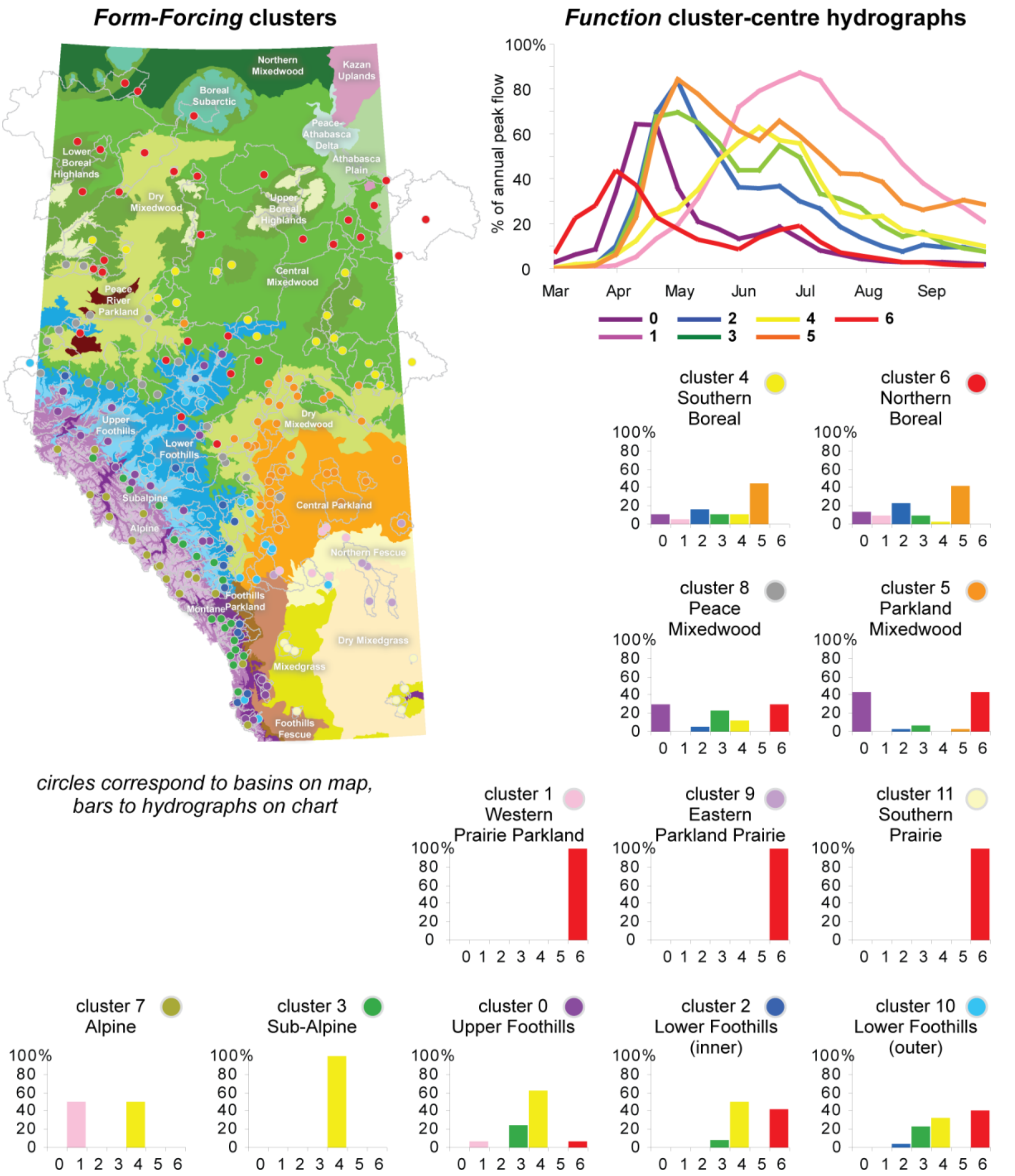


Figure 3: Spatial distribution of the twelve *Form-Forcing* classes (contextualised by NSRs), and fractional representation of the seven *Function* classes within each

4. Conclusion

This paper describes the potential of SOMs to identify relationships between climatological inputs, physiographic processing, and streamflow outputs, using readily-available data from disparate sources. Importantly, it distinguished between catchment clusters possessing similar hydrometric profiles but contrasting physiographic / climatological attributes. While this amounts so far only to the identification of broad associations between relatively coarse groups of catchments and flow distributions, subsequent refinement is expected to deliver more informative links between catchment descriptions and specific regimes in this geographic context. It has also provided independent support for the association of catchment descriptions based on ecological regions with flow regimes, as suggested by research conducted in the neighbouring Canadian province of British Columbia (Trubilowicz *et al.*, 2011). By extending similar analyses further afield and continuing to develop these techniques, it may be possible to make progress towards rigorous and objective classification schemes which are valid at continental or even global scales.

5. Acknowledgements

This paper describes part of a study submitted as a dissertation for the degree of MSc in GIS, supervised jointly by Dr A. Heppenstall (University of Leeds) and Dr J. Leyland (University of Southampton). The study was generously supported by funding made available by Dr A. Anderson of the Foothills Research Insitute, Hinton, Alberta.

6. Biography

- BSc, Geology, Bristol (1986)
- Developer of 'hydrogeoinformatics' software, primarily for the NERC Centre for Ecology and Hydrology / Wallingford HydroSolutions (1993 – 2013)
- MSc, GIS (Online Distance Learning), Leeds / Southampton (2013)
- Now a (rather elderly) PhD Candidate: University of Northern British Columbia, Prince George, BC, Canada

References

Alahakoon D., Halgamuge S.K. and Srinivasan B. (2000) Dynamic self-organizing 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-organizing map for high dimensional data clustering *Proceedings of the Fourth IEEE International Conference on Hybrid Intelligent Systems, December 2004*: pp. 216-221

András K. (2008) Linguistic pattern recognition *Mathematical Linguistics* Chapter 8: pp 201-217 Springer London

Beven K.J. and Kirkby M.J. (1979) A physically based, variable contributing area model of basin hydrology *Hydrological Sciences* 24,1,3

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

Ehret U. and 20 others (2014) Advancing catchment hydrology to deal with predictions under change *Hydrology and Earth System Sciences* 18: pp. 649-671

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://www.albertaparks.ca/media/2942026/nrsrcomplete_may_06.pdf (Retrieved 8th January 2015)

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

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

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

Lvovich M.I. (1938) Opyt Klassifikatsii Rek SSSR *Trudy GGI*, Vyp. 6, Leningrad

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

Pardé M. (1933) (sometimes quoted as a later edition, 1955: 5th edition 1968)
Fleuves et Rivières Paris: Armand Colin

Sivakumar B.M., Singh V.P., Berndtsson R. and Khan S.K. (2015) Catchment Classification Framework in Hydrology: Challenges and Directions *Journal of Hydrologic Engineering* 20(1): SPECIAL ISSUE: Grand Challenges in Hydrology, A4014002

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

Troch P.A., Carrillo G.A., Heidbüchel I., Rajagopal S., Switanek M., Volkmann T.H.M. and Yaeger M. (2008) Dealing with Landscape Heterogeneity in Watershed Hydrology: A Review of Recent Progress toward New Hydrological Theory *Geography Compass* 2 (2008)

Trubilowicz J.W., Moore R.D. and Buttle J.M. (2013) Prediction of stream-flow regime using ecological classification zones *Hydrological Processes* 27(13): pp. 1935-1944

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: Review of Spatial Patterns in Catchment Hydrology: Observations and Modelling (Grayson R. and Bloeschl G. – eds) *Hydrological Processes* 16: pp. 1111–1113