A strategy for the identification of areas of consistent hydrologic character by means of dimensionless numbers

M.L.V. Martina(1), J.S. Selker(2), D.E. Rupp(2), M.P. Wright(2), R. Haggerty(2), A. Nolin(2), J.J. McDonnell(2), G. Grant (3)

(1) Department of Earth and Geo-Environmental Sciences, University of Bologna, Italy, (2)Oregon State University, Corvallis, Oregon U.S.A., (3) USDA Forest Service, Pacific Northwest Research Station, Corvallis, Oregon, U.S.A - Corresponding author: mario.martina@unibo.it

Abstract

The prediction of the hydrologic response of a landscape is of critical importance to water resource and hazard planning. For these purposes it is not practical, nor we believe necessary, to gauge each stream in a basin. Rather, with the extensive knowledge of climate, soils, geology, topography, and ready access to long-term gauges in hydrologically similar conditions, it should be practical to develop accurate predictions of hydrologic response. The goal of this work is to lay out an approach to achieve these ends illustrated using data from the Willamette Valley in Oregon.

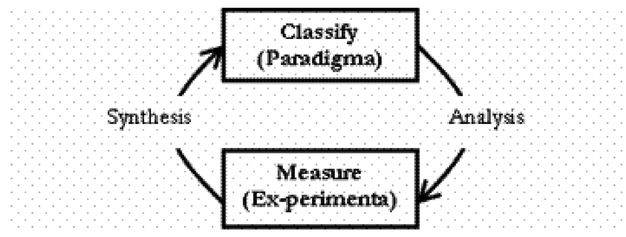
A number of dimensionless indexes (here called predictands), obtained from the combination of geomorphological and climatic characteristics of the basin, has been used to infer the "signature" of 152 subcatchments i.e. some features of the hydrograph (called predictors) such as the annual average runoff, some percetiles of the duration curve, the parameters of the hydrograph recession curve. On this purpose a linear and log-linear regression analysis has been performed. Some predictand/predictor plots show interesting threshold relations. For each plot the threshold behaviour is identified, and apparent threshold values are computed. The identification of these thresholds is approached graphically here, though ultimately this process should be made completely objective through statistical measures of the degree of parameter control. Being the overall objective of the work to identify subcatchments of consistent hydrologic character, a classification three-shaped scheme has been filled on the basis of the thresholds defined on the predictors. The leaves of the three result to be group of hydrologically similar catchments.

In conclusion, rather than attempting to present a final taxonomic system, we suggest a strategy and the roadmap for how this approach could be alternatively developed.

Why?

Why do we need a classification?
Because it is our instinctive way of understanding, learning, being conscious of our lacks of knowledge. But also in natural sciences it is the first step of the process of analyzing nature. It is a way to organize our existing knowledge.
There could be different reasons for a classification:
- Finding similarities = studying properties (inference)
- A classification helps at reducing the variability / the complexity by assuming Heterogeneity between the classes / Homogeneity within a class.
- Some of the theories work better in a subset of the cases, because our theories (models / assumptions) are still immature and not-universal. E.g. the Darcy flow, the Hortonian runoff, Saturation excess, Kinematic wave

If we look at the understanding process as a cycle: we built our hypothesis / paradigm /classification scheme, we make our experiments / measure - going through an analysis process - to verify/falsify our paradigms, we then re-built new paradigms / classification schemes by a synthesis process.



In other scientific disciplines ...

Chemistry: periodic table
[Same number of protons = similar properties]

Biology: the animal kingdom
[Same species = same features]

Geology: the geological eras / the morphology of the rocks (sedimentary, metamorphic, volcanic...)

Astronomy: Galaxies / solar systems / planets [Same galaxy = same

Fluid dynamics: fluid regimes [Same regime = same behavior]

The Need for a Quantitative System of Hydrologic Taxonomy

The prediction of the hydrologic response of a landscape is of critical importance to water resource and hazard planning. For these purposes it is not practical, nor we believe necessary, to gauge each stream in a basin. Rather, with the extensive knowledge of climate, soils, geology, topography, and ready access to long-term gauges in hydrologically similar conditions, it should be practical to develop accurate predictions of hydrologic response. It is the goal of this paper to lay out an approach to achieve these ends illustrated using data from the Willamette Valley in Oregon. Rather than attempting to present a final taxonomic system, we suggest a strategy and the roadmap for how this approach could be developed.

Taxonomy has provided an organizational framework for the organization of the fields of biology, chemistry, geology, soils, to name but a few. The goals of taxonomy are diverse, but may be summarized as a qualitative systemization of elements with similar features and behaviors. The first requirement when setting out to predict the hydrologic response of a basin is the identification of the dominant hydrologic processes: is the basin outflow dominated by deep aquifers, shallow interflow, or direct surface runoff? Is the area dominated by rainfall or is there significant snow accumulation? Is the area arid or humid? Is it steeply sloping or flood-plain-flat? The answers to these questions guide expectations of the hydrologic behavior, and even more so, the selection of how to quantitatively model the basin. Though these questions are stated in qualitative terms here, fundamentally these must be addressed through quantitative measures of the relative magnitudes of the constituent processes. Eventually there may be synthesized hydrologic models which have the capability to simulate all of these conditions, but even then there will need to be a procedure to identify which of the model features is selected for inclusion.

To date there has been a dearth of attention to quantitative taxonomy in hydrology. Most authors and scientists select modeling strategies and components of simulation based on intuitive basis, or without any explicit explanation at all. This leads to a fundamental non-reproducibility of results since the most fundamental aspects of the modeling strategy are hidden. Since independent reproducibility is a fundamental requirement of scientific validity, this represents a critical obstacle to the advancement of hydrologic science.

The potential use of the dimensionless number as predictors is to compare them to threshold values which differentiate among different hydrological classes/properties/regimes,... However the selected dimensionless number have not associated an a priori threshold value, i.e. the value which potentially could differentiate among different hydro regimes is not implicitly defined in the parameter expression. (not like with the Froude number, ...)

There are two possible ways to define those thresholds values: (1) A priori, based on experts' knowledge, (2) A posteriori, based on a sensitivity analysis (when the data are available)

In the present application, given the availability of stream data and being the first attempt to use some of the dimensionless numbers (we don't know their potential diagnostic power), we followed the latter way.

How?

Taxonomy: The Asymmetric Hierarchical Tree

There are many approaches which may be used to identify sets of similarly responding watersheds. As discussed above, in the case of n binary (e.g., control/non-control) predictors, then it would be possible to divide into 2n groupings, which quickly becomes so large to be impractical for either construction or application of the taxonomy. Thus we seek to identify *hierarchically* dominant factors which eliminate any further need to discriminate within a particular class of watersheds. An extreme example is that of the Atacama desert, where there are points that have experienced no precipitation in over 100 years (citation). Here the aridity is completely dominant, and we can ignore questions of snow, potential ET, etc. This leads to *asymmetry* in the taxonomy, in that the two branches following one division between classes will be made up of differing numbers of subsequent classes.

Although the dominance of particular factors will certainly vary widely so that no single hierarchy will be universal, we posit that a hierarchy is required, and propose a general hierarchy of Climate; Human intervention; Geology; Topography; and Vegetation. Climate incorporates temperature, precipitation amount and timing, and potential evaporation. Climate is considered to be dominant in that it controls the occurrence of water (P) and the potential for its direct dissipation (ET). Human intervention classifies the nature of direct (e.g. dams) and indirect (impervious surfaces) factors that would effect hydrologic response. This is taken as the second hierarchical factor in so far as dams in particular can alter behavior to the point of completely eliminating flow without respect to any other factor. In our analysis we specifically selected for channels not dominated by human intervention, and so do not explore this predictor, but dammed rivers might be taken as the most extreme example of asymmetry in that if a river is dammed all further taxonomic description is obviated.

Geology, which here is taken to include both soil and sub-soil properties, is taken as the third level control, in that these factors dictate the amount of water that can be stored between events for potential ET, and within a particular climatic regime largely control the division between surface and sub-surface routing. Topography, indicating slope and altitude, is influential in establishing the partitioning between surface and subsurface flow, as well as the time between impulse and response. Finally vegetation classifies the extent and type of plant coverage which determines canopy interception, root depth, and seasonality in transpiration. Though clearly very significant, vegetation occupies the lowest hierarchical position in that it is itself largely dependent on climate, soils, and topography, and so could not be logically placed above these factors.

Identification of classes of fluid behavior based on dimensionless parameters (e.g., Reynolds and Froude numbers) is a cornerstone of fluid mechanics. Many of the delineations made in classical fluid mechanics are relatively abrupt or dramatic in comparison to those required in hydrology. In place of determining if the flow will be turbulent or laminar, hydrologists must determine of the snow is plays a decisive role in stream flow. In the former case, the criteria will be universal, while in the later case the fact that snow accumulation is a large part of the precipitation may be irrelevant in later summer in Oregon, or in Greenland regardless of season. With this caution in mind, it is yet appealing and intuitively reasonable to seek objective criteria for when a climate acts arid or humid; when snow is significant or irrelevant; when soil water storage and permeability influence the hydrograph.

The list of potentially useful dimensionless numbers, while not rigorously limitless, would have many hundreds of entries based on the list of parameters which have potential importance in an environment. The selection of which to consider should include considerations of the strength of the underlying physical principles, the availability of the data, and the clarity of the correlation to predictands. In assembling our list of dimensionless parameters we purposefully avoided many dimensionless groups presented in hydrologic papers that were used to scale governing equations. Generally these dimensionless groups describe the sensitivity to particular parameters *within* a particular hydrologic framework rather than being representative of transitions *between* differing governing relationships. For instance dimensionless time for infiltration allows infiltration of differing duration to be plotted on common axes, but in itself does not assist in the identification of the relative importance of infiltration versus runoff processes. Clearly this is not a matter of black and white, as some dimensionless parameters used to make equations scale independent are useful to identify process transitions, however it is the latter characteristic which would suggest they should be included here.

Aims

The purpose of the proposed classification is to predict some features of the hydrograph at the outlet of a catchment based on data and information available even on ungauged (or poorly gauged) basins. The classification strongly depends on what one would consider *predictands* (= things to be predict) and *predictors* (= things we use to predict). Regardless of the defined predictands/predictors, requirements of the classification are to be scientifically defensible (by means of a data validation when available), globally extensible and flexible.

The definition of the *predictands/predictors* should reflect the knowledge affordable from the available data. The classification could be an efficient tool to organize the existing data/knowledge/information in order to be intelligible by hydrologist and other scientist and allow them to predict (before a more accurate analysis or before gauging the catchment) important hydrograph features and characterize the relative importance of the hydrological processes expressed by that particular catchment.

Predictors		Predictands
For the Willamette basin several maps are available both for a geomorphological and climatic characterization: the DEM, the soil types (with an indication of the soil depth, hydraulic conductivity, field capacity,..), the monthly precipitation map, the monthly PET (Potential EvapoTranspiration), the geological map (with an indication of the geological unit transmissivities, ...). In order to generalize the results of the classification and make it transferable to other catchments, the predictors has been defined as dimensionless numbers which should be relative scale independent and useful to identify dominant processes.		In the Willamette basin are installed more than 170 stream stations operative since 1975 although discontinuously. Based on these available data we defined the predictands as some characteristics of the hydrograph computed directly from the time series. These characteristics are intended useful to understand the processes which control the hydrograph (snow-dominated, groundwater dominated, ...). For the recession analysis, we referred to the relation: $\frac{dQ}{dt} = -a\,Q^b$ where Q is the discharge and dQ/dt is the time derivative, and the parameters a and b are computed only for the recession limb of the hydrograph.
Aridity index	$Aridity = \frac{ET_0}{P}$	Annual Average Discharge
Snow fraction index	$Snow\_fr = \frac{D_{snow}}{P}$	5th percentile of the discharge distribution
Snowmelt index	$Snow\_melt = \frac{R_{sw}}{\lambda_f\, D_{snow}\, P}$	95th percentile of the discharge distribution
Drainage Density index	$DDensity = \frac{l^2}{A}$	Annual maxima values of discharge
Ground water index	$GW = \frac{P - ET_a}{T\, i\, \sqrt{A}}$	Parameters of the hydrograph recession curve
Infiltration Excess index	$Inf = \frac{P_{max}}{K_s}$	
ETAP index	$Evap = \frac{ET_a}{P}$	
Soil capacity index	$Soil = \frac{P}{\theta_{max}\, Z}$	
Precipitation Variability Index	$Precip = \frac{\frac{\Delta P}{\Delta x}\sqrt{A}}{P}$	

We select dimensionless parameters that are useful to identify dominant processes, but are still computable with the available data. In assembling this list we purposefully avoided many dimensionless groups presented in hydrologic papers that were used to scale governing equations. Generally these dimensionless groups describe the sensitivity to particular parameters within a particular hydrologic framework, in the context of a particular dominant process. Clearly this is not a matter of black and white, as some dimensionless parameters used to make equations scale independent are useful to identify process transitions, however it is the latter characteristic which would suggest they should be included here.

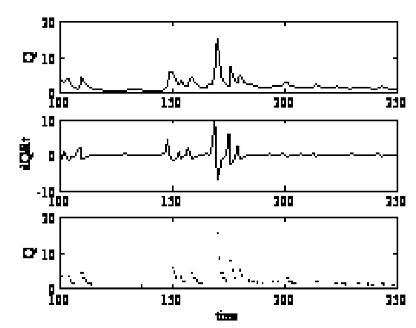


Figure 1 - Example of the recession curve selection. (1) Discharge ime series, (2) Discharge time-derivative, (3) Selection of the data in the recession phase

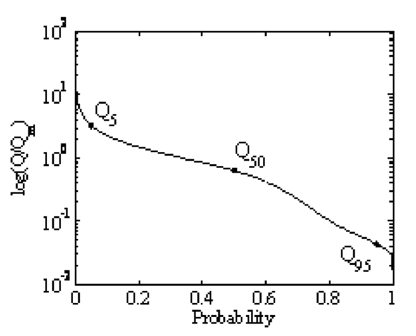


Figure 2 - Example f the duration curve for discharge normalized by the mean (Qm)

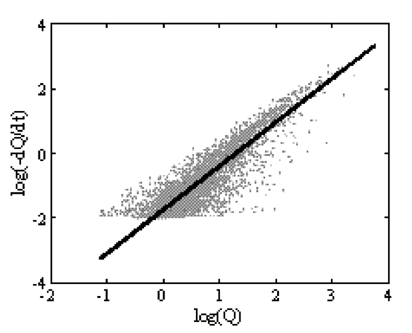
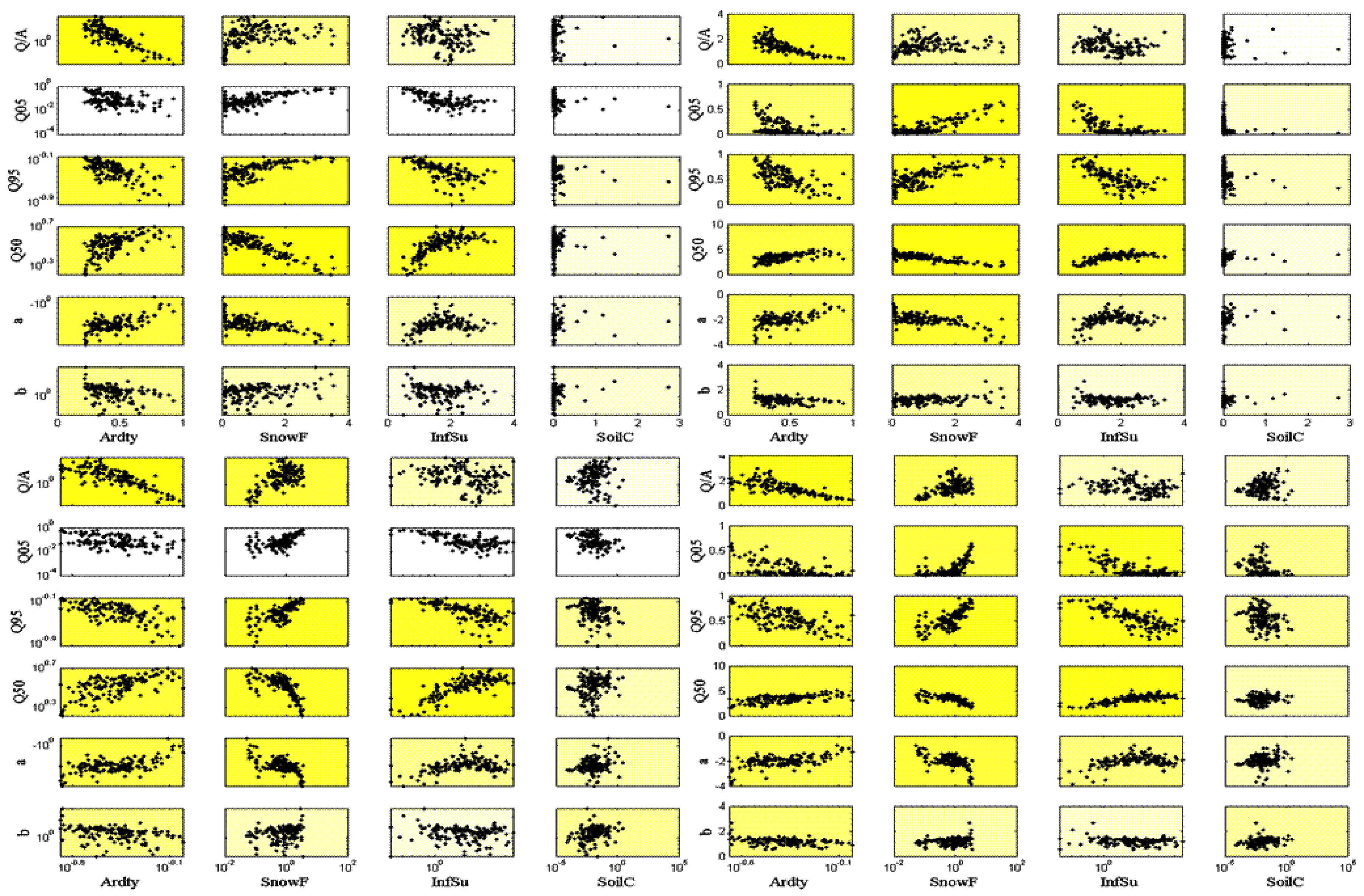
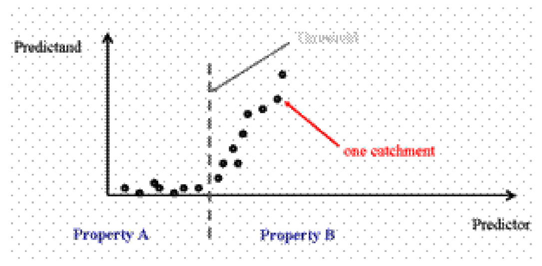
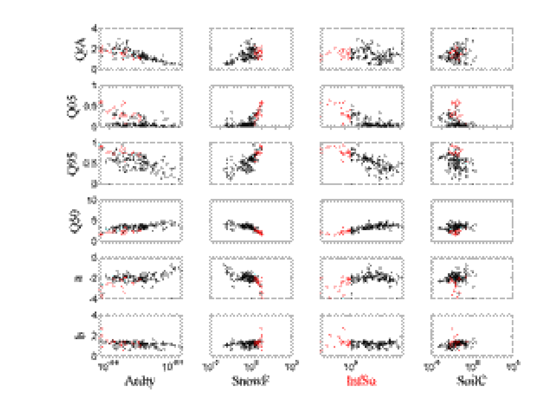


Figure 3 - Example of the regression for the analysis of the recession curve

Analysis





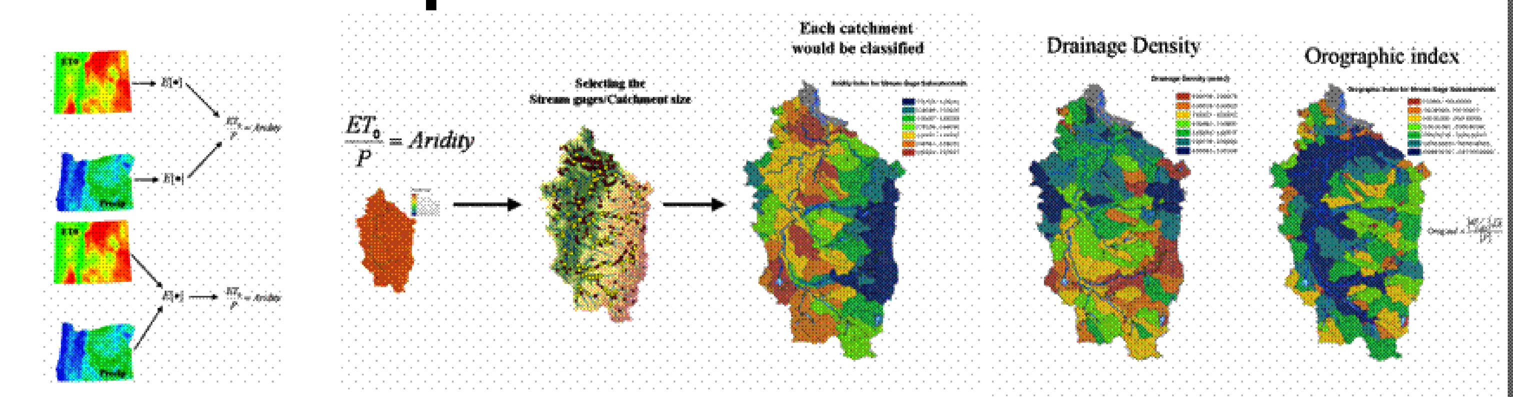


For the purposes of this paper we seek to divide basins into between five and ten classes which have similar hydrologic character. There are many strategies that could be adopted. A simple strategy would be to seek that each of the classes had approximately equal number of basins, in which case the thresholds would divide evenly based on the median value of the parameter. While appealing from a statistical perspective (common degrees of freedom), it would make the taxonomic system dependent on the base data employed (i.e., the variability within the population of basins being considered), which is not a desirable characteristic. A second approach would be to code basins which are deemed by hydrologists to be hydrologically similar, and place the threshold so that basins of similar character fall on the same side of a threshold. This approach takes advantage of expert insight, but is fundamentally subjective in nature, which violates our basic objectives. A third approach would be to plot the predictor against the predictants and look for statistically significant changes in the relationship. This could be undertaken based on statistical significance of differences between trend-lines on either side of the threshold, making it an objective approach, and makes use of actual observed P/P relationships which is always necessary when attempting to apply dimensionless variable to discriminate between phase-transitions. We therefore employ this third approach.

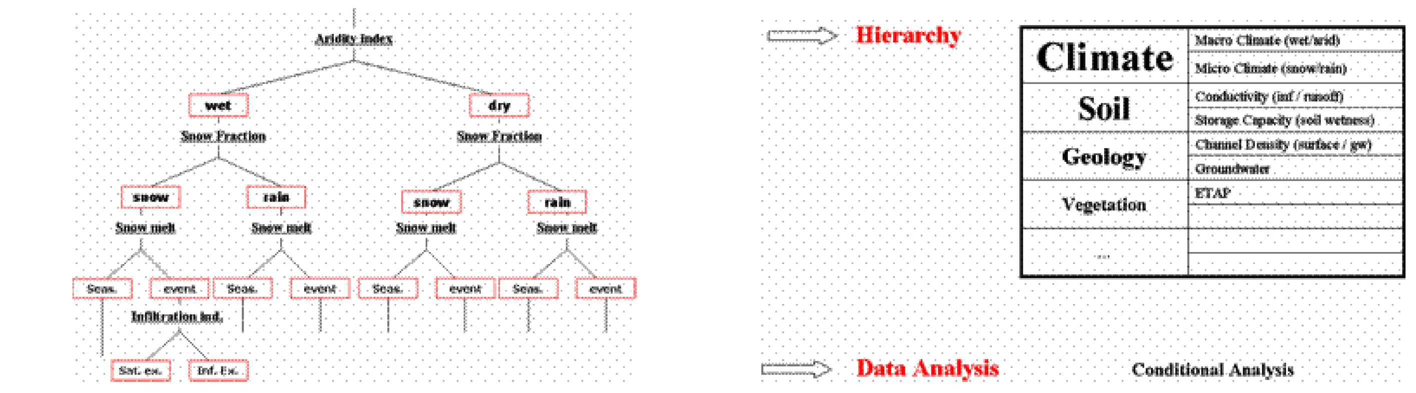
Within this approach it is apparent that for any one predictant, each P/P pair may indicate a different critical threshold. In this case we adopt a "no violations" rule, wherein a threshold can only be adopted if there is no clear violation of this value presented by one or more predictors.

Many dominant features controlling hydrologic response are correlated. Thus when seeking to identify a threshold that separates hydrologic regimes it would be natural to explore the redundancy between indices, as well as the possible non-linear combination of factors to discriminate between classes. Eventually such thresholds should make explicit use of multiple predictors simultaneously in making taxonomic groupings. For instance, when discriminating between arid and humid environments it is essential to consider the precipitation and potential ET, but it is further reasonable to include the influence of the temporal distribution of rainfall, vegetation, and soil storage capacity to identify which locations might act arid without respect to the total volume of rainfall. Since the goal of this paper is to demonstrate a conceptual approach we leave such important refinements to future work.

How to maps indexes



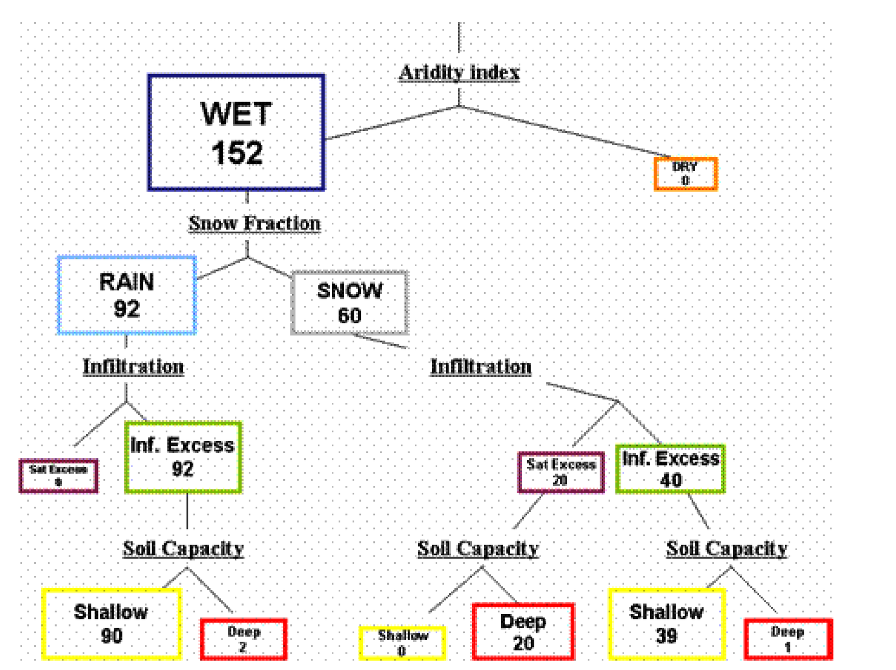
Classification Scheme



Results

The Asymmetric Hierarchical Tree

By followinfg the methodology above described, it is possible to give the shapes to the leaves and the branches of an Asymmetric Hierachical Tree which reflects the different hydrological features of the subcatchments included in the Willamette basin.



References

Grant, G. E. A geomorphic basis for interpreting the hydrologic behavior of large river basins, in River Quality, Dynamics and Restoration, edited by A. Laenan, pp. 105-116, CRC Press, Boca Raton, Fla, 1997

Grant, G. E., and C. L. Tague, The source of the river: Geologic control of the hydrologic regime, Willamette River, Oregon, Geol. Soc. Am. Abstr. Programs, 34(5), A104, 2002

Marshall, S., R. J Oglesby and A. W. Nolin, The predictability of winter snow cover over the western United States, J. Climate, 16, 1062-1073, 2003.

Martina M.L.V., Entekhabi D., Identification of runoff generation spatial distribution using conventional hydrologic gauge time series, Water Resources Research, vol.42, W08431, doi:10.1029/2005WR004783, 2006

Rupp, D.E. and J.S. Selker. Information, artifacts, and noise in dQ/dt - Q recession analysis. Adv. Wat. Resour. 29(2):154-160. 2006.

Rupp, D. E., and J. S. Selker (2006), On the use of the Boussinesq equation for interpreting recession hydrographs from sloping aquifers, Water Resour. Res., 42, W12421, doi:10.1029/2006WR005080.

Tague, C., and G. E. Grant, A geological framework for interpreting the low-flow regimes of Cascade streams, 24 Willamette River Basin, Oregon, Water Resour. Res., 40, 2004