Nonparametric direct mapping of rainfall-runoff relations*

by Ion Iorgulescu[1] & Keith Beven[2]
[1]EPFL, Lausanne, Switzerland, [2]Lancaster University, UK

(**WRR, in press*)

A new modelling and data analysis technique that differs radically from previous approaches:

1. Only (linear) combinations of inputs as predictor variables (no 'state' variables)
2. Regression trees, a non-linear non-parametric model based on recursive binary partitioning as identification algorithm.

Predictor variables

Type	Description	Mathematical formulation
I	Cumulative precipitation (P) over the previous K time steps	$\sum_{i=0}^{K} P(t-i\Delta t) \quad K=0,1,\ldots m$
II	Cumulative potential evapotranspiratio (PET) over the previous K time steps	$\sum_{i=0}^{K} PET(t-i\Delta t) \quad K=0,1,\ldots m$
III	Cumulative P-PET over the previous K time steps	$\sum_{i=0}^{K} (P(t-i\Delta t) - PET(t-i\Delta t)) \quad K=0,1,\ldots m$
IV	Sum of exponentially decaying P over the available record (API)	$\sum_{i=0}^{r} \exp\left(-\frac{i\Delta t}{K}\right) P(t-i\Delta t) \quad K \in [0, m\Delta t/2]$
V	Sum of exponentially decaying P-PET over the available record	$\sum_{i=0}^{r} \exp\left(-\frac{i\Delta t}{K}\right) (P(t-i\Delta t) - PET(t-i\Delta t)) \quad K \in [0, m\Delta t/2]$

Regression trees algorithm

Grow

- split selection: Least Squares (or LAD)+partition diversity
- rule for assigning a value (prediction) to samples in a partition: partition mean (or median)
- rules to determine when a partition is terminal: homogeneity, size

Prune

test on an independent data set: if prediction error of a partition is less than that of the sub-tree rooted in it then prune (REP)

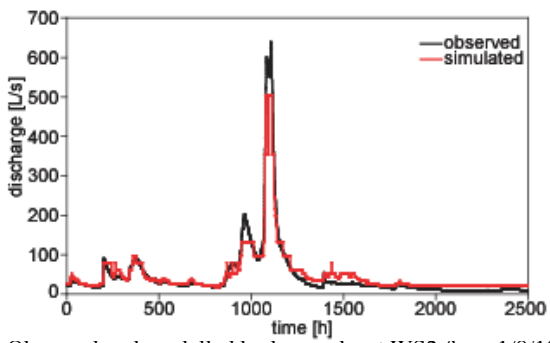
Study site and data

- Andrews Experimental Forest Watershed : WS1 and WS2
- 41 years of 15' rainfall and runoff; PET guestimate (850 mm/y sinusoidal)

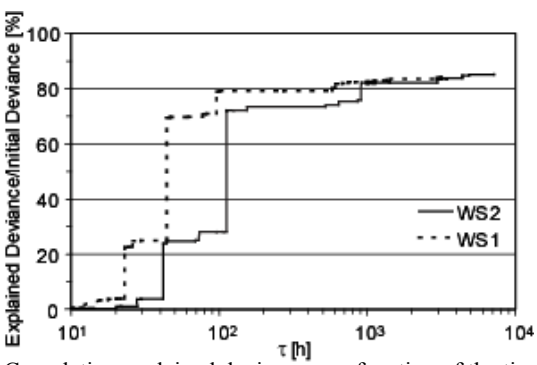
Main Results

- Performance: over 80% of variance explained in both calibration and validation; competitive with state-of-the-art hydrological models
- Allows to identify heteroscedascity & outliers

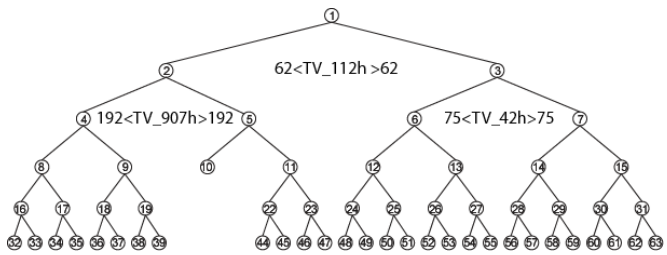
  (related to snow melt/accumulation in this case)



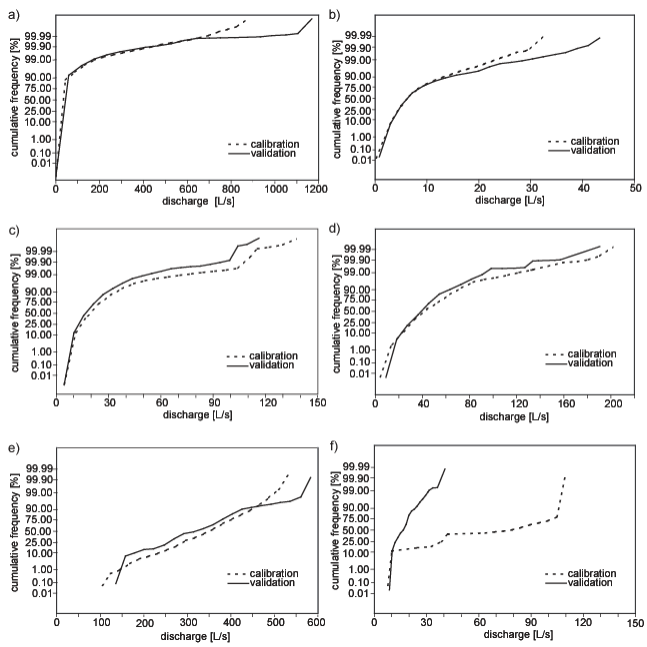
Observed and modelled hydrographs at WS2 from 1/8/1986 to 4/22/1986 (validation period); same level 'stairs' belong to a same partition.



Cumulative explained deviance as a function of the time scale (τ) characterizing split variables.



Depth-five regression tree for WS2; 29 terminal partitions



Cumulative distributions for the calibration and validation periods (normal probability paper): a) partition #1 (whole period); b) partition #33; c) partition #10; d) partition #46; e) partition #62; f) partition #20 (pruned).

Perspectives

- Expand the set of investigated predictor variables.
- Better account for autocorrelations in variables.
- Build hybrid models & implement bagging random forests.

Applications

- Data quality control (outlier detection)
- Catchment comparative and change analysis
- Multicriterion evaluation of parametric 'state' hydrological models.