Modeling flash floods in ungauged mountain catchments of China: A decision tree learning approach for parameter regionalization

Silvan Ragettli[1], Jian Zhou[2], Haijing Wang[1], Changjun Liu[3,4], Liang Guo[3,4]

[1]hydrosolutions ltd, Zurich, Switzerland [2]Northwest Institute of Eco-Environment and Resources, CAS, Lanzhou, Gansu, P.R.China, [3]China Institute of Water Resources and Hydropower Research, Beijing, PR China, [4]Research Center on Flood & Drought Disaster Reduction of the Ministry of Water Resources, Beijing, PR China

AGU FALL MEETING
New Orleans | 11-15 December 2017
AGU2017-H53D-1477

hydrosolutions

Introduction and goals

Rainstorm flash floods are a common and serious phenomenon during the summer months in many mountainous regions of China. 1970-2010 disaster areas cover more than 12% of the total land area. In this study we develop a modeling strategy for the simulation of flood events in small river basins. One of the main challenges is finding appropriate parameter values for simulating flash floods in ungauged catchments. The main goals of this study are:

I. To develop a new parameter regionalization method that is based on decision tree learning. We use decision trees to identify catchment similarities that are related to good parameter transfer performance. We assess if this method can effectively support decision making for hydrological modeling in ungauged catchments.
II. To identify the dominant controls on successful hydrologic parameter transfer in mountainous mesoscale catchments of China
III. To assess the applicability of the hydrological model PRMS-OMS for modeling rainstorm flash floods in China.

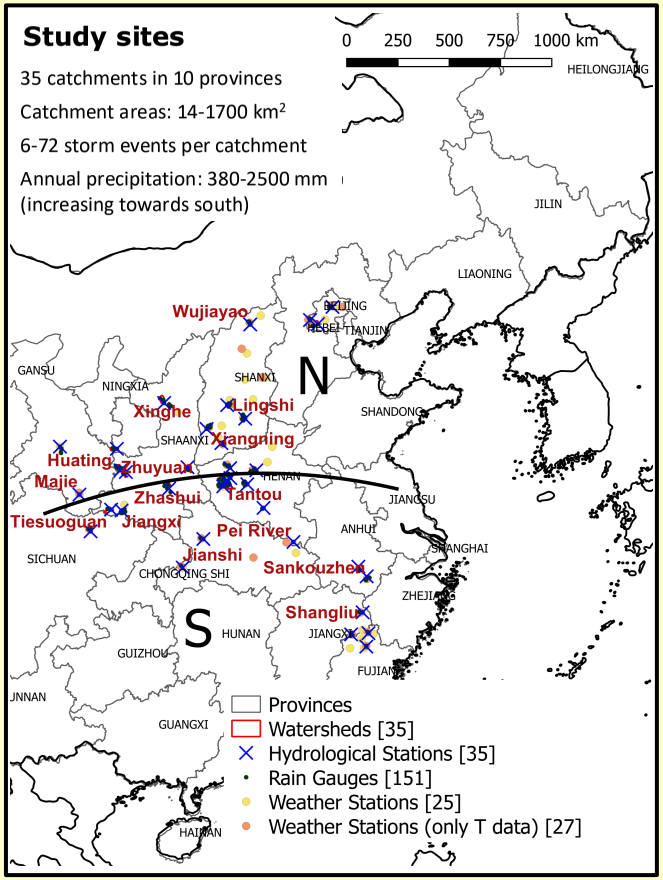
Study sites

35 catchments in 10 provinces

Catchment areas: 14-1700 km²

6-72 storm events per catchment

Annual precipitation: 380-2500 mm (increasing towards south)



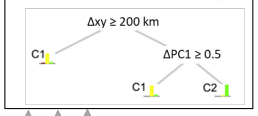
Study design

1. Model calibration (each catchment separately). Automated procedure: Shuffled Complex Evolution technique

2. Perform parameter transfer across all catchments

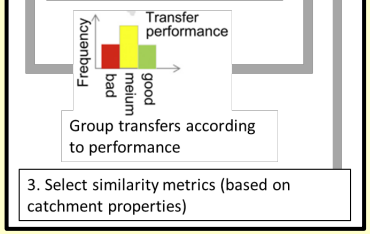
3. Select similarity metrics (based on catchment properties)

4. Relate high/low similarities with good/bad transfer performance
   - I. Regression tree analysis
   - II. Classification tree analysis



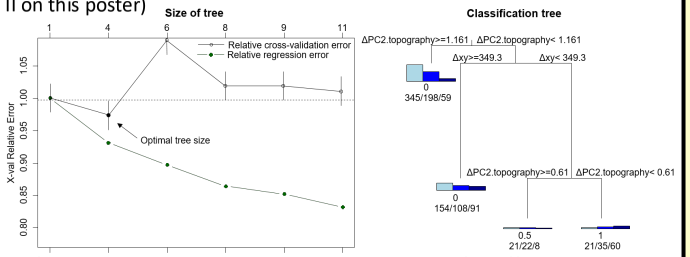
5. Apply splitting rules of classification/regression trees leading to the branches with best transfer performance to identify parameter sets for ungauged catchments

6. Evaluate model performance for assumed ungauged catchments



I. Decision tree learning

Decision tree learning is a predictive modelling approach used in statistics, data mining and machine learning. We use classification and regression trees (CART) to explore the catchment similarities – model performance space after we transfer parameter sets between the study catchments. The training set of data consists of 1190 parameter transfers (35 receiver catchments with each 34 donor catchments) and 15 catchment similarity metrics (see section II on this poster)

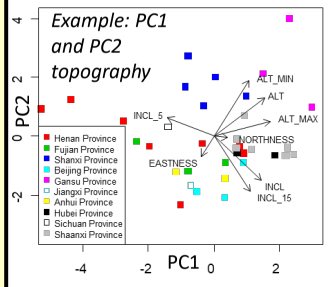


Left: Relative cross-validation and regression errors for different tree sizes. Right: Classification tree relating clusters of different transfer performance with catchment similarity metrics.

II. Variable Importance

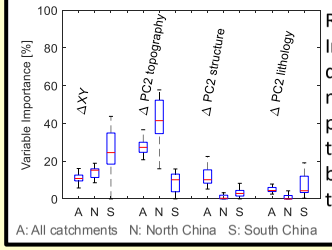
Name	Characteristics (Abbreviation)
Topography Landform	ALT, ALT_MIN, ALT_MAX, INCL, INCL_5, INCL_15, EASTNESS, NORTHNESS
Lithology	PLAIN, MED. GRADIENT HILL, MED. GRADIENT MOUNTAIN, HIGH. GRADIENT HILL, HIGH. GRADIENT MOUNTAIN
Soil texture Landuse	GRANITE, GRANODIORITE, BASALT, IGNIMBRITE, GNEISS, SLATE, SANDSTONE, SILTSTONE, SHALE, LIMESTONE, CALCAREOUS, FLUVIAL, BAUXITE
	SAND, SILT, CLAY
	CROPS_IRRIGATED, CROPS_RAINFED, CROPLANDS_MOSAIC, VEGETATION_MOSAIC, EVERGREEN_CLOSED2OPEN, EVERGREEN_CLOSED, FOREST_CLOSED2OPEN, SHRUBLAND_MOSAIC, GRASSLAND_MOSAIC, SHRUBLAND_CLOSED2OPEN, GRASSLAND_CLOSED2OPEN, ARTIFICIAL
Climate	PREC_ANNUAL, PREC_MAX1H, PREC_MAX24H, T_MEAN, T_MIN_MONTHLY, T_MAX_MONTHLY, ARIDITY, PETJP
Drainage	TOPWET, STREAM_DENSITY, 1st_ORDER, 2nd_ORDER, 3rd_ORDER, STRAHLER_MAX
Structure	AREA, HRU_SIZE, CHANNEL_LENGTH

58 catchment attributes (divided in 8 categories) that may affect catchment hydrology.



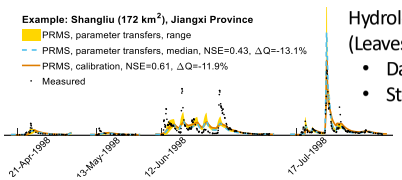
Principal component (PC) analysis to extract uncorrelated similarity metrics. The PCs that explain at least 66.6% of the total variance per category are retained.

14 catchment similarity metrics plus Δxy (the spatial distance) are considered for the regression analysis.



Results: Importance of different similarity metrics to assess parameter transferability, based on decision tree analysis.

III. Hydrological model performance

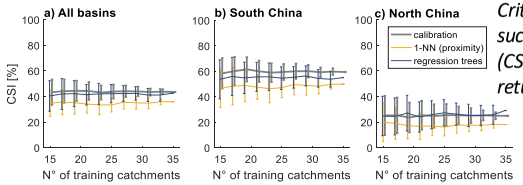


Hydrological model: PRMS-OMS (Leavesley et al. 2006)
- Daily mode: 24-h time steps
- Storm mode: 1-h time steps

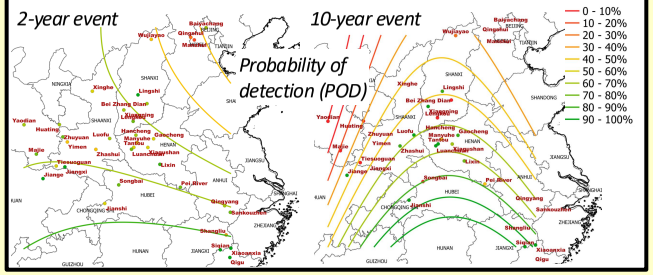
Contingency scores to evaluate the models' ability to accurately issue warnings for events that exceed a given return period threshold:

$$POD = \frac{H}{H+M} \qquad CSI = \frac{H}{H+M+FA}$$

H: HITS
M: MISSES
FA: FALSE ALERTS



Critical success index (CSI), 10-year return period



Conclusions

- Decision tree learning optimally utilizes the information content of available similarity descriptors to identify suitable gauged donor catchments for parameter transfer to to ungauged catchments.
- CART offer flexible rules for selecting suitable donor catchments, since several decision-making paths can be followed. Our approach is most useful for sparsely gauged and topographically complex environments: spatial proximity can be used as a selection criteria but is skipped in the case where no similar gauged catchments are in the vicinity.
- PRMS-OMS can be used to simulate rainstorm flash floods in China. However, for the semi-arid northern catchments, the probability of detection of a 10-year flood is still less than 50%. For an operational use of PRMS-OMS, therefore, further model improvements are recommended to increase calibration and regionalization performance.

References:
Leavesley, G., Markstrom, S., Viger, R. (2006): USGS Modular Modeling System (MMS) - Precipitation-Runoff Modeling System (PRMS); Watershed Models. CRC Press, 159-177.
Ragettli et al. (2017): Modeling flash floods in ungauged mountain catchments of China: A decision tree learning approach for parameter regionalization. Journal of Hydrology, Vol. 555, 330-346

ragettli@hydrosolutions.ch