Catchment-scale spatial importance for better understanding the discharge generation processes

Tomasz Berezowski and Tomasz Bieliński

Gdansk University of Technology, Faculty of Electronics, Telecommunication and Informatics, Department of Geoinformatics (tomberez@eti.pg.edu.pl)

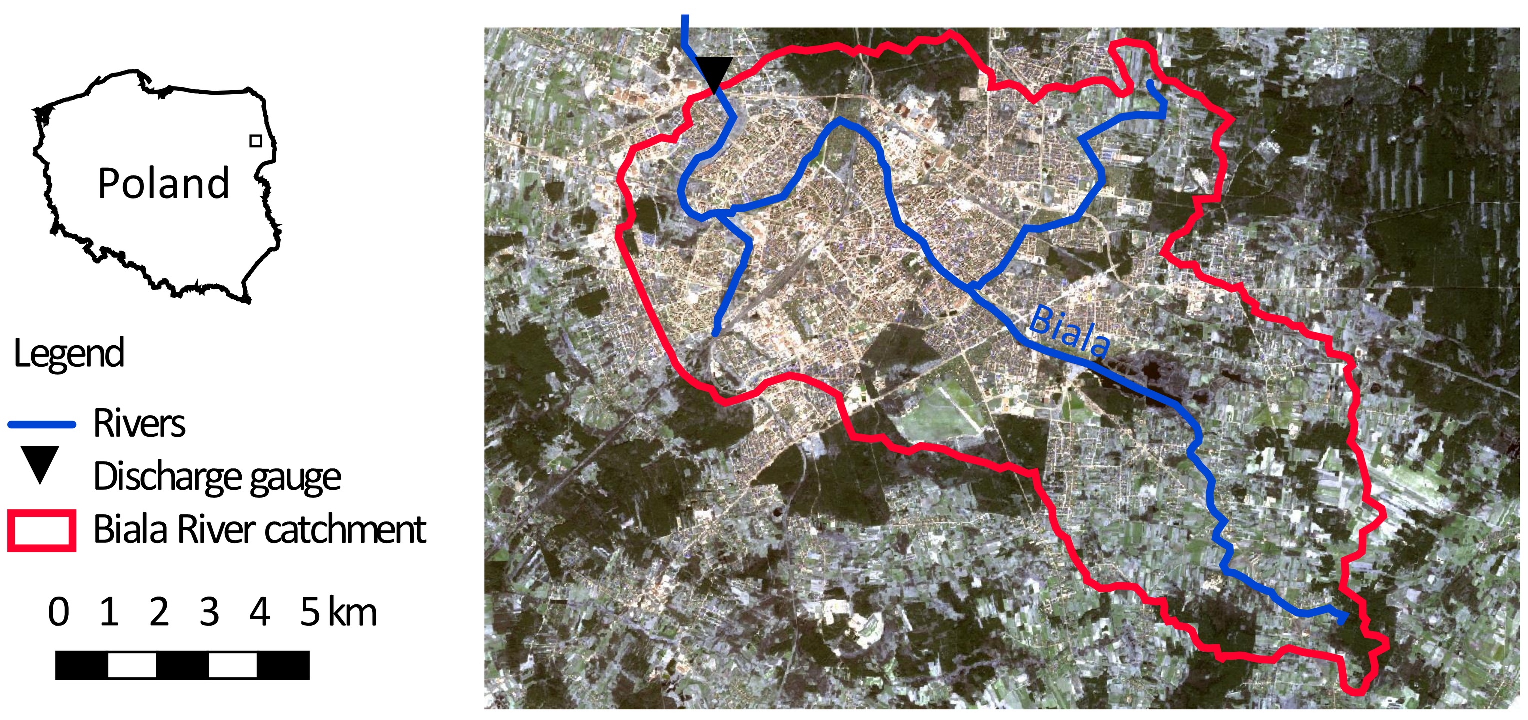
Introduction

Identification of discharge generation hotspots is important for undertaking actions aimed at flood protection or drought reduction. Such hotspots may be identified by integrated hydrological models (especially in natural areas), sewer system hydrodynamic models (especially in urban areas) or using a spatial sensitivity analysis method. Each of the mentioned methods requires a trained and validated hydrological in order to obtain meaningful results. This is often costly in terms of required data, labor and computation time.

In this study we propose a simple method for identification of discharge generation hotspots at a catchment scale using a machine learning approach and a big dataset. Our approach uses a functionality of the Random Forests algorithm that allows to quantify predictors importance in a model. The predictors, however, are not lumped (e.g. catchment average rainfall), but spatially distributed (e.g. rainfall field in an appropriate spatial resolution). Hence, a spatial pattern of the predictors importance for the model output (e.g. discharge at a catchment outlet) is obtained.

Methods

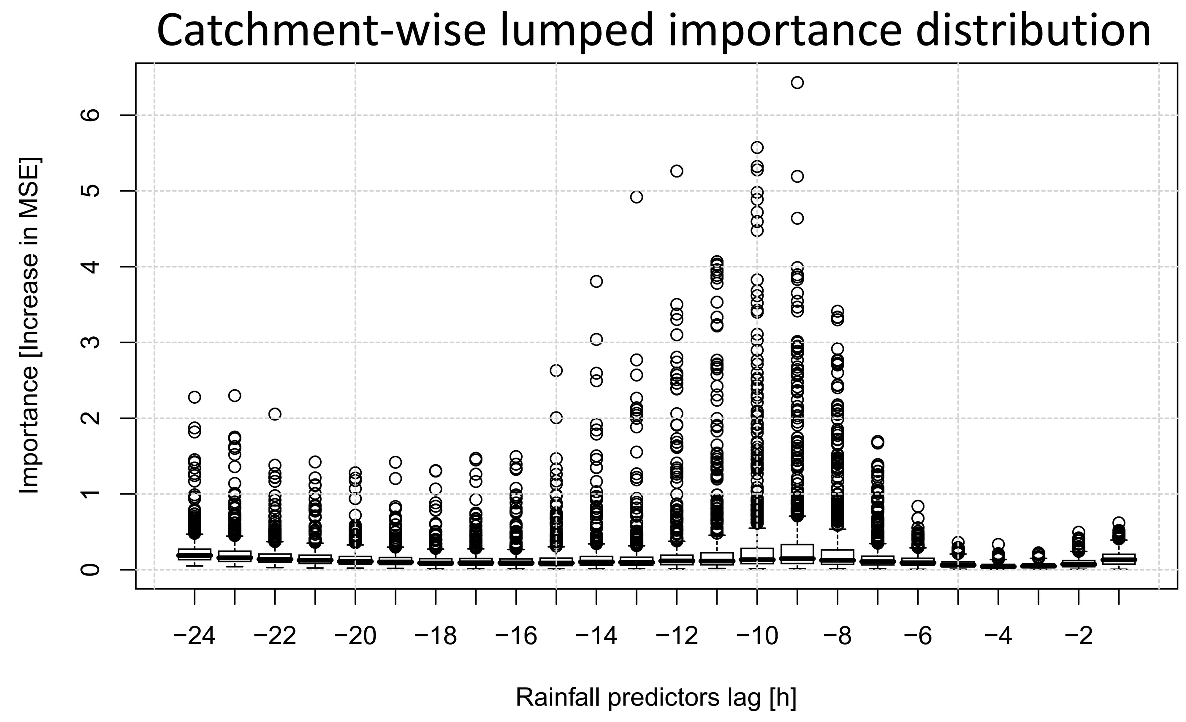
In this study we use a machine learning algorithm - Random Forest - for discharge prediction in 1h resolution in the period VI 2015 –VI 2016 in a 105.9 km^2 urbanized catchment in NE Poland - Biala River. The meteorological data used as the predictors are rainfall fields obtained from meteorological radars in 10 min temporal and 1 x 1 km spatial resolutions. Each grid cell in the radar data is a predictor used to explain discharge in a Random Forest model. The Random Forest built-in importance feature is used to highlight important grid cells – i.e. Important for discharge generation zones in the catchment



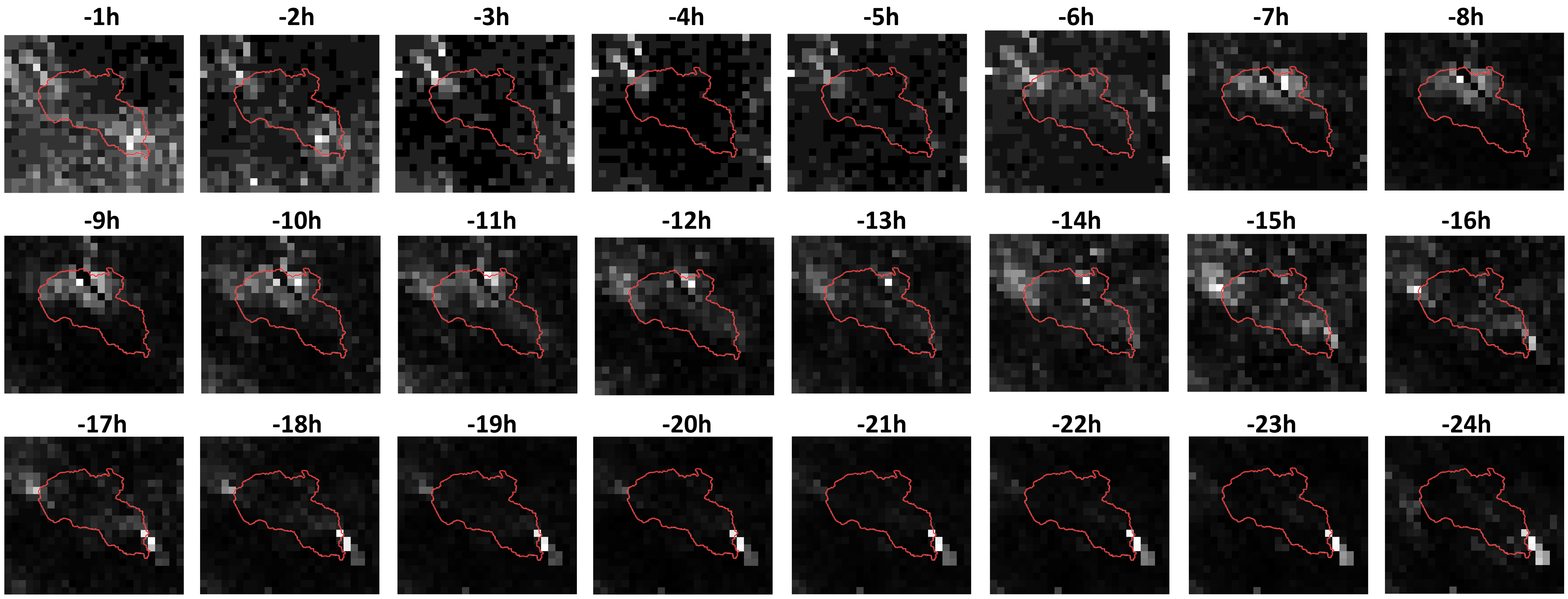
Results

- Catchment-wise lumped importance shows clear temporal pattern, showing that rainfall delayed 8-13 h is the most important for discharge modeling.
- The grid-wise distributed importance shows that important clusters appears in different parts of the catchment while rainfall predictors lag time advances.
- Some grid cells show high importance outside the topographical catchment. This is likely due to presence of sewer systems that are not taken into account for topographical catchment delineation and due to the fact that radar data is resampled from lower resolution into 1x1 km grid.

Catchment-wise lumped importance distribution



Grid-wise importance maps for subsequent rainfall predictors temporal lag



Conclusions

The spatial importance pattern clearly relates to land-cover features like impervious surfaces and show temporal dependence. This preliminary study shows that it is possible to look at the runoff generation process using Random Forests derived importance, however, this topic requires further research focused at the spatial scale of the problem and sources of noise in the data and results.