Lomonosov Moscow state university, Faculty of Geography, Hydrology Department

Objective classification of changes in water regime types of the Russian Plain rivers utilizing machine learning approaches

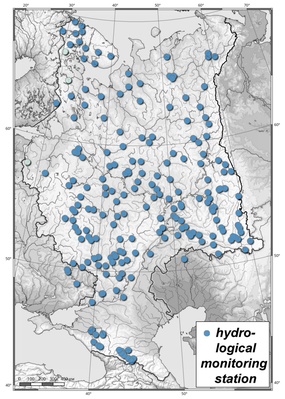
Alexander Ivanov, Timophey Samsonov, Natalia Frolova, Maria Kireeva, and Elena Povalishnikova

The study was supported by the Russian Science Foundation (grant No.19-77-10032) in methods and Russian Foundation for Basic Research (grant No.18-05-60021) for analyses in Arctic region

MAIN GOAL: establishment of a river classification based on water runoff dynamics with contemporary machine learning and clusterization methods

OBJECTIVES:

- To create an objective river classification algorithm based on machine learning algorithms
- To establish how did water runoff dynamics change over the course of last 70 years and the impact of Climate Change on it
- To renew the hydrological characteristics of European Russia rivers taken into account in hydro-meteorological engineering survey and structure design



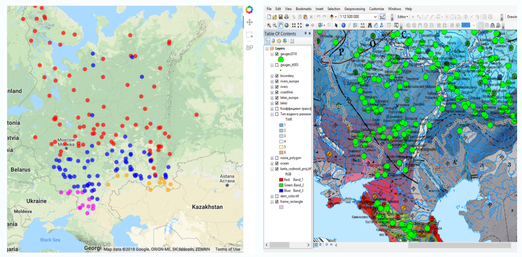
DATA:

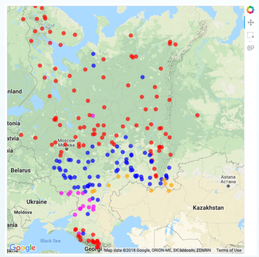
Monthly discharge for 232 representative watersheds, evenly located on the European part of Russian Federation

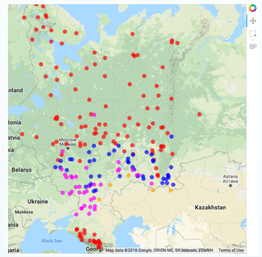
(shown on figure to the left)

1 Classification Based On Existing Runoff Types

Classification using Gradient Boosting: results comparable to human-made. Classificator identifies changes in water runoff types for southern rivers







Top Left: River classification by a human based on reference runoff dynamics types

Top Right: Reference color map (Evstigneev et.al,1990)

Bottom Left: Gradient Boosting classifier output for an aggregated data for 1945-1977

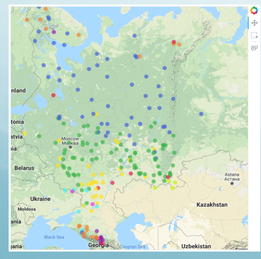
Bottom Right: Gradient Boosting classifier output for an aggregated data from 1978 up to 2006

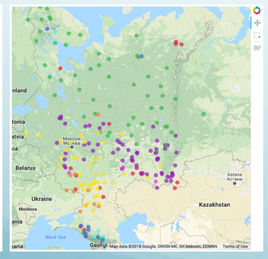
By comparing top and bottom left images you can clearly see that automatic classification results are very similar to the human-made. The most notable difference is several misrepresented yellow points in the lower part of the map (model identified 90% of datapointsa correctly).

By comparing bottom left and right we see that automated classification shows changes in runoff dynamics for pink and blue classes. These changes correspond to the ones we see in real life in Don river basin.

2 Clustering on Aggregated Data

DBSAN algorithm shows best results. Clustering also show changes in runoff types in southern regions





Left: Results of clustering using DBSCAN on aggregated data for 1945-1977

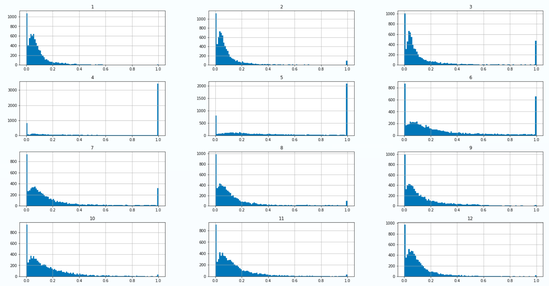
Right: Results of clustering using the same algorithm on aggregated data from 1978 up to 2006 (colors doesn't match)

Different clustering algorithm (K-means, DBSCAN, agglomerative) were used. Due to the nature of data DBSCAN showed the best results.

According to the results 19 river in south Russia (green on the left) started to consistently behave more like southern ones (yellow and pink on the right) after 1977. These changes are more drastic than when using classifier from (1). However during the course of the work it was also determined that several rivers aren't' fixed to the single cluster and several runoff types aren't really different from each other when using average data over long period. Therefore it was determined that yearly data should be used for clustering approaches.

3 Clustering on Non-Aggregated Data to determine year-to-year changes for different gauges

As a follow-up to clustering on aggregated data we decided to make an approach using non-aggregated data for the last 80 years

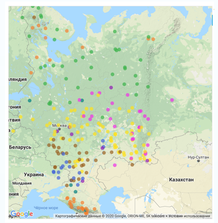


A histogram for monthly runoff values of all rivers in European Russia over the course of the last 70 years normalized over maximum yearly value for each year.

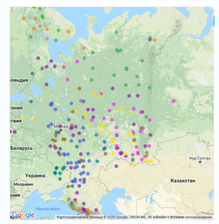
To determine changes in water regime over the course of year we decided to make an attempt at establishing inter-year clustering. By doing this we also try to verify a hypothesis that runoff types stay the same with the climate change, but river gauges can exhibit behavior of the different regime.

As seen on the histograms to the left data for March-June while displaying clear peaks also shows continuous stream of values that makes naive attempts at clustering difficult. There're no clear borders between rivers with different runoff types due to year-to-year variations.

We made two different attempts at establishing inter-year clustering: 1) Using K-means with PCA 2) Using cosine similarity metric.



1966

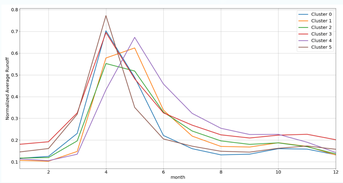


2007

Clusters determined using K-Means algorithm

For cosine similarity it was possible to determine such a cutoff that resulting clusters seem to represent real classes for the year (judging by the visualization of river hydrographs). For K-Means with PCA optimal number of clusters and components were set according to the best silhouette score values. Clusters were averaged over the 600 iterations of algorithm.

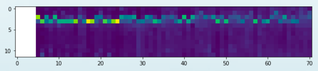
Main hydrograph types automatically determined by algorithm can be seen below



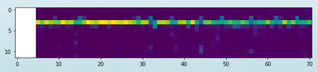
Transfers between hydrograph types can be seen for each year and for each of 232 stations and used to establish global trends.

4 Visualizing Changes in Runoff Types

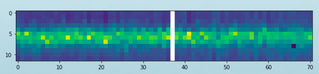
During the course of work a fast way to understand if a river runoff type has changed has been developed



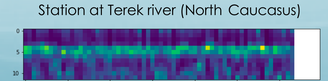
Station at Don river basin



Station at lower Volga river basin



Station at Terek river (North Caucasus)



Station at Vuoksi river

It's hard to answer if the river has changed it's runoff type or it's just a fluctuation in this year runoff. Visualization often helps. We found the following approach most useful:

- First, we normalize each year runoff values by dividing them by the maximum value of the same year. Therefore each value for each year lies in [0; 1] range.

- Next we stack them up along a time axis with most recent years on the right and past on the left of x axis.

- Finally we color-code the resulting set setting maximum value to bright yellow and lowest value to dark purple (standard matplotlib approach).

Looking at Don river station it's clear that maximum runoff became unstable around year 30 from the start of observation. The force of spring flood diminished over time and its quite possible that river changed its runoff type in the recent years. Volga river runoff has also changed, but overall type looks the same or quite close to the values at start of the observation.

5 Conclusions

- An automated classifier for established runoff types was built
- By using the classifier we can determine that southern runoff river types can be registered further north than in the period of 1945-1977
- Classic clustering algorithms identify stronger changes, but aren't as reliable.
- A better approach to clustering can be made using non-aggregated data and either K-Means with PCA or Cosine Similarity metric. As a result we can track changes in river regimes on a year-to-year basis