Estimation of the Aral Sea state predictability based on the open data sources and the unique field observations

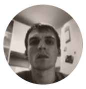
Poster · April 2016

DOI: 10.13140/RG.2.1.5096.4887

4 authors, including:

Alexander Izhitskiy

P.P. Shirshov Institute of Oceanology



Georgy Vladimirovich Ayzel

Russian Academy of Sciences

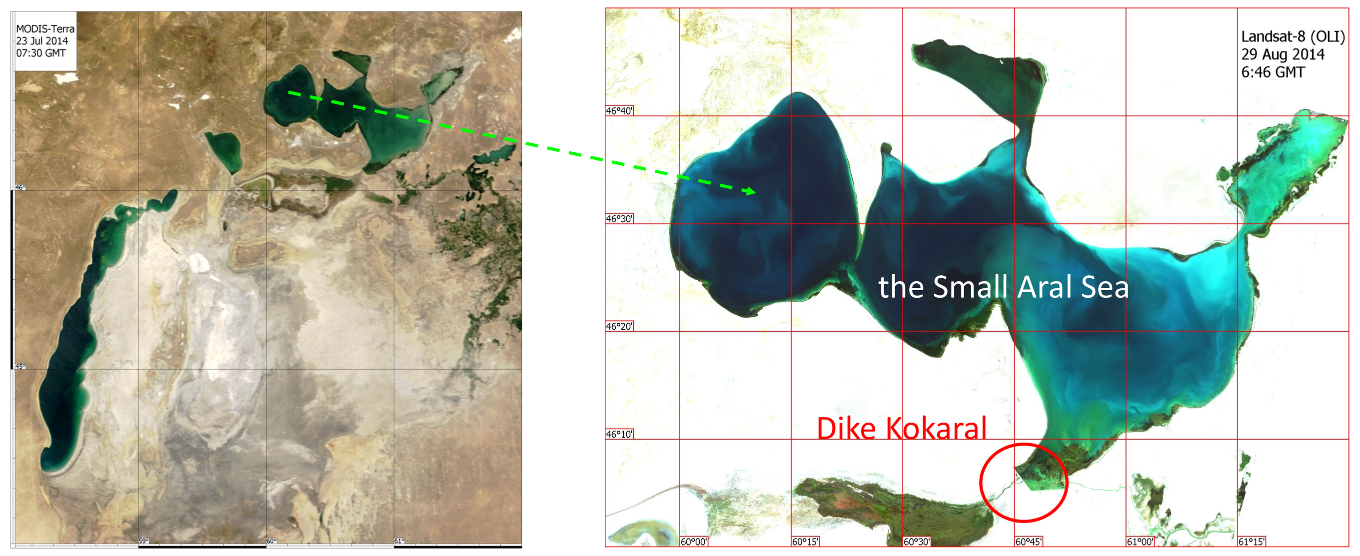
Estimation of the Aral Sea state predictability based on the open data sources and the unique field observations

Alexander Izhitskiy[1], Georgy Ayzel[2], Peter Zavialov[1] and Abilgazi Kurbaniyazov[3]

1 - Shirshov Institute if Oceanology Russian Academy of Sciences, Moscow, Russia; 2 - Institute of Water Problems Russian Academy of Sciences, Moscow, Russia; 3 - Yasavi International Kazakh-Turkish University, Turkestan, Kazakhstan.

1. Introduction

- The Aral Sea, formerly one of the four largest lakes in the world, has lost over 90% of its volume during the dramatical dessication mainly caused by the severe alteration of water budget of the basin.



- The Small Aral separated from the main part of the sea in 1988-1989. Subsequently, several attempts to build a dam in the Kokaral strait were undertaken. Final construction was finished in 2005.
- This retained the Syr Daria river discharge within the Small Aral and facilitated an increase in the lake's water level.
- The main idea of the study is to estimate the possibility of prediction of the Small Aral Sea state using basic geoanalysis tools and state-of-art machine learning techniques.

2. Data

- Water level variability (the Small Aral Sea): from the open Database for Hydrological Time Series of Inland Waters (**DAHITI**) for period 1992-2014 (3 times a month) [1]
- River Discharge (Syr Daria): historical data (monthly, 1979-1986) provided by the Global Runoff Data Center (**GRDC**) [2]
- Atmospheric forcing: Era-Interim, 1x1 degree resolution, daily [3]
- Salinity: historical data (annually from 1992 to 2002) for the Small Aral Sea presented in [4]; values for the Small Aral Sea water obtained during the last field surveys of Shirshov Institute of Oceanology in 2014-2015 [5]

2. Methods: Machine Learning

Key modeling concept is based on implementation of a simple Decision Tree model in case of regression task. Typical Decision Tree model is a "white box" consists of the range of boolean classifiers which split our samples to tiny "leaf" nodes where all samples constantly refers to the one target value. Single tree-based implementation of Decision Tree algorithm faced with the case of overfitting and robustness lack that lead to limited using in real world examples. In our work we used three cutting-edge machine learning techniques based on ensemble approach to predictions which are based on ensembles of simple Decision Tree models.

Random Forest and Extra Trees:

Each tree in ensemble builds on a random split sample of training data. In addition, Extra Trees have one more randomization procedure inside the tree in the moment of choosing the best split among the features.

pros:
- powerful algorithm allows to recover complex relations in data;
- does not require features scaling;
- robust to noisy data;
- simple parallelization of calculations (each tree can grow independently);
- have a few hyperparameters with simple interpretation;
- ncreasing of number of trees in ensemble does not lead to overfitting.

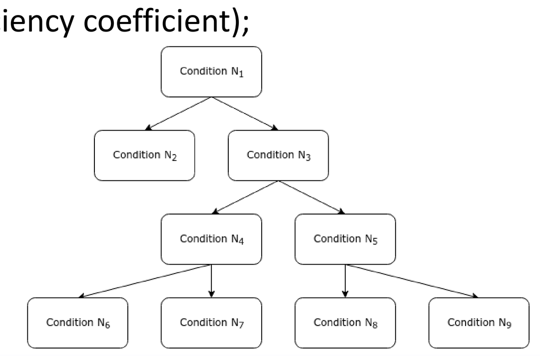
cons:
- needs a lot of time for training …
- and as a result - provide "freezing" calculations on a new data;
- hard to fit on complex quality measure (when differ from default R2);
- does not work well on sparse data.

Gradient Boosting:

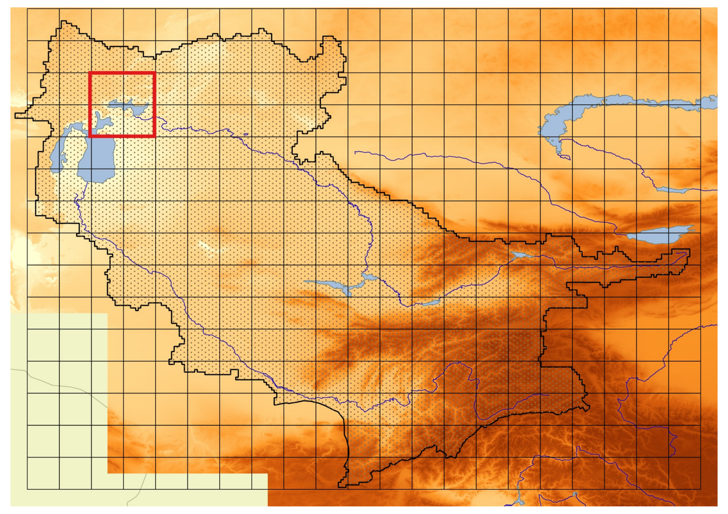
Each tree in ensemble learns consequentially on the residuals of previous tree predictions (directed training). Trees for ensembles can be weak learners with small depth (3-6 levels).

pros:
- the most powerful algorithm in machine learning (Kaggle.com competitions winner);
- does not require features scaling;
- robust to noisy data;
- can fit on every differentiable loss function (e.g. Nash-Sutcliffe efficiency coefficient);
- fast learning algorithm;

cons:
- a lot o f sensitive hyperparameters;
- requires high attention to number of trees to avoid overfitting;
- does not work well on sparse data;
- can not be parallelized.



4. Data driven models



Syr Darya monthly runoff predictions

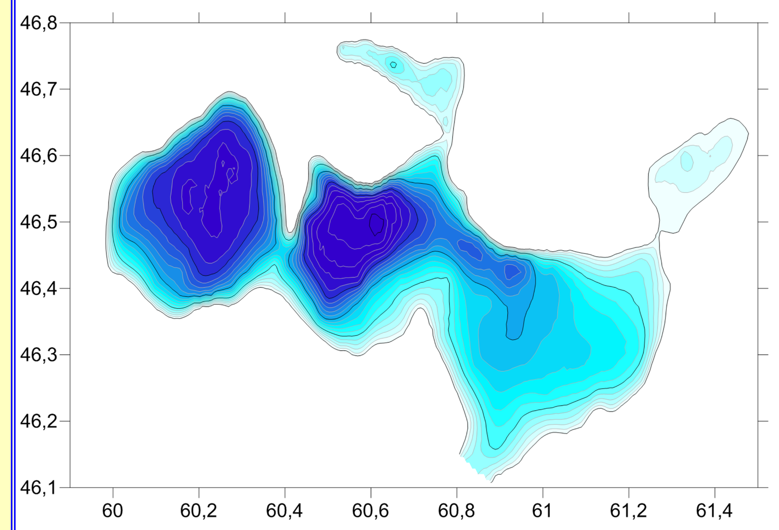
- **Gauge station**: Kazalinsk (130 km from the mouth);
- **Features (predictors)**: monthly air temperature and total precipitation for the current and six previous months from Era-Interim data averaged over the Aral Sea basin;
- **Target**: monthly runoff values in Kazalinsk (Kazakhstan);
- **Training period**: July 1979 - December 1985 (7 years);
- **Prediction period**: January 1986 - September 2015 (30 years).

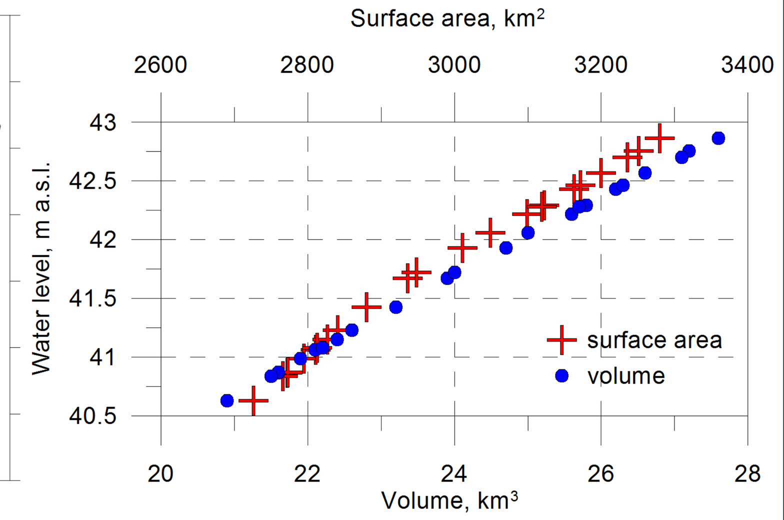
The Small Aral Sea volume predictions

- **Features (predictors)**: monthly air temperature, sea-surface temperature, total precipitation, wind velocity data from Era-Interim and Syrdarya inflow data from our runoff formation model for current month without any shift;
- **Target**: Small Aral Sea volume derived from the recalculation of water level provided by DAHITI project and digital bathymetry map;
- **Training period**: October 1992 - November 2014 (23 years);
- **Prediction period**: December 2014 - September 2015 (1 year).

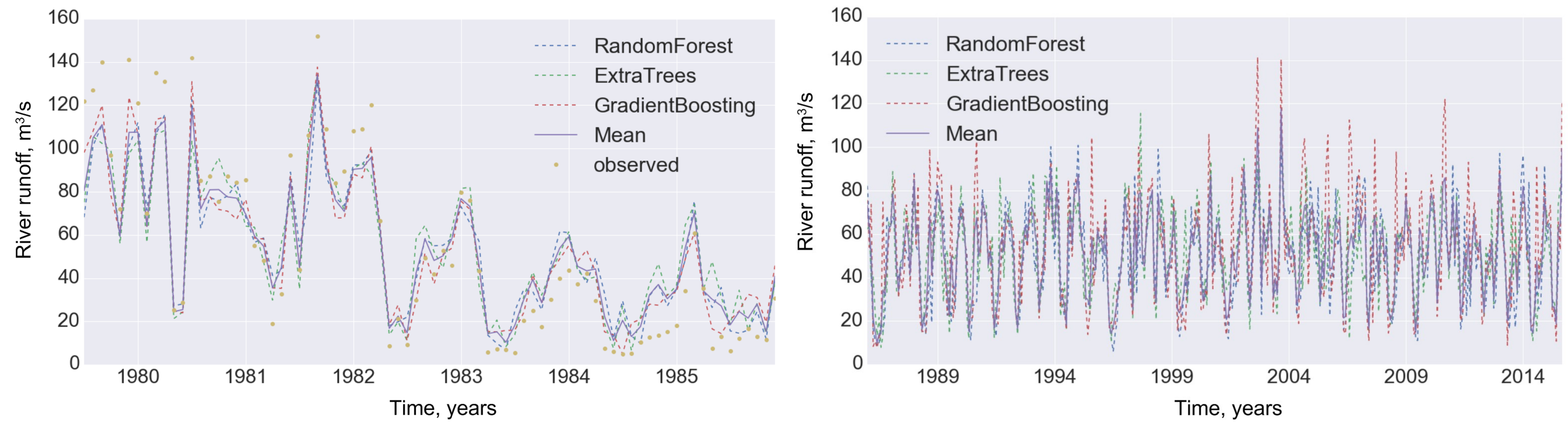
5. Morphological settings

- The hypsometric relation for the Small Aral Sea was obtained. Volume if the reservoir was calculated using the detailed bathymetry for the range of water level variability presented in DAHITI database.
- Linear regression connecting volume of the Small Aral and its water level variability was derived and subsequently implemented in a model experiment.

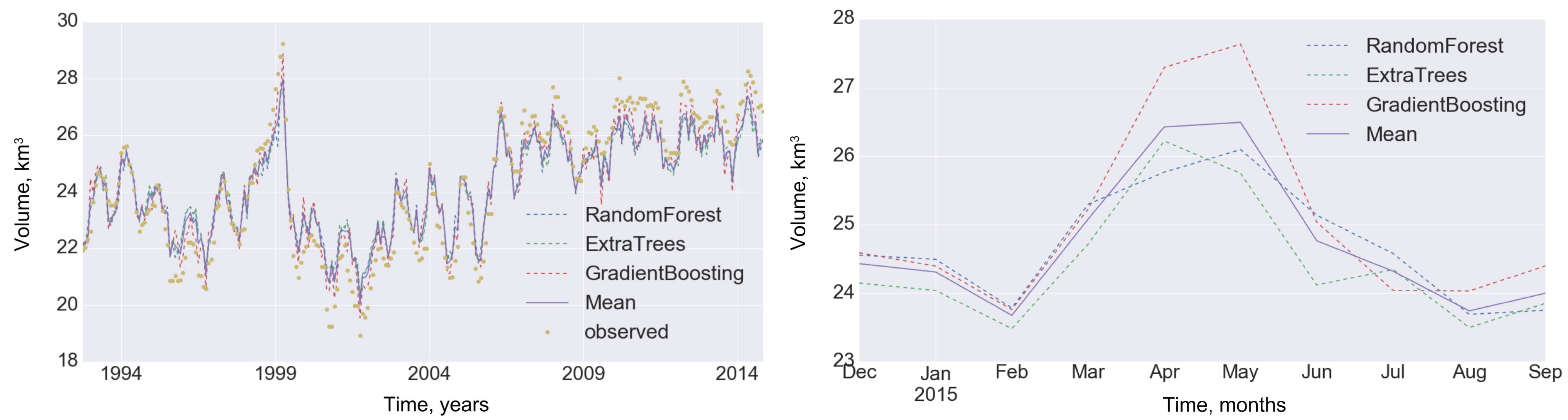




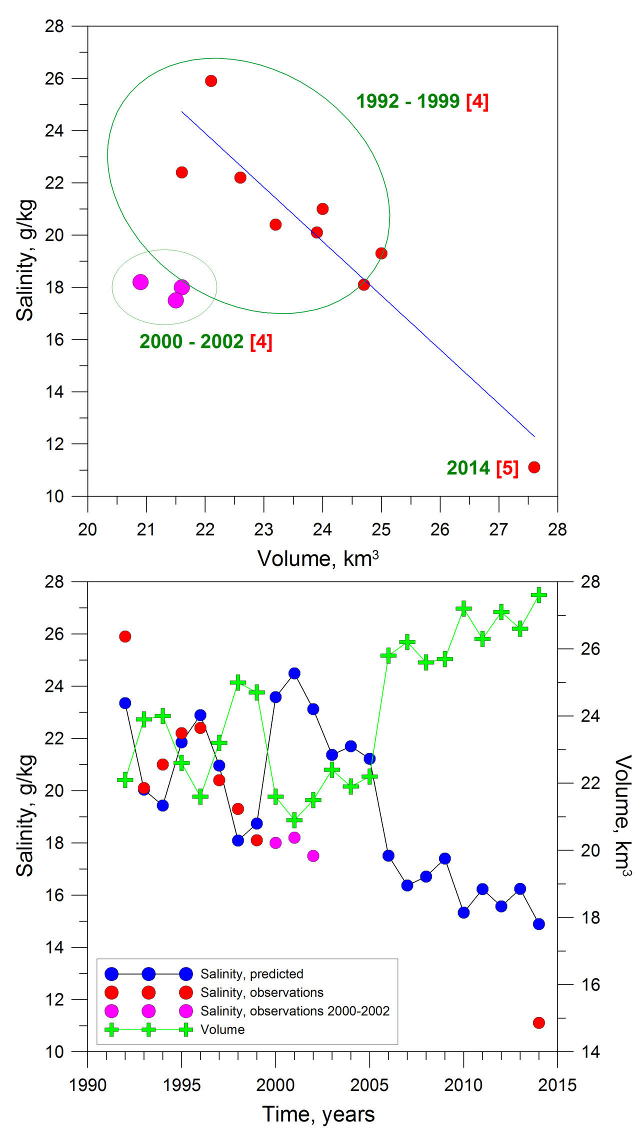
6. Runoff prediction results



7. Small Aral Sea volume prediction results



8. Salinity variations



8. Conclusions

- Prediction of the hydrological state of the Aral Sea faces a lot of difficulties due to severe lack of direct observations of principal characteristics of water balance, such as water level, river inflow etc.
- Despite the fact, approach presented provides sufficiently accurate prediction of the Small Aral Sea volume basing on the open-access data sources.
- Implementation of non-considered dam factor to the data-driven model shall reduce an error of lake volume prediction. However, there are no available data on water dropping through the dyke and irrigation system over the basin.
- Presented methodic could be coupled with the hydrodynamics models providing input data for the model simulations and predicting the future scenarios of the Aral's hydrological system evolution.

References

1. *Schwatke, C. et al.* DAHITI - an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry, Hydrol. Earth Syst. Sci., 19, 4345-4364, 2015
2. Monthly runoff data. The Global Runoff Data Centre, 56068 Koblenz, Germany. www.bafg.de
3. *Dee, D.P. et al.* The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137: 553–597. doi: 10.1002/qj.828
4. *Aladin, N. et al.* Modern hydro-biological state of the Small Aral sea. Environmetrics, 16, 375-392, 2005
5. *Izhitskiy, A.S. et al.* Present state of the Aral Sea: diverging physical and biological characteristics of the residual basins. Scientific Reports, 6, 23906 (2016) doi:10.1038/srep23906