Open Access
PIAHS

Innovative water resources management – understanding and balancing interactions between humankind and nature

Coupling physically based and data-driven models for assessing freshwater inflow into the Small Aral Sea

Georgy Ayzel[1,2] and **Alexander Izhitskiy**[3]

[1]Institute of Earth and Environmental Science, University of Potsdam, 14476 Potsdam, Germany
[2]Institute of Water Problems, Russian Academy of Sciences, 119333 Moscow, Russia
[3]Shirshov Institute of Oceanology, Russian Academy of Science, 117997 Moscow, Russia

Correspondence: Georgy Ayzel (ayzel@uni-potsdam.de)

Abstract. The Aral Sea desiccation and related changes in hydroclimatic conditions on a regional level is a hot topic for past decades. The key problem of scientific research projects devoted to an investigation of modern Aral Sea basin hydrological regime is its discontinuous nature – the only limited amount of papers takes into account the complex runoff formation system entirely. Addressing this challenge we have developed a continuous prediction system for assessing freshwater inflow into the Small Aral Sea based on coupling stack of hydrological and data-driven models. Results show a good prediction skill and approve the possibility to develop a valuable water assessment tool which utilizes the power of classical physically based and modern machine learning models both for territories with complex water management system and strong water-related data scarcity. The source code and data of the proposed system is available on a Github page (https://github.com/SMASHIproject/IWRM2018).

1   Introduction

The Aral Sea and its basin are among the highly recognizable examples of significant environmental changes which took place in the Central Asia during the last decades (Izhitskiy et al., 2016; Micklin, 2007; Raskin et al., 1992; Zavialov et al., 2003). Induced by river runoff exploitation across huge irrigation systems the Aral Sea level has significantly decreased and run irreversible ecosystem and water balance shifts (Zmijewski and Becker, 2014). Nowadays the Small Aral Sea has a limited (by the Kokaral Dam) hydrological connection with dying southern sea basins and tends to stay a separate part under current social and political situation in the region. It is extremely important to devote scientific attention to this region as a real live example of the human-induced impact on water balance and its response (Immerzeel and Bierkens, 2012).

The main volume of the freshwater inflow into the Small Aral Sea is formed on the Syr Darya river basin which is among the largest and highly vulnerable river basins in the Central Asia. There are thirteen large reservoirs and much local water management related installations on the Syr Darya river and its tributaries which utilize full freshwater poten-

tial for irrigational, industrial, recreational, and social needs. This complex structure of water management system coupled with the total absence of data describes its functioning is a challenge for any approach directed to the accurate assessment of the Small Aral Sea freshwater budget formation and evolution across the basin (Lutz et al., 2012a; Raskin et al., 1992; Sorg et al., 2014).

There are three main categories of scientific literature devoted to the identification of modern water balance shifts in the Aral basin. The first group accumulates research directed to large-scale heat and water flux changes assessment based on remote sensing, climate modeling and re-analysis data for a whole basin area (López et al., 2017; Shi et al., 2014; Zmijewski and Becker, 2014). These investigations help us to identify patterns and key factors affect long-term hydrological changes and its trends (geographical approach), but cannot be easily scaled for providing quantitative predictions. The second group focuses attention mostly on the upstream Aral Sea basin area (mountainous zone and Ferghana Valley) by the reason of high-altitude glaciers and the biggest reservoirs presence here. Pereira-Cardenal et al. (2011), Siegfried et al. (2012),

Hagg et al. (2006, 2007), Gan et al. (2015) and Lutz et al. (2012b) modeled runoff in glacierized catchments and its contribution into underlying reservoirs inflow using conceptual and physically based models (NAM, HBV-ETH, OEZ, SWAT, AralMountain). Apel et al. (2017) evaluated the skill of simple statistical models for seasonal runoff forecast in this region. Radchenko et al. (2017) examined historical runoff for 18 river basins in Ferghana Valley using HBV-light model and estimated projected changes in streamflow characteristics according to the A1B climatic scenario for these basins. For the extensive review of hydrological modeling studies in glacierized catchments of Central Asia please refer to Chen et al. (2017). The third (and the least) group of papers being conducted on developing end-to-end hydrological modeling system for the whole Small Aral Sea basin. A simplified approach for assessing annual freshwater inflow based on hypothetical and general circulation model based scenarios of future climate temperature and precipitation has been applied in Shibuo et al. (2007) and Jarsjö et al. (2012) using Porflow model without any parameters calibration. The most comprehensible routine for model-based assessment of water balance components of Syr Darya river basin was proposed in Lutz et al. (2012a) and utilizes coupling of conceptual runoff formation model AralMountain (Lutz et al., 2012b) with Water Evaluation And Planning model (WEAP) which has been already implemented for the former Aral Sea basin in 1989 (Raskin et al., 1992).

In presented work we have tried to combine best practices in an existing scientific literature and modern advances in the field of machine learning to develop continuous hybrid hydrological model for investigating both runoff generation processes using physically based models and runoff transformation through one of the most complex water management systems in the world using machine learning algorithms and models. With our research, we want to fill a modern gap in developing a continuous runoff prediction system for the entire Syr Darya river basin domain using a combination of state-of-the-art modeling techniques. Our research does not pretend to cover the problem of freshwater inflow predictions in the Small Aral Sea in high details, but it is an attempt to map the efficiency level of the runoff prediction system which has been built only on open data sources.

2   Materials

2.1   Study area

The main part of the Small Aral Sea basin (Fig. 1) is occupied by the Syr Darya river and its tributaries which contribute around $40 \, \text{km}^3$ of freshwater inflow annually (Radchenko et al., 2017). About 70 % of the runoff of the Syr Darya river basin originates in the Kyrgyzstan mountain ranges and the main contribution of this volume corresponds to Ferghana Valley river basins (Belyaev, 1995; Radchenko et al., 2017). In our research, we have selected 24 basins which run to

the Ferghana Valley as the main source of hydrological insights and information about runoff generation in the freshwater formation zone of the Small Aral Sea (Fig. 1). These basins are highly contrasting in geographical and hydroclimatic conditions, and cover a range of areas from 150 to $24\,000 \, \text{km}^2$. For a detailed geographical description of Ferghana Valley river basins please refer to Radchenko et al. (2017).

2.2   Runoff and meteorological forcing data

Observed runoff data for selected basins were provided by the Global Runoff Data Centre (GRDC; http://www.bafg. de/GRDC/). Only for 2 basins of 24 there were daily observed runoff time series, therefore in our work we used only monthly observations for holding methodological consistency. Runoff data availability is the main limit for developing and validation of our methodology by the reason of the majority of available observations lie in the interval from 1975 to 1985. For the modern studies related to contemporary water resources assessment on vast territories, it is essential to use global gridded data products as the only spatial and temporal continuous source. For this reason, all models were driven by precipitation and temperature data from the ERA-40 reanalysis (1957–2002, 0.5° spatial resolution, http://apps.ecmwf.int/datasets/, Uppala et al., 2005). Potential evapotranspiration is another required forcing variable for all models and it was derived based on Oudin et al. temperature-based equation (Oudin et al., 2005).

3   Methods and software

3.1   Hydrological models

The HBV (Hydrologiska Byråns Vattenbalansavdelning, in Lindström et al., 1997), the GR4J (modele du Genie Rural a 4 parametres Journalier in Perrin et al., 2003), and the SIMHYD (in Chiew et al., 2009) models were used in this study according to its wide implementation for different hydrological applications, flexibility, proven effectiveness for runoff predictions in different geographical conditions, and numerous successful applying for prediction in ungauged basins related studies (Beck et al., 2016; Oudin et al., 2008; Reichl et al., 2009). All listed models have a typical conceptual, bucket type with lumped parameters representation of runoff formation processes at basin scale with daily timestep. GR4J and SIMHYD models have been updated by adding Cema-Neige snow module (Valéry et al., 2014). Models' source code is freely available as a component of LHMP tool (Lumped Hydrological Models Playground, http://github.com/hydrogo/LHMP, Ayzel, 2016). Models' parameters were automatically calibrated by maximizing Nash-Sutcliffe criteria (NSE, Nash and Sutcliffe, 1970) for a whole period of observations using differential evolution algorithm
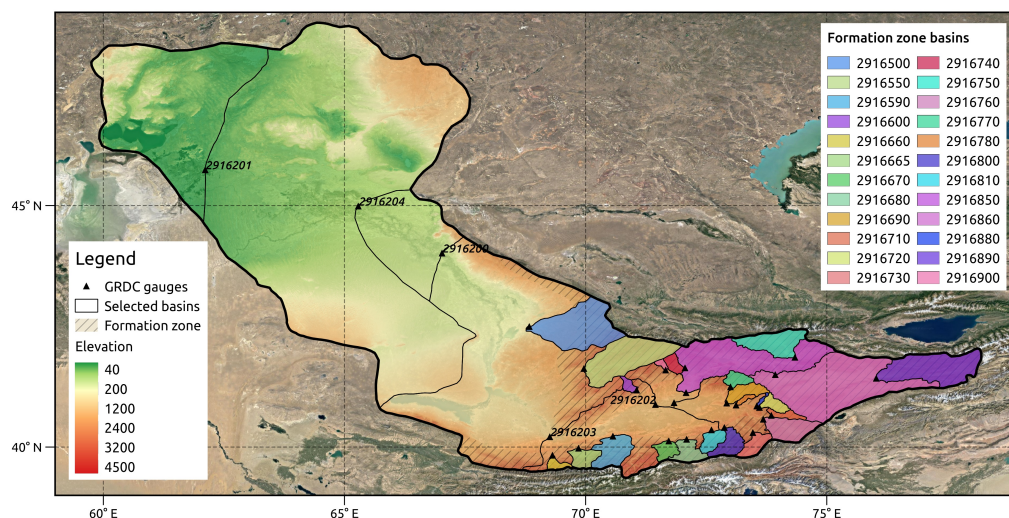
Figure 1. The Small Aral Sea basin and selected river basins.

which finds the global minimum of a multivariate function (http://docs.scipy.org, Storn and Price, 1997).
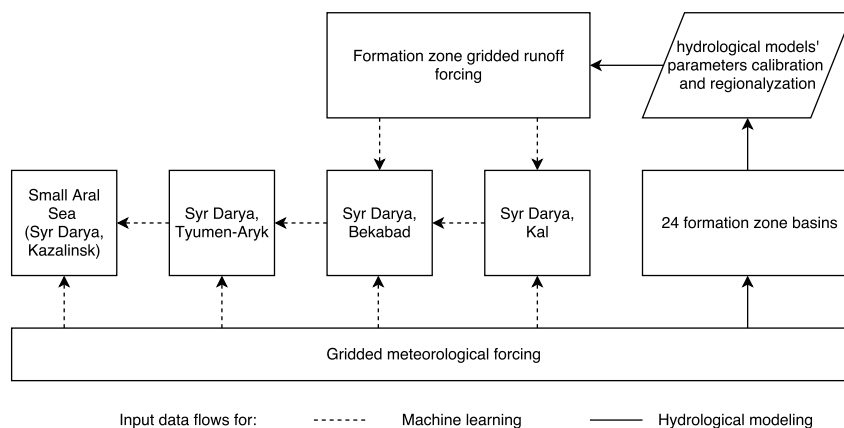
3.2 Machine learning models

For runoff modeling, we have used different machine learning models starting from simple MLP (Multiple Linear Regression) and wide-based decision tree ensembles of ETR (Extra Trees Regression) to the most complicated depth-based tree ensembles of LGB (Light Gradient Boosting machine) and XGB (eXtreme Gradient Boosting machine). MLP and RFR were implemented by using the Scikit-learn package (https://github.com/scikit-learn/scikit-learn, Pedregosa et al., 2011), LGB was taken from the LightGBM package (https://github.com/Microsoft/LightGBM, Zhang et al., 2017), LGB was taken from XGboost package (https://github.com/dmlc/xgboost, Chen and Guestrin, 2016). Machine learning model parameters tuning requires a lot of expertise and experimentation and cannot be resolved automatically because of high computational complexity (Snoek et al., 2012), therefore we had calibrated required parameters manually. For deriving predictions in ensemble manner and approaching realism in the model setting we have used leave-one-out cross-validation technique for machine learning model performance assessment (Ayzel, 2017; Hastie et al., 2001) – as a result, we evaluated model performance on every observational point independently and produced ensemble realization according to the amount of runoff observation we use. This setting provides us most comprehensive evaluation protocol for machine learning models and uncertainties related to models' structures.

3.3 Feature engineering

Feature engineering is an essential part of any routine of machine learning model developing. The general idea of feature engineering is to map already existed features of data to the new representation (dimension). Two (among others) classical implementations of these techniques are extending data with adding some features shifted in time (further referred as LAGS) and shrinking data dimensionality with principal component analysis (PCA) orthogonal transformation algorithm (Hastie et al., 2001). We have tested performances of our machine learning models with default input features, using LAGS and PCA separately and in a coupled setting, and then select the best combination in term of runoff predictions accuracy.

3.4 Workflow representation

The main idea of the presented work was to extract the value using all freely available hydrological information available for the Small Aral Sea basin. On the first stage of our research workflow (Fig. 2) we have calibrated parameters of three hydrological models for 24 rivers run to Ferghana Valley. During the calibration stage, every model had been running at daily temporal resolution then predicted runoff was aggregated at monthly scale for consistency with observational data for loss function calculation. On the second stage, we have implemented common spatial proximity based model parameters regionalization technique (Oudin et al., 2008) for transferring optimal sets of model parameters to meteorological forcing grid cells centroids. On the third stage we have run our models in a grid cell wise mode – for computing runoff in every grid cell in previously delineated formation zone (Fig. 1). As a result we have developed daily gridded multi-model runoff database for the Small Aral Sea

Figure 2. Runoff modeling workflow.

runoff formation zone which serves us as additional input data source for runoff modeling using machine learning models: for the first gauge in our cascade on the Syr Darya river – Kal – we have used both gridded meteorological and formation zone runoff forcing as input, the same for the next gauge in a cascade – Bekabad – but with added mean ensemble modeled runoff realization from Kal. For the remaining two gauges in a cascade (Tyumen Aryk and Kazalinsk), we used only meteorological forcing and mean ensemble modeled runoff realization from overlying gauge in a cascade.
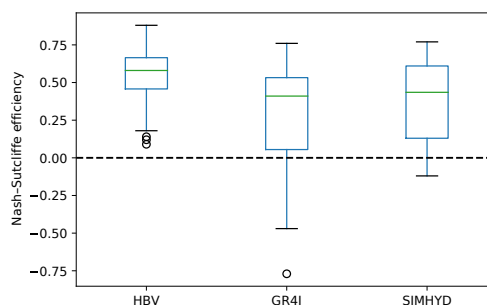
Figure 3. Boxplot of NSE for formation zone basins.

4 Results and discussion

Model calibration results differ from model to model and from different complexity of optimization algorithm. Only one setting with HBV model and the most computationally expensive realization of differential evolution algorithm (number of iteration equals 25) provides positive values of NSE for every single basin (Fig. 3) and we have decided to use only this set-up for further investigations. Only five of selected basins have an NSE less than 0.45 – all of them (GRDC ids: 2916590, 2916660, 2916665, 2916670, 2916810) are located on a north exposition of Alay range. These low efficiencies can be explained by errors in GRDC observational runoff data or errors in basins' metadata (wrong coordinates of outlets, basin areas) which are quite hard to detect and check in the open literature and web sources. Inter-comparison of obtained modeling results with different studies (Lutz et al., 2012b; Pereira-Cardenal et al., 2011; Radchenko et al., 2017; Siegfried et al., 2012) shows high consistency among different approaches for modeling water balance in the upper and mountainous part of the Syr Darya river. Therefore we showed positive value of using freely available data sources for water balance modeling in the study area and tried to transfer this value to the gridded runoff database of the Small Aral Sea (runoff) formation zone using the most robust way of model parameters region-

alization (Ayzel et al., 2017). Using developed gridded runoff dataset for extracting runoff realization for 24 selected basins in a semi-distributed manner shows good consistency with realizations were produced by lumped model setting with optimal parameters.

There is no silver bullet in machine learning field in taking a priori decision about the best data preprocessing routine, the best model, the best validation technique, and the best measure of failure (or success) of the proposed approach. In our research, we have tried to investigate the most widespread solutions for tackling regression problems in machine learning using different state-of-the-art techniques. Results (Table 1) show a good efficiency of different machine learning models to predict monthly runoff alongside a cascade of gauges on the Syr Darya river. The high variance between models' efficiencies from gauge to gauge is explained by the various complexity of water management infrastructure and runoff formation complexes located between those gauges. The worst results for Kal and Kazalinsk gauges and the best efficiency for Tyumen Aryk gauge highly correspond with a complexity of runoff formation/transformation processes we want to map with our input data. The only inflow from the upper Bekabad gauge is enough for robust mapping to a runoff in Tyumen Aryk using simple linear regression
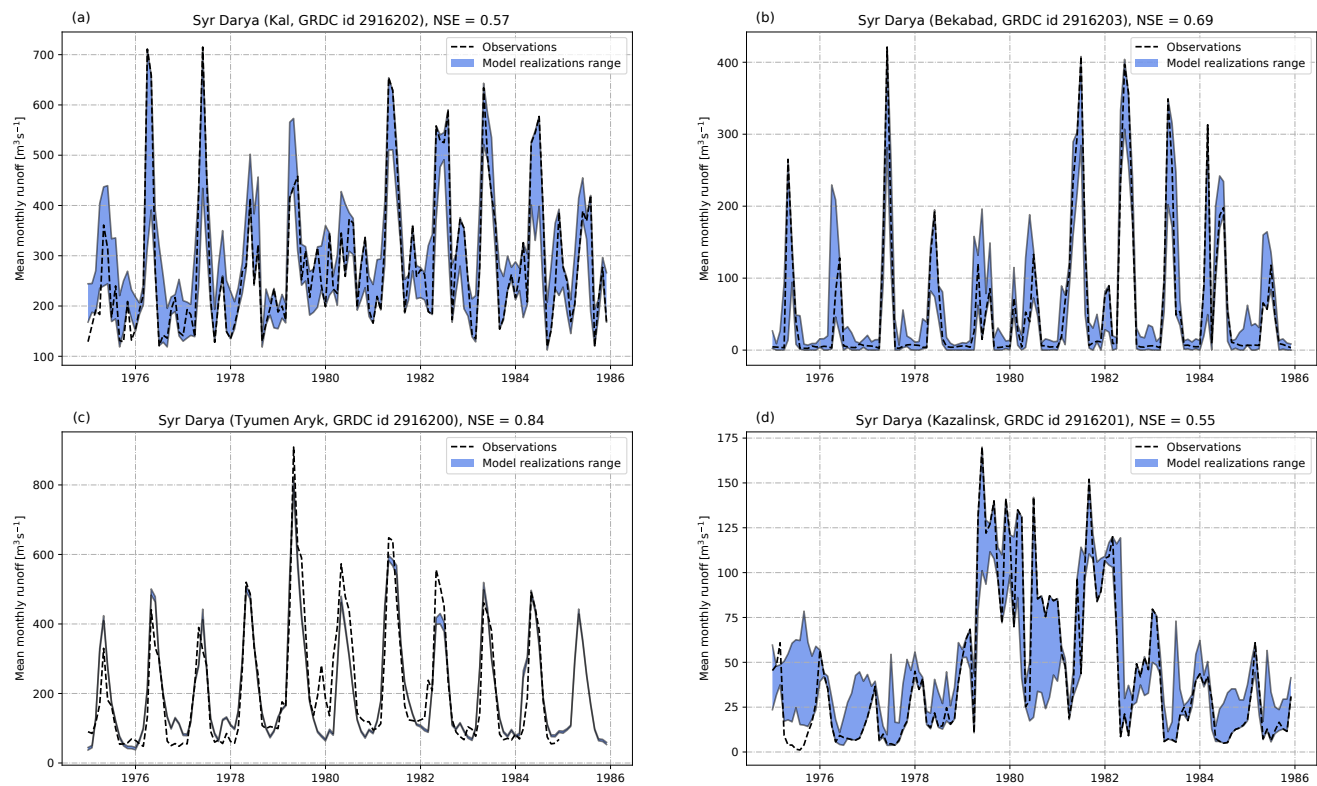
Figure 4. Predictions of machine learning model ensembles.

Table 1. Runoff modeling results.

Gauge, GRDC id	Machine learning models efficiency (NSE)				Feature engineering
	MLR	ETR	LGB	XGB	
Kal (2916202)		0.56	0.56	0.58	LAGS, PCA
Bekabad (2916203)		0.67	0.66	0.73	
Tyumen Aryk (2916200)	0.84				
Kazalinsk (2916201)		0.56		0.54	LAGS

model because of little factors of water redistribution in this local region. But we need more (in terms of quality, quantity, and diversity) data to map more complicated relationships for Kal and Kazalinsk because of high load of water management system on runoff formation/transformation processes. There is no clear pattern in models' efficiency for specific gauges – models rank differently in different settings. Nevertheless, we consider that the best solution in our case is to use both linear MLR and non-linear XGB with different feature engineering techniques for maximizing the spread of possible solutions. Obtained results are in the upper part of NSE range as in Gudmundsson and Seneviratne (2015) who provided monthly runoff predictions for a set of Europe river basins using Random Forest model and Watch Forcing Data (http://www.eu-watch.org/data_availability) as in-

put forcing – this result underlines the crucial value of added gridded runoff information to machine learning model inputs which allows comparable model performance with European basins.

Ensemble runoff predictions produced by machine learning models (Fig. 4) depict significant rate of model-related uncertainties which highly correlates with model complexity. This highlights a statement "the simpler – the better" regarding scientific model robustness issues, but we have to mention that high prediction uncertainties are fair pay for the ability of complex model map input features to relevant output. It is also clear that obtained efficiency correlates with overall complexity of observed system – for Kal gauge station (Fig. 4a) wide range of overlying tributaries and water management rules on them significantly contribute to com-

plexity of processes we have to consider, the same is relevant for the gauge station in Kazalinsk which affected by many, often fuzzy and unclear water management practices (Fig. 4d). This result is also confirmed by the complexity of preprocessing routine – for the simplest cases (Bekabad and Tyumen Aryk gauge stations) we do not need to implement either PCA or LAGS for mapping features for a different dimension.

There is about 100 km from Kazalinsk gauge station to the actual Syr Darya delta, and there are a lot of channels, ponds, and other water management infrastructure units (e.g. the Aklak water regulation station) which can affect total freshwater inflow in the Small Aral Sea basin, but for consistency with previous studies (Lutz et al., 2012a; Raskin et al., 1992) we consider the equality between observed runoff in Kazalinsk and freshwater inflow to the sea. Only a brief look on the observed runoff time series in Kazalinsk (Fig. 4d) gives any researcher a clear representation of high complexity of runoff formation system behavior here – we can only detect simple seasonal pattern with maximum water availability during winter, but it is impossible to generalize remaining runoff amplitude according to natural reasons. Nevertheless, XGB and ETR models utilize this complexity well due to their native algorithmic structures based on simple binarized decision rules which try to mimic decision-making process takes place in many real-life situations. Despite the clear attraction of observations to lower and upper boundaries of our prediction interval which may identify unstable system behavior, there is an obvious correlation between observed and modeled runoff.

Despite the limited observed runoff data availability for this region (mainly for 1975–1985) which was the main constraint in implementing comprehensive routines for proposed methodology validation, obtained machine learning model-based ensemble realizations of freshwater inflow into the Small Aral Sea for the period of 1958–2002 (alongside the forcing data availability) could form the basis for further "Soviet-driven water management" scenario predictions which help us better understand modern shifts in water resources distribution in post-Soviet time.

5 Conclusions

The complex structure of the Small Aral Sea basin water management system coupled with the total absence of data describes its functioning is a challenge for any approach directed to the accurate assessment of the freshwater budget formation and evolution across the basin. Our work shows the possibility to tackle these challenges by coupling hydrological models with the state-of-the-art machine learning techniques. In detail, we have evaluated the significant value of using physically based models for runoff predictions in ungauged upper part of the Syr Darya river for developing gridded runoff database which can be used as an additional fea-

ture for machine learning model in a coupled setting. Results show a positive skill and a high flexibility of the proposed methodology, and in our perspective, it can be used widely as a baseline approach for water balance research studies in arid, ungauged areas, with complex water management system and strong water-related data scarcity.

We understand that an equality between freshwater inflow into the Small Aral Sea and observed runoff in Kazalinsk is quite a rough assumption, and in the further studies, we will try to assess real inflow by coupling simple seawater balance model to our existing modeling system.

The code and data we have developed are totally open and freely accessible. We hope that this supports reproducibility of our research and provides easy access to the community to test, criticize, or apply our findings.

References

Apel, H., Abdykerimova, Z., Agalhanova, M., Baimaganbetov, A., Gavrilenko, N., Gerlitz, L., Kalashnikova, O., Unger-Shayesteh, K., Vorogushyn, S., and Gafurov, A.: Statistical forecast of seasonal discharge in Central Asia for water resources management: development of a generic linear modelling tool for operational use, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-340, in review, 2017.

Ayzel, G.: LHMP: First major release, https://doi.org/10.5281/zenodo.59680, 2016.

Ayzel, G.: Use of machine learning techniques for modeling of snow depth, Ice and Snow, 34–44, https://doi.org/10.15356/2076-6734-2017-1-34-44, 2017.

Ayzel, G. and Izhitskiy, A.: Data, code, and results for the paper "Coupling physically based and data-driven models for assessing freshwater inflow into the Small Aral Sea", https://doi.org/10.5281/zenodo.1161906, 2017.

Ayzel, G. V., Gusev, E. M., and Nasonova, O. N.: River runoff evaluation for ungauged watersheds by SWAP model. 2. Application of methods of physiographic similarity and spatial geostatistics, Water Resour., 44, 547–558, https://doi.org/10.1134/S0097807817040029, 2017.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52, 3599–3622, https://doi.org/10.1002/2015WR018247, 2016.

Belyaev, A. V.: Water Balance and Water Resources of the Aral Sea Basin and Its Man-Induced Changes, GeoJournal, 35, 17–21, 1995.

Chen, T. and Guestrin, C.: XGBoost: Reliable Large-scale Tree Boosting System, available at: https://arxiv.org/abs/1603.02754 (last access: 13 February 2018), 2016.

Chen, Y., Li, W., Fang, G., and Li, Z.: Review article: Hydrological modeling in glacierized catchments of central Asia – status and challenges, Hydrol. Earth Syst. Sci., 21, 669–684, https://doi.org/10.5194/hess-21-669-2017, 2017.

Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C., and Viney, N. R.: Estimating climate change impact on runoff across southeast Australia: Method, results, and implications of the modeling method, Water Resour. Res., 45, W10414, https://doi.org/10.1029/2008WR007338, 2009.

Gan, R., Luo, Y., Zuo, Q., and Sun, L.: Effects of projected climate change on the glacier and runoff generation in the Naryn River Basin, Central Asia, J. Hydrol., 523, 240–251, https://doi.org/10.1016/j.jhydrol.2015.01.057, 2015.

Gudmundsson, L. and Seneviratne, S. I.: Towards observation-based gridded runoff estimates for Europe, Hydrol. Earth Syst. Sci., 19, 2859–2879, https://doi.org/10.5194/hess-19-2859-2015, 2015.

Hagg, W., Braun, L. N., Weber, M., and Becht, M.: Runoff modelling in glacierized Central Asian catchments for present-day and future climate, Nordic Hydrology, 37, 93–105, https://doi.org/10.2166/nh.2006.001, 2006.

Hagg, W., Braun, L. N., Kuhn, M., and Nesgaard, T. I.: Modelling of hydrological response to climate change in glacierized Central Asian catchments, J. Hydrol., 332, 40–53, https://doi.org/10.1016/j.jhydrol.2006.06.021, 2007.

Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, Springer Series in Statistics, Springer New York Inc., https://doi.org/10.1007/978-0-387-84858-7, 2001.

Immerzeel, W. W. and Bierkens, M. F. P.: Asia's water balance, Nat. Geosci., 5, 841–842, https://doi.org/10.1038/ngeo1643, 2012.

Izhitskiy, A. S., Zavialov, P. O., Sapozhnikov, P. V., Kirillin, G. B., Grossart, H. P., Kalinina, O. Y., Zalota, A. K., Goncharenko, I. V., and Kurbaniyazov, A. K.: Present state of the Aral Sea: diverging physical and biological characteristics of the residual basins, Scientific Reports, 6, 23906, https://doi.org/10.1038/srep23906, 2016.

Jarsjö, J., Asokan, S. M., Prieto, C., Bring, A., and Destouni, G.: Hydrological responses to climate change conditioned by historic alterations of land-use and water-use, Hydrol. Earth Syst. Sci., 16, 1335–1347, https://doi.org/10.5194/hess-16-1335-2012, 2012.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, https://doi.org/10.1016/S0022-1694(97)00041-3, 1997.

López, O., Houborg, R., and McCabe, M. F.: Evaluating the hydrological consistency of evaporation products using satellite-based gravity and rainfall data, Hydrol. Earth Syst. Sci., 21, 323–343, https://doi.org/10.5194/hess-21-323-2017, 2017.

Lutz, A., Droogers, P., and Immerzeel, W.: Climate Change Impact and Adaptation on the Water Resources in the Amu Darya and Syr Darya River Basins. FutureWater Report 110, Tech. rep., FutureWater, available at: http://www.futurewater.eu/wp-content/uploads/2013/01/CC_Downstream_Report_V51.pdf (last access: 13 February 2018), 2012a.

Lutz, A., Immerzeel, W., and Droogers, P.: Climate Change Impacts on the Upstream Water Resources of the Amu and Syr Darya River Basins. FutureWater Report 107, Tech. rep., FutureWater, available at: http://www.futurewater.eu/wp-content/uploads/2012/03/Upstream_Report_FW_web.pdf (last access: 13 February 2018), 2012b.

Micklin, P.: The Aral Sea Disaster, Annu. Rev. Earth Planet. Sc., 35, 47–72, https://doi.org/10.1146/annurev.earth.35.031306.140120, 2007.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling, J. Hydrol., 303, 290–306, https://doi.org/10.1016/j.jhydrol.2004.08.026, 2005.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resour. Res., 44, W03413, https://doi.org/10.1029/2007WR006240, 2008.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Pereira-Cardenal, S. J., Riegels, N. D., Berry, P. A. M., Smith, R. G., Yakovlev, A., Siegfried, T. U., and Bauer-Gottwein, P.: Real-time remote sensing driven river basin modeling using radar altimetry, Hydrol. Earth Syst. Sci., 15, 241–254, https://doi.org/10.5194/hess-15-241-2011, 2011.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, 2003.

Radchenko, I., Dernedde, Y., Mannig, B., Frede, H.-G., and Breuer, L.: Climate change impacts on runoff in the Ferghana Valley (Central Asia), Water Resour., 44, 707–730, https://doi.org/10.1134/S0097807817050098, 2017.

Raskin, P., Hansen, E., Zhu, Z., and Stavisky, D.: Simulation of Water Supply and Demand in the Aral Sea Region, Water Int., 17, 55–67, https://doi.org/10.1080/02508069208686127, 1992.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S.: Optimization of a similarity measure for estimating ungauged streamflow, Water Resour. Res., 45, W10423, https://doi.org/10.1029/2008WR007248, 2009.

Shi, W., Wang, M., and Guo, W.: Long-term hydrological changes of the Aral Sea observed by satellites, J. Geophys. Res.-Oceans, 119, 3313–3326, https://doi.org/10.1002/2014JC009988, 2014.

Shibuo, Y., Jarsjö, J., and Destouni, G.: Hydrological responses to climate change and irrigation in the Aral Sea drainage basin, Geophys. Res. Lett., 34, l21406, https://doi.org/10.1029/2007GL031465, 2007.

Siegfried, T., Bernauer, T., Guiennet, R., Sellars, S., Robertson, A. W., Mankin, J., Bauer-Gottwein, P., and Yakovlev, A.: Will climate change exacerbate water stress in Central Asia?, Climatic Change, 112, 881–899, https://doi.org/10.1007/s10584-011-0253-z, 2012.

Snoek, J., Larochelle, H., and Adams, R. P.: Practical Bayesian Optimization of Machine Learning Algorithms, in: Advances in Neural Information Processing Systems 25, edited by: Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., 2951–2959, Curran Associates, Inc., 2012.

Sorg, A., Mosello, B., Shalpykova, G., Allan, A., Hill Clarvis, M., and Stoffel, M.: Coping with changing water resources: The case of the Syr Darya river basin in Central Asia, Environ. Sci. Policy, 43, 68–77, https://doi.org/10.1016/j.envsci.2013.11.003, 2014.

Storn, R. and Price, K.: Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, J. Global Optim., 11, 341–359, https://doi.org/10.1023/A:1008202821328, 1997.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

Valéry, A., Andréassian, V., and Perrin, C.: "As simple as possible but not simpler": What is useful in a temperature-based snow-accounting routine? Part 1 – Comparison of six snow accounting routines on 380 catchments, J. Hydrol., 517, 1166–1175, https://doi.org/10.1016/j.jhydrol.2014.04.059, 2014.

Zavialov, P. O., Kostianoy, A. G., Emelianov, S. V., Ni, A. A., Ishniyazov, D., Khan, V. M., and Kudyshkin, T. V.: Hydrographic survey in the dying Aral Sea, Geophys. Res. Lett., 30, 1659, https://doi.org/10.1029/2003GL017427, 1659, 2003.

hang, H., Si, S., and Hsieh, C.-J.: GPU-acceleration for Large-scale Tree Boosting, available at: http://arxiv.org/abs/1706.08359 (last access: 13 February 2018), 2017.

Zmijewski, K. and Becker, R.: Estimating the Effects of Anthropogenic Modification on Water Balance in the Aral Sea Watershed Using GRACE: 2003–12, Earth Interact., 18, 1–16, https://doi.org/10.1175/2013EI000537.1, 2014.