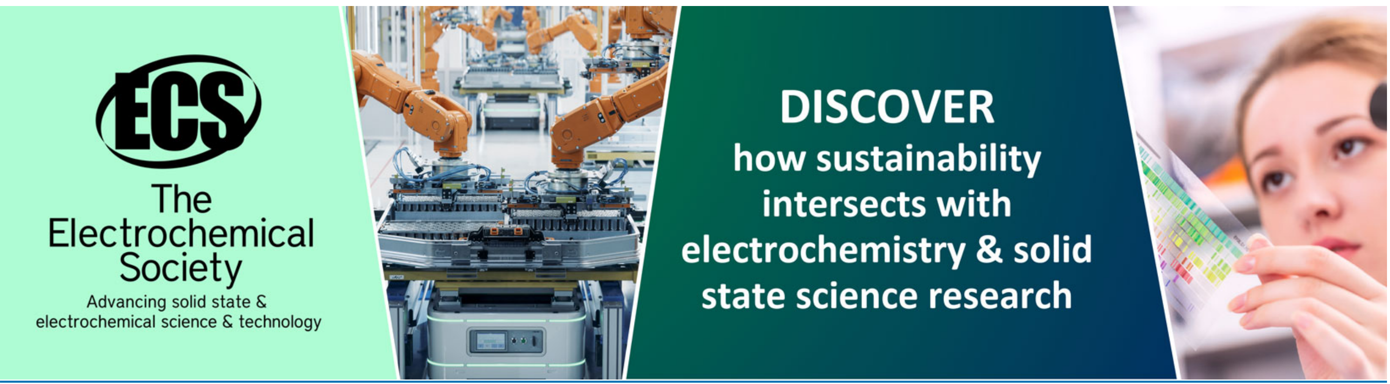
IOP Conference Series: Earth and Environmental Science

PAPER • **OPEN ACCESS**

Development of multi-model ensemble climate estimates using open-source environment

To cite this article: V V Klimenko and E V Fedotova 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **211** 012084



Development of multi-model ensemble climate estimates using open-source environment

V V Klimenko[1, 2] and E V Fedotova[1, 2]

[1]Moscow Power Engineering Institute, ul. Krasnokazarmennaya 14, Moscow, 111250 Russia, www.mpei.ru

[2]A.M. Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Pyzhevskiy per., 3, Moscow, 119017 Russia, www.ifaran.ru

Abstract. We demonstrate an application of an R open source environment to climatological problems, namely, to the construction of multi-model ensemble estimates. The R ecosystem is used to develop a computation code which automates an ensembling procedure. We apply the code to the CMIP5 runoff data in the Russian territory for a preliminary analysis of the long-term natural variability. It has been found that a multi-model projection demonstrates a non-monotonic behaviour which contradicts the classic "wet getting wetter" pattern, but perfectly corresponds to the long-term natural variability. A detailed analysis of reproducibility of the runoff temporary structure seems to be unavoidable for successful forecasting of the long-term runoff trends.

1. Introduction

The open source paradigm has been one of the main trends of software development during the last ten years and has considerably changed the development standards. The open source R development environment [1] has become the leading technology for the data science purposes in both production and research.

We have taken advantage of R to create a computation tool for the solution of a routine climatology problem, namely, the construction of multi-model climate estimates. The ensembling procedure is a common approach to decrease simulation uncertainties when using climate simulation results.

A discussion of the datasets used and implementation approaches are presented in the next section. Application of the developed calculation tool to a real practical problem is demonstrated in the last part of the paper.

2. Implementation of the data processing approach

The simulation results of the Coupled Model Intercomparison Project Phase 5 (CMIP5) were used as an input for our calculations [2]. Nowadays the CMIP5 dataset is the most comprehensive source of initial data to construct climate projections on a long-term scale. The common workflow includes utilisation of the CMIP5 data to build up an ensemble estimation with further application of statistic or dynamic downscaling methods.

The CMIP5 data are represented in the network common data format (netCDF). R provides a number of tools to work with netCDF files. Now the main netCDF-processing includes **raster** [3], **RCMIP5** [4], and **netcdf4** [5] packages.

The **raster** package is integrated into the eco-system of R geospatial analysis tools and fits perfectly to read, manipulate, and visualize gridded geographical datasets. However, the **raster** tends to hide from the user steps of the datasets processing and meta-information of the datasets. This feature increases the probability of mistakes during data processing and makes the raster a nonoptimal solution for our tasks.

The **RCMIP5** package was created for fast and easy processing of the CMIP5 data. Unfortunately, **RCMIP5** is not intended to provide much flexibility to the user, especially when working with monthly-aggregated data.

The **ncdf4** package is a universal tool to process data of the netCDF format giving to the user a high level of flexibility and control. We took **ncdf4** as the main tool and built it up with a programming code which allows both efficient and accurate processing of the CMIP5 files. One of our main ideas was to apply linear interpolation on the real calendar time for the monthly-aggregated data.

3. Application example

We have applied the developed approach to one of the challenging problems of modern climatology, namely, to long-term projection of the runoff. Hydropower is nowadays a leading renewable power branch in Russia accounting for about 18% of the total national electricity production. A combination of large-scale power plants with moderate head-differences determines high complexity of water use control. Competition between energy and non-energy users may lead to serious problems.

The climate change has contributed to a significant shift of the runoff regimes across the country during the last twenty years. Particularly, the seasonal runoff distribution is getting more uniform mainly due to an increase of the winter runoff. Changes of the annual runoff are still within the range of natural variability. However, the increasing trend of the air temperature combined with the changes of the precipitation patterns may cause non-stationarity of the annual runoff time-series as well.

The observed and probable future hydrological changes and tightening of the environmental requirements in Russia bring the question of applicability of the current water use regulations. Therefore, long-term projections of the runoff could be of highest practical use. Unfortunately, the results of runoff-projections for the next 50-80 years are quite uncertain [7]. Apart from the complex nature of the hydrological processes themselves, lack of available observational data contributes greatly to this uncertainty.

One of the recent findings of hydrological research in Russia was an experimental evidence of long-term periodicities in the runoff-records. The runoff was proved to demonstrate approximately thirty-year phases of increase and decrease both in the European and Siberian parts of Russia [8]. However, to the best of our knowledge, all published works completely ignore this fact when making long-term projections of the runoff in Russia. This gap is connected, among other issues, with the complexity of computational approaches that have been applied.

Our automated approach allows us to facilitate the computational workflow dramatically and makes possible a more detailed analysis.

First of all, we made our own quick check confirmed existence of long-term periodicities in the runoff records. Our explanatory analysis discovered some linking of these periodicities with the global atmospheric processes (Figure 1), of which the Atlantic Multidecadal Oscillation (AMO) seems to have the most pronounced impact. The phases of high and low runoff across the country correspond, most likely, to the opposite phases of the AMO.

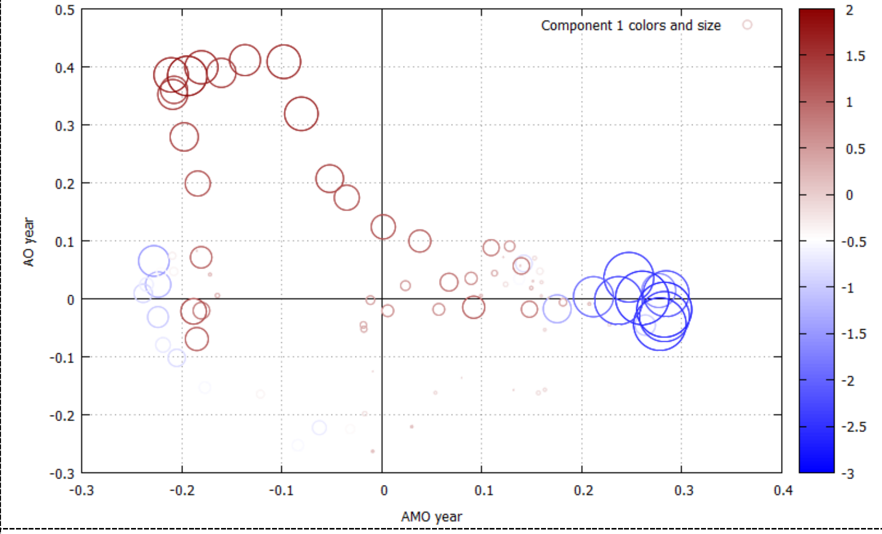


Figure 1. Interconnection of the first principal component of the Volga-Kama runoff with the AO and AMO indices.

The next step of our research was validation of the CMIP5 simulation results. We used as inputs the gridded runoff dataset [9] and the observation records of the Global River Data Centre [10]. A satisfactory correspondence between the observation and the simulated data was found for the values smoothed on multi-year time scales (Figure 2). Different models demonstrate different skills depending on the choice of a validation data set, making multi-model ensemble averaging a good strategy. That is why we decided not to exclude any models for our preliminary countrywide analysis.

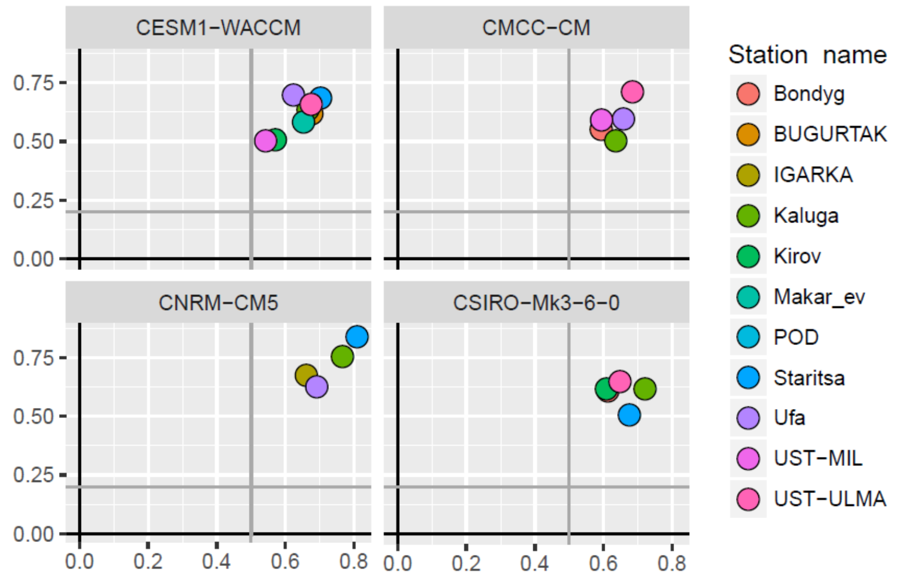
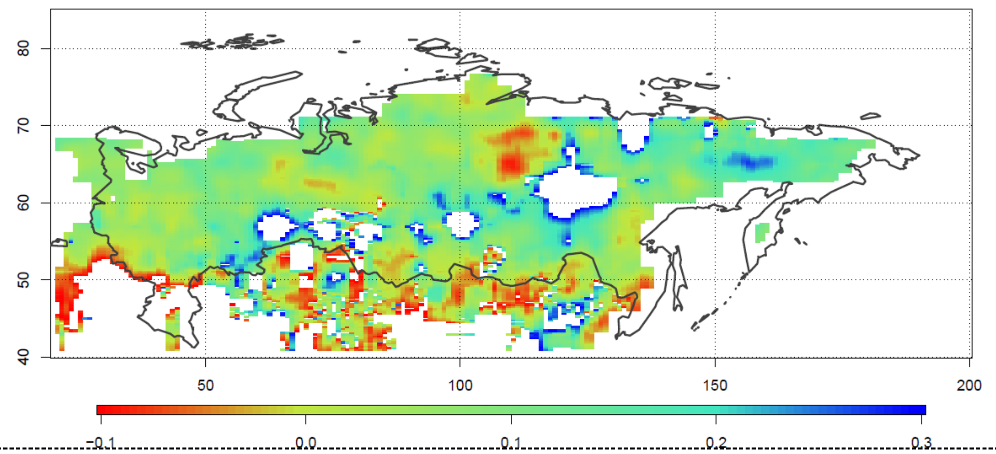


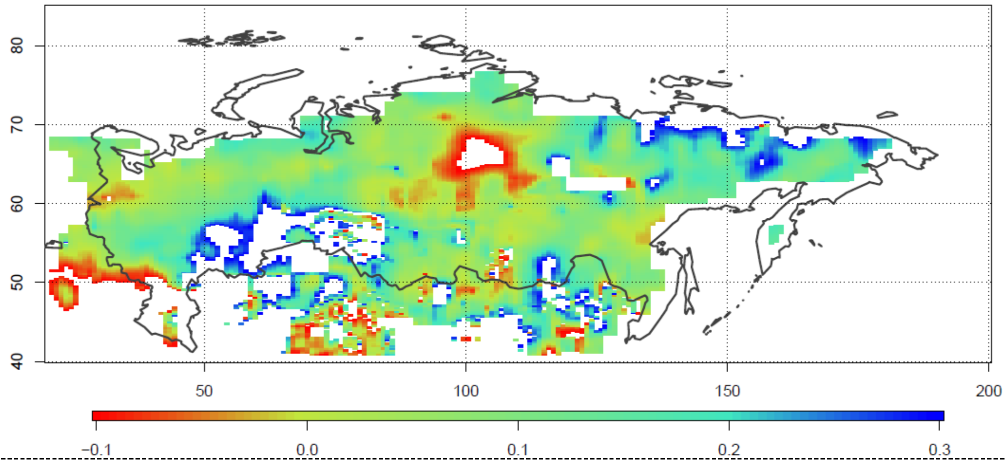
Figure 2. Validation diagrams: correlation of simulated runoff results with observation data, both smoothed with a ten-year moving average. The mean correlation in ten-degree area around a station is computed; only correlation values more than 0.5 are shown.

Finally, we have constructed some multi-model ensembles for different time horizons. Two of them are shown in Figure 3. The first 2045 to 2054 period will likely correspond to the anticipated negative AMO phase; the second one, from 2071 to 2080, to the positive one.

Generally, the simulated changes correspond to the classic “wet-getting-wetter” pattern. But it is interesting that the obtained runoff distributions have a non-monotonic change of the runoff field in some regions. Particularly, the north-western part of European Russia gets wetter in the first period and noticeably drier during the second one. The changes in the area around northern Urals demonstrate an antiphase pattern. Such spatial peculiarities, namely, different wet and dry phases of the Kama and upper Volga basins are well known to the hydrology engineers. Apart of that, such natural variability on the thirty-year period perfectly corresponds to the both synthesis of the observations [8] and up-to-date theoretical results [11].



a



b

Figure 3. Ensemble estimates of the percent change of the annual runoff for rcp 4.5 scenario as compared with 2007-2016 reference period: (a) for 2045-2054, (b) for 2071-2080. CMIP5 data [12] were processed by the authors.

The results of our study clearly demonstrate that natural variability should be necessarily taken into account when developing long-term runoff-projections. Detailed analysis of the reproducibility of the runoff temporary structure seems to be unavoidable for successful forecasting of long-term trends.

Acknowledgments
This study was supported by the Russian Science Foundation (project No 16-17-00114) in part of methodology development and under Presidential Young Scientists grant MK-1494.2017.8 for E.

Fedotova (Kasilova) in part of runoff estimations. The authors are grateful to the modelling groups of the CMIP5 project and the World Data Centre for Climate in Hamburg for providing access to the CMIP5 simulation data.

The authors also thank Prof. A.Yu. Aleksandrovskii for insightful discussions on water management practices.

References

[1] R Core Team 2013 R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (Vienna, Austria) http://www.R-project.org/

[2] Taylor K E, Stouffer R J and Meehl G A 2012 An Overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc*. **93** 485–98

[3] Robert J. Hijmans 2017 Raster: Geographic Data Analysis and Modeling. R package version 2.6-7. *CRAN Repository* (https://CRAN.R-project.org/package=raster

[4] Todd-Brown K and Bond-Lamberty B 2016 Package ‘RCMIP5’ Tools for Manipulating and Summarizing CMIP5 Data *CRAN Repository https://CRAN.R-project.org/package=RCMIP5*

[5] Pierce D 2017 ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R package version 1.2.0 *CRAN Repository https://CRAN.R-project.org/package=ncdf4*

[6] *The Second Evaluation Report on Climate Change and Its Consequences on the Territory of the Russian Federation* 2014 Ed. by V. M. Kattsov and S. M. Semenov (Rosgidromet, Moscow) (in Russian).

[7] *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* 2014 ed Core Writing Team, R. K. Pachauri and L. A. Meyer (Intergovernmental Panel on Climate Change, Geneva, Switzerland)

[8] Popova V V and Georgiadi A G 2017 Spectral estimations of the interconnections of the Volga runoff variability with the Northern Atlantic Oscillation in 1882-2007 *Izvestiya Rossijskoj Akademii Nauk. Seriya geograficheskaya* **2** 47-59 (in Russian)

[9] Gudmundsson L and Seneviratne S 2016 Observation-based gridded runoff estimates for Europe (E-RUN version 1.1) *Earth Syst. Sci. Data*. **8** 279-295

[10] GRDC Timeseries Data 2018 *Global Runoff Data Centre* (Koblenz, Federal Institute of Hydrology (BfG))

[11] Semenov V A and Cherenkova E A 2018 Evaluation of the Atlantic Multidecadal Oscillation Impact on Large-Scale Atmospheric Circulation in the Atlantic Region in Summer *Doklady Earth Sciences* 2 **478** 263–267

[12] Knutti R 2014 IPCC Working Group I AR5 snapshot: The rcp45 experiment *World Data Center for Climate* doi:10.1594/WDCC/ETHr4