Bias-corrected precipitation data for South Siberia

A A Ryazanova¹, N N Voropay¹,² and E A Dyukarev¹,³

¹ Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk, Russia
² Sochava Institute of Geography SB RAS, Irkutsk, Russia
³ Yugra State University, Khanty-Mansiysk, Russia

E-mail: r.ann.1194@gmail.com, voropay_nn@mail.ru, dekot@mail.ru

Abstract. Assessing changes in extreme hydroclimatic phenomena requires accurate precipitation estimates with high temporal and fine spatial resolutions. Significant differences were obtained between reanalysis data and observations at weather stations. Bias correction method applied to ERA5 total precipitation product with spatial resolution 0.25×0.25 ° is reported in the present paper to improve the reproducibility of the local weather conditions. Hourly ERA5 precipitations were aggregated and monthly precipitation sums were processed. Monthly precipitations for 133 weather stations in Siberia for May-September from 1979 to 2017 were used as reference data. Comparative analysis has shown that long-term average monthly precipitations from ERA5 reanalysis are higher than precipitation estimations from the observation data. The number of weather stations where the difference between two data sets exceed 60% vary from 3 in July to 37 in May. To eliminate significant differences between reanalysis and observation data, we have developed a specific procedure for correcting reanalysis data. As a result of the adjustment, the deviations of the reanalysis data from the observation data has decreased. The difference after correction is lower than 60 % at all weather stations, and from 94 (July) to 118 (August) weather stations has the difference lower than 30%.

1. Introduction
Atmospheric precipitation is one of the most important climate characteristics. They are the main source of moisture for the land surface. In climatology, precipitation is represented by a large number of climatic indicators. In contrast to other meteorological quantities, the daily variation of atmospheric precipitation is not considered. Decadal, monthly, seasonal, and annual amounts are calculated, their Statistical characteristics of precipitation time series (mean, coefficients of variations and asymmetries, daily maximum for different periods, the number of days by gradation, the duration of continuous rains, etc.) are determined to characterize the long-term dynamics.

Many researchers show that the intensity of atmospheric precipitation increases with climate warming observed in recent decades [1-3]. Despite the increase in the amount of atmospheric precipitation over land in mid-latitudes [4-6], its changes in the territory of the Russian Federation are generally not statistically significant [7, 8]. At the same time, there is a significant increase in the seasonal intensity and frequency of extreme precipitation in many regions of the country [9, 10]. On the other hand, in recent decades, there has been an increase in the frequency and intensity of droughts in summer in the European territory of Russia [11, 12] and in Siberia [13-16]. In connection with the foregoing, current representative data on atmospheric precipitation are the basis for qualitative
analysis and forecasting of extreme hydroclimatic events, which include droughts and periods of excessive moisture.

First of all, climatologists use observation data from weather stations to analyze the hydroclimatic regime. Unfortunately, on the territory of Siberia, the network of weather stations of Roshydromet is very spaced. Often, especially in the northern regions, the distance between stations exceed 100 km. An additional problem is the complex relief of the territory. At the same time, in mountainous areas, weather stations are located mainly in river valleys.

Assessing changes in characteristics of extreme events requires accurate precipitation estimates with high temporal resolution and fine spatial scale. In this context, long-term global atmospheric reanalysis datasets and climate model calculations are an important source of information, mainly due to their global coverage and long-term data series exceeding 30-35 years.

Recent years many researchers have used various reanalyses data for accessing climatic changes. However, the coarse spatial resolution (0.25-0.5 degrees) introduces a significant error in the assessment of local risks of droughts and floods, carried out for individual river basins. Good quality (convergence of observation and reanalysis data) has been confirmed for air temperature, atmospheric pressure and some other meteorological characteristics [17, 18], the observation data of which are assimilated in the process of making forecasts and reanalysis. Unfortunately, this statement does not apply to atmospheric precipitation. At first glance, reanalyses reproduce the precipitation field with good quality. Precipitation products from reanalysis show a contrast between upland and plain areas, leeeward and windward slopes, water bodies and land. However, when compared with the observations at weather stations, significant differences are discovered [19-23]. Therefore, downscaling and bias correction of the results of calculations of global climate models are often applied to improve the reproducibility of local weather conditions [24-27]. There are two basic approaches to scaling raw grid data. With dynamic scaling, when a regional climate model is built into the coarse grid of the global climate model, which reproduces the physics of the atmosphere at a higher resolution within a limited research area. Statistical scaling establishes statistical relationships between large-scale weather characteristics and observed local-scale weather [28].

There are many statistical approaches to scaling large-scale data [29-31]. In its simplest form, the idea of statistical scaling involves finding a relation between a weather characteristic of a large (or larger) scale and the expected value of a characteristic of a local scale [28]. In addition, reanalysis and forecast data from global climate models are scaled using parametric [32] or non-parametric statistical schemes based on empirically derived quantile regression relationships. The scaled data should match very well with monitoring data at weather stations and the local monitoring network [33] and, therefore, can be used to assess the characteristics of extreme events in the area.

2. Methods
In previous studies, we compared the Era Interim reanalysis data with observation data at weather stations and corrected the monthly sums of atmospheric precipitation at the reanalysis grid nodes for the territory of Southern Siberia [23, 34]. In this work, to study the hydrothermal conditions, it was decided to use the data of the Era5 reanalysis [35], since this reanalysis was released relatively recently and replaced the Era-Interim reanalysis [36]. In January 2019, access to the archive with data for the period from 1979 to the present was opened, new data is posted regularly and available for download through the Copernicus system. The spatial resolution of the Era5 reanalysis is 0.25×0.25°, which is approximately 27.7 km in latitude and 16.5 km in longitude for the territory of Siberia. The time resolution of the data is one hour, and larger time periods (daily, monthly) are also available for download. Since in our study we are interested in what changes occur during the growing season in Siberia, it was decided to use the archive with monthly data for May-September from 1979 to 2017.

The first step in our work was to compare the data of this reanalysis with the data of meteorological observations. Since the used reanalysis is one of the most modern products of retrospective analysis with a high spatial resolution, a comparison with the measured meteorological characteristics at weather stations should reliably show how accurately this reanalysis reproduces the meteorological
characteristics in different regions and landscapes of Siberia.

As verification data of observations, we used an archive with monthly homogeneous series of data of the required characteristics for 133 stations in Siberia for the growing season (May–September) from 1979 to 2017. Weather station observation data are available for free download through the VNIIGMI-MCD portal (www.meteo.ru).

To eliminate significant differences between reanalysis data and observation data, we have developed a special procedure for correcting reanalysis data, which is described below:

- Reanalysis data $R_{grd}$ (at $0.25 \times 0.25^\circ$ grid) were interpolated using the method of simple bilinear interpolation to the points of location of weather stations $s$ for each time interval (month) $m$ and year $t$.

- Comparison (correlation coefficient, bias value, graphical presentation of data) of interpolated reanalysis data $R_{s,m}^t$ and observation data $O_{s,m}^t$ at each station $s$ and for each month $m$ is performed, where $s = [1 ... S]$, $m = [1 ... M], t = [1 ... T]$ where $S$ is the number of stations (in our case 133 stations), $M$ is the number of time intervals (in our case, 5 months from May to September), $T$ is the number of years for which the study is conducted (in our case 39 years, 1979-2017) (table 1, 2).

- Relative absolute deviation for each station $s$ and month $m$ were calculated:

$$D_{s,m} = \frac{1}{T} \sum_{t=1}^{T} \left( R_{s,m}^t - O_{s,m}^t \right) / O_{s,m}^t \cdot 100\%$$

- Then the calculated characteristics are analyzed, and if significant deviations are observed, then the procedure for correcting the reanalysis data using the data of weather stations is performed.

**Table 1.** Number of weather stations within Correlation Coefficient ranges of precipitation (observations / reanalysis).

| Range of Correlation Coefficient | V  | VI | VII | VIII | IX |
|---------------------------------|----|----|-----|------|----|
| 0.28-0.40                      | -  | 1  | 3   | -    | -  |
| 0.41-0.53                      | 5  | 7  | 7   | 3    | 2  |
| 0.54-0.67                      | 14 | 19 | 30  | 35   | 7  |
| 0.68-0.80                      | 33 | 64 | 60  | 70   | 36 |
| 0.81-0.96                      | 81 | 42 | 33  | 25   | 88 |

**Table 2.** Number of weather stations within relative deviation (%) of precipitation (reanalysis - measurement data) before correction.

| Relative deviation (%) | V  | VI | VII | VIII | IX |
|------------------------|----|----|-----|------|----|
| 0-30                   | 25 | 39 | 59  | 62   | 51 |
| 30-60                  | 71 | 86 | 71  | 67   | 65 |
| 60-90                  | 19 | 5  | 1   | 2    | 9  |
| 90-200                 | 16 | 2  | 1   | 1    | 6  |
| >200                   | 2  | 1  | 1   | 1    | 2  |

According to the calculated deviation, the stations were divided into 5 groups: 1 group – stations where the deviation of the reanalysis precipitation data does not exceed 30%; 2 group – stations where the deviation is from 30 to 60%; 3 group – stations where the deviation is from 60 to 90%; 4 group – stations where the deviation is from 90 to 200%; 5 group – stations where the deviation is more than 200%.
The next step was the division of the original sample into the training and control samples. Usually, 20% of the data from the original sample is taken for the control sample respectively, 80% of the data make up the training sample. This 20% was randomly selected from each group of stations, and the total number of stations in the control sample was 26 stations. After selection of the control sample, the training part of the sample was used to correct the reanalysis data.

The algorithm of the correction procedure was the following:

- For each weather station and for each time interval, a simple linear regression without an intercept equation were written $R_{s,m} = a_{s,m} O_{s,m}$, where $s = [1 \ldots S]$, $m = [1 \ldots M]$, $S$ – the number of stations (in our case, 133 stations from the training part of the sample), $M$ – the number of time intervals (in our case, 5 months, May-September), $R_{s,m}$ – vector with interpolated reanalysis data at the $s$-th point of the weather station location for the $m$-th time interval, $O_{s,m}$ – vector with data from the $s$-th weather stations for the $m$-th time interval.
- Using the least squares method, the coefficient $a_{s,m}$ was estimated.
- The obtained non-uniform field of coefficients $a_{s,m}$ was interpolated to a uniform grid (in our case 0.25x0.25°) using the simple Kriging method [37, 38].
- The obtained uniform field (in our case at grid 0.25°x0.25°) of $a_{s,m}$ coefficients was multiplied by the initial field of reanalysis values $R_{grd}$ for each month $m$ of $t$-th year.

3. Results
As a result of this procedure, the corrected reanalysis field $R_{corr}$ of the investigated meteorological characteristic is obtained. Since this procedure is universal and can be applied for any meteorological characteristic and dataset, in the future it is planned to integrate it into the CLIMAT web GIS [http://climate.scert.ru/] as an additional software module. In this work, the entire software implementation was performed in the Python 3 programming language.

Then, to check the quality of the correction, the comparison procedure was again applied to the obtained field of the corrected reanalysis (table 3). Interpolation was carried out at all points of the stations, including the control sample.

### Table 3. Number of weather stations within relative deviation (%) of precipitation (reanalysis data – measurement data) after correction.

| Relative deviation (%) | Month |
|------------------------|-------|
|                        | V     | VI    | VII   | VIII  | IX    |
| 0-30                   | 106   | 102   | 94    | 106   | 118   |
| 30-60                  | 27    | 31    | 39    | 27    | 15    |
| 60-90                  | -     | -     | -     | -     | -     |
| 90-200                 | -     | -     | -     | -     | -     |
| >200                   | -     | -     | -     | -     | -     |

Then, the resulting deviations of the control sample were compared with the deviations obtained before correction. As a result of the correction procedure, the number of stations in groups 1 and 2 has increased. Groups 3-5, selected from comparison initial reanalysis product, do not include any station. Relative deviation at any studied station of bias corrected precipitation form observation data does not exceed 60%. Essential amount of stations was attributed to group 1 with relative deviation less than 30%.

4. Conclusion
Precipitation is one of the most important characteristics of the climate. They are the main source of moisture for the earth's surface. In climatology, precipitation estimates are represented by a large number of climatic indicators (amount of precipitation, intensity, frequency, etc.). Long-term global
atmospheric reanalysis datasets and climate model calculations are an important source of information, mainly due to their global coverage and large data range over 30-35 years. Comparing reanalysis data with observation data at weather stations, significant differences for precipitation sums were obtained. Extreme hydrothermal phenomena are characterized by different indices, which are an integral characteristic of the conditions of heat and moisture supply of the territory. When assessing climatic extremes, special attention should be paid to the quality of the baseline information. Correction of precipitation sums using linear regression coefficients makes it possible to correctly estimate the distribution of hydro-thermal extremes (droughts or waterlogging conditions) in the study area. In this paper, we have corrected the ERA5 reanalysis monthly precipitation data. As a result of the bias correction, the deviations of the reanalysis data from the observation data have decreased. The relative difference after correction is lower than 60% at all weather stations, and from 94 (July) to 118 (August) weather stations have difference lower than 30%. The corrected datasets provide accurate and reliable information with the high spatial resolution required for proper analysis of climate change at the mesoscale in the regions of Siberia with a sparse weather station network.

Acknowledgments
The reported study was funded by the Russian Foundation for Basic Researches and Tomsk Region Administration, according to the research project grant 18-45-700015 and SB RAS Basic research program AAAA-A17-117013050037-0.

References
[1] Mokhov I I, Khon V Ch, Timazhev A V, Chernokulsky A V and Semenov V A 2014 Hydrological anomalies and trends in the Amur river basin due to climate change Extreme Floods in the Amur River Basin: Reasons, Forecasts, Recommendations (Moscow: Roshydromet) p 81-120
[2] Allen M R and Ingram W J 2002 Constraints on future changes in climate and the hydrologic cycle Nature 419 224-32
[3] Trenberth K E 1999 Atmospheric moisture recycling: Role of advection and local evaporation J. Clim. 12 1368-81
[4] Groisman P Ya, Knight R W, Easterling D R, Karl T R, Hegerl G C and Razuvayev V N 2005 Trends in Intense Precipitation in the Climate Record J. Clim. 18 1326-50
[5] Karl T R, Kukla G, Razuvayev V N, Changery M J, Quayle R G, Heim R R, Easterling D R and Fu C B 1991 Global warming: Evidence for asymmetric diurnal temperature change Geophys. Res. Lett. 18 2253-56
[6] Folland C K et al. 2001 Observed climate variability and change Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge: Cambridge University Press) p 99-181
[7] Second Assessment Report of Roshydromet on Climate Change and its Consequences on the Territory of the Russian Federation 2014 vol I (Moscow: Roshydromet) pp 18-235
[8] IPCC 2014 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 (Geneva: IPCC) p 151
[9] Zolina O G and Bulygina O N 2016 Current climatic variability of the characteristics of extreme precipitation in Russia Fundamental and Applied Climatology 1 84-103
[10] Zolotokrylin A Í and Cherenkova E Á 2017 Seasonal changes in precipitation extremes in Russia for the last several decades and their impact on vital activities of the human population Geography, Environment, Sustainability 4(10) 69-82
[11] Zolotokrylin A N, Vinogradova V V and Cherenkova E A 2007 Drought dynamics in European Russia in the context of global warming Problems of Ecological Monitoring and Modeling of Ecosystems 21 160-81
[12] Cherenkova E A 2007 The dynamics of dangerous atmospheric drought in European Russia Meteorology and Hydrology 11 14-25
[13] Voropay N N, Maksyutova E V and Balybina A S 2011 Contemporary climatic changes in the Predbaikal region Environ. Res. Lett. 6 045209 doi:10.1088/1748-9326/6/4/045209
[14] Voropay N N, Maksyutova E V and Riazanova A A 2016 Hydrothermal conditions at the south of East Siberia during the ongoing warming IOP Conf. Series: Earth and Environmental Science 48(1) 012003 doi:10.1088/1755-1315/48/1/012003
[15] Ryazanova A A and Voropay N N 2017 Droughts and Excessive Moisture Events in Southern Siberia in the Late XXth – Early XXIst Centuries IOP Conf. Series: Earth and Environmental Science (EES) 96(1) 012015 doi:10.1088/1755-1315/96/1/012015
[16] Voropay N N and Ryazanova A A 2018 Atmospheric droughts in Southern Siberia in the late 20th and early 21st centuries IOP Conf. Series: Earth and Environmental Science 211 012062 doi:10.1088/1755-1315/211/1/012062
[17] Kokorev V A and Shersyukov A B 2015 Meteorological data for studying the current and projected for the future climate change in Russia Arctic XXI Century. Natural Sciences 2 5-23
[18] Kharyukina E V, Martynova Y V and Loginov S V 2018 Atmospheric precipitation and its anomalies in Western Siberia against the background of global climate change from reanalysis and numerical modeling data IOP Conf. Series: Earth and Environmental Science 211 012009 doi:10.1088/1755-1315/211/1/012009
[19] Aleshina M A, Toropov P A and Semenov V A 2018 Changes in the temperature and humidity regime of the Black Sea coast of the Caucasus in the period 1982-2014 Meteorology and Hydrology 4 41-53
[20] Groisman P Ya, Bogданова Е G, Alexeev V A, Cherry J E and Bulygina O N 2014 Impact of snowfall measurement deficiencies on quantification of precipitation and its trends over Northern Eurasia Ice and Snow 2(126) 29-43
[21] Zolina O G 2011 Changes in the duration of synoptic rainy periods in Europe from 1950 to 2008 and their relation to extreme precipitation Dokl. Earth Sci. 436 279-83
[22] Zolina O G, Kapala A, Simmer C and Gulev S K 2004 Analysis of extreme precipitation over Europe from different reanalyses: A comparative assessment Glob. Planet. Change 44 129-61
[23] Voropay N N and Ryazanova A A 2018 A comparative assessment of the aridity indices for analysis of the hydrothermal conditions IOP Conf. Series: Earth and Environmental Science 190 012041 doi:10.1088/1755-1315/190/1/012041
[24] Fowler H J, Blenkinsop S and Tebaldi C 2007 Linking climate change modelling to impact studies: Recent advances in downscaling techniques for hydrological modelling Int. J. Climatol. 27 1547-78
[25] Hewitson B and Crane R 1996 Climate downscaling: Techniques and application Clim. Res. 7 85-95
[26] Maraun D et al. 2010 Precipitation Downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user Rev. Geophys. 48 RG3003
[27] Prudhomme C, Reynard N and Crooks S 2002 Downscaling of global climate models for flood frequency analysis: Where are we now? Hydrol. Processes 16(6) 1137-50 doi:10.1002/hyp.1054
[28] Maraun D and Widmann M 2018 Statistical Downscaling and Bias Correction for Climate Research (Cambridge: Cambridge University Press) p 360
[29] von Storch H and Zorita E 2019 The History of Ideas of Downscaling – From Synoptic Dynamics and Spatial Interpolation Front. Environ. Sci. 7(21) 00021 doi: 10.3389/fenvs.2019.00021
[30] Christensen J H, Boberg F, Christensen O B and Lucas-Picher P 2008 On the need for bias correction of regional climate change projections of temperature and precipitation *Geophys. Res. Lett.* **35** L20709 doi:10.1029/2008GL035694

[31] Hanssen-Bauer I, Achberger C, Benestad R, Chen D and Forland E 2005 Statistical downscaling of climate scenarios over Scandinavia *Clim. Res.* **29**(3) 255-68

[32] Mamalakis A, Langousis A, Deidda R and Marrocu M 2017 A parametric approach for simultaneous bias correction and high-resolution downscaling of climate model rainfall *Water Resour. Res.* **53** doi:10.1002/2016WR019578

[33] Langousis A, Mamalakis A, Deidda R and Marrocu M 2016 Assessing the relative effectiveness of statistical downscaling and distribution mapping in reproducing rainfall statistics based on climate model results *Water Resour. Res.* **52** 471-494 doi:10.1002/2015WR017556

[34] Ryazanova A A, Voropay N N, Okladnikov I G and Gordov E P 2016 Development of computational module of regional aridity for web-GIS “Climate” *IOP Conf. Series: Earth and Environmental Science* **48**(1) 012032 doi:10.1088/1755-1315/48/1/012032

[35] Copernicus Climate Change Service (C3S): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS) Income accessed online on 03th of September 2019 via https://cds.climate.copernicus.eu/cdsapp#!/home

[36] Dee D P et al. 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system *Quarterly Journal of the Royal Meteorological Society* **137** 553-97

[37] Webster R and Oliver M A 2007 Geostatistics for Environmental Scientists. Second edition, (Chichester: John Wiley & Sons Ltd.) p 330

[38] Shukla K, Kumar P, Mann G S and Khare M 2020 Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi *Sustainable Cities and Society* **54** 101997