



River runoff modeling in the European territory of Russia

O. Yermolaev^{a,b}, S. Mukharamova^b, E. Vedeneeva^{b,*}

^a Belgorod State National Research University, Belgorod, Russia

^b Department of the Landscape Ecology Institute of Ecology and Environmental Sciences, Kazan Federal University, 420097, Kazan, 5 Tovarisheskaya str., Russia

ARTICLE INFO

Keywords:

River runoff
Geoinformation database
Landscape zones
Modeling
Multiple regression

ABSTRACT

The paper describes river runoff modeling for a plains region of the European territory of Russia (ETR), as well as a prediction for ungauged drainage basins. The study of river runoff is one of the key research objectives in determining the patterns of sediment yield formation. Among many other zonal factors, river runoff is considered to be the main factor in sediment yield formation in a humid climate. In this study, modeling results for the entire European territory of Russia and various landscape zones are presented via the use of multiple regression methods. Multiple regression methods do not require the mathematical description of the main physical processes of runoff formation in terms of their spatial heterogeneity. At the same time, such methods can be distinguished by their simplicity in terms of determining parameters and providing clear interpretations of the results. The research methodology in this work is based on a drainage basin approach. Initial data for the river runoff and its formation factors are presented in the open-access geoinformation database "Drainage basins of the European territory of Russia", which has been created earlier by the authors. The river runoff geodatabase was formed with results from 1440 gauging stations. The independent variables, such as the relief morphometric characteristics, climatic indicators reflecting average values, scale, seasonal variations, extreme values of temperature and precipitation, percentage of forest and swamp cover, plowing, percentage of meadows, assessment of the anthropogenic impact on the drainage basin, geographical coordinates of the centroid, prevailing soil type, type of soil-forming rock, and class of pre-Quaternary deposits are used for modeling here. Data processing and model development is conducted using the R software environment. Models obtained by linear and nonlinear methods explain about 85–88% of data variability and are well interpreted in terms of the water balance equation. It is found here that the most significant predictors in the model are annual precipitation, the sum of the active temperatures (characterizing runoff losses via evaporation), average slope gradient, and the forest cover of the catchment. For Environmental Resources Management, it is required that data for river runoff are collected at the local (municipal) level. The results for the extrapolation of the river runoff values to ungauged river basins in a plains region of the European territory of Russia are presented here. Calculations of predicted values for the river runoff are given based on the obtained discharge per unit area logarithm model. The model and its cartographic representation reflect the patterns of the spatial distribution of river runoff for the level of spatial detail accepted in the study. The methods applied in this study and the results obtained could be used for similar studies of plains territories across the world.

1. Introduction

The study of river runoff is a relevant research task due to the irregular placements of gauging posts and stations and their decreasing number in Russia within recent decades. The application of modern methods for runoff modeling allows statistically reliable and physically proven dependencies of river runoff to be obtained. These results are based on the factors of formation, followed by a further prediction of

runoff values for ungauged drainage basins.

The first observations of the river runoff in the ETR were conducted in the second half of the 19th century. Gauging station development began on the largest rivers of the region, such as the Volga, Northern Dvina, Don, and Kuban. Intensive development of the gauging station network started in the 1920 s and continued after the Second World War, reaching the number of stations in the 1970 s ([Resources of surface waters of the USSR, 1973](#)). By the end of the 20th century, the number of

* Corresponding author.

E-mail address: vedeneeva-evgeniya@mail.ru (E. Vedeneeva).

<https://doi.org/10.1016/j.catena.2021.105327>

Received 26 May 2020; Received in revised form 16 March 2021; Accepted 19 March 2021

Available online 8 April 2021

0341-8162/© 2021 Elsevier B.V. All rights reserved.

stations began to decline due to the economic downturn in the country, which has led to budget cuts at the Federal Service for Hydrometeorology and Environmental Monitoring ([Overview of hydrometeorological observations, data processing and preparation of information products, 2017](#)).

The accumulation of sufficient hydrological information for statistical processing has allowed the first spatial generalizations of river runoff to be created. The first river runoff map for the European part of the USSR was created in 1927 by Kocherin ([Kocherin, 1927](#)). To create the map, data from 34 gauging stations were used. The distribution of river runoff across the territory was shown using smooth isolines. Despite the small amount of initial data, Kocherin's map accurately reflects the zonal changes in runoff ([Voskresensky, 1962](#)).

Later, USSR river runoff maps were created by Zaikov using data from 987 gauging stations ([Zaykov and Belinkov, 1937](#), [Zaykov, 1946](#)). In 1961, a river runoff map for the European part of the USSR was created by Voskresensky, who used data from 1677 gauging stations. This map was significantly different from the previous maps in the northern and southeastern regions, although it mostly retains the same isoline character in the other areas. However, the river runoff values calculated using such maps are zonal and do not reflect the impact of local factors of runoff formation ([Voskresensky, 1962](#)).

Currently, the river runoff research using modern observation data is in progress. Studies in this field include studies of runoff characteristic changes ([Koronkevich, 1990](#); [Dzhamalov et al., 2015a, 2015b](#)), the re-evaluation of surface and groundwater resources ([Dzhamalov et al., 2014](#)), runoff prediction for the largest rivers in the ETR via physical and mathematical calculation ([Gordeeva and Malinin, 2018](#)). There are studies of the river runoff in the ETR that have been conducted at the individual catchment level. There are studies related to the research of annual runoff estimates in the Volga River drainage basin ([Alekseevskij et al., 2013](#), [Frolova et al., 2014](#)), the Oka and Don drainage basins ([Grigoriev et al., 2018](#)), the Northern Dvina River drainage basin ([Belchikov et al., 2013](#)). The results of these studies are presented in river runoff distribution maps using the isoline method or the cartographic method, summarizing data for the catchments of the largest rivers, natural zones, and administrative units.

Mathematical modeling methods are widely used to estimate the runoff values and perform forecasting. The developed models vary in terms of the details describing the hydrological processes, the required initial information, the method of setting the mathematical structure, feasibility, accuracy, and reliability ([Kuchment et al., 2013](#)).

One of the most frequently used methods of river runoff modeling is multiple regression. The application of multiple regression to the study of the river runoff formation conditions has been successful at various levels, from the local level, covering individual catchments, to regional and global levels ([Duan Limin et al., 2010](#); [Barbarossa et al., 2017](#); [Zengin et al., 2017](#)).

Various natural and anthropogenic factors (predictors) of river runoff formation can be used as input data for regression models. It is noteworthy that only climatic factors, e.g., precipitation, are often used to model runoff in individual catchments, while factors such as the topography and land cover type are purposely ignored ([Patel et al., 2016](#)). This is usually done so that runoff values may be quickly determined. The relationship between runoff data and climatic factors can be studied at different time scales. An example application of a multiple regression method is investigating the response of river runoff to climate change ([Chong-Li et al., 2014](#)).

Several studies have assessed the impact of a single factor on runoff. This study ([Živković et al., 2015](#)) describes the impacts of different vegetation types (forests, meadows, and arable land) on river runoff. There are factors that could be used in this regard, such as forest cover percentages and vegetation indices ([Ning et al., 2017](#)). The so-called landscape factor can also be used as a predictor, representing the proportions of areas occupied by different landscapes from the total catchment area ([Terski et al., 2017](#)). In this case, the landscape factor is

an integrated factor that combines the geological structure of the area, as well as the topography, soil, and vegetation conditions. Multiple regression methods could be successfully applied using indicators such as the catchment shape and terrain characteristics ([Zengin et al., 2017](#)).

The usage of multiple regression allows relevant models of runoff formation to be obtained at the regional level alongside river runoff modeling in individual catchments. In this case, data for climatic indicators and underlying surface characteristics are considered as predictors ([Reimers, 1990](#); [Vogel et al., 1999](#); [Tran et al., 2015](#)). Such studies are relevant due to the lack of hydrological monitoring in the territory as a result of the low availability of gauging stations. Existing studies suggest that the results of regional runoff modeling can be successfully implemented in geoinformation databases.

With the development and implementation of geoinformation systems, coupled with the availability of digital data, the acquisition of various indicators representing runoff formation factors becomes far more efficient and allows runoff simulation with higher accuracy ([Zhu and Day, 2009](#)).

It should be noticed that regression models allow establishing runoff relationships and catchment characteristics at the global level. The study by [Burgers et al. \(2014\)](#) presents runoff formation models that use the catchment area, runoff length, and annual precipitation as independent variables. The study suggests that by adding the precipitation indicator to the model, its quality may be improved from 40% to 56%. The study ([Barbarossa et al., 2017](#)) shows successful implementation of a multiple regression method for creating a global model of runoff. This model allows calculating mean annual discharge at any point in a river network on a global scale. Furthermore, in addition to developing a model itself, this study provides a comparison with the existing hydrological model PCR-GLOBWB. The results of the analysis reveal that the developed regression model has a higher quality for river runoff forecasting with a lower standard deviation.

The purpose of this research is to model the river runoff dependencies in terms of the conditions of formation in various landscape zones in the ETR plain's part using a drainage basin approach.

2. Description of the natural conditions of the river runoff formation within the ETR territory

The total area of the ETR is about 3.8×10^6 m², and the plains area (excluding the Caucasus and the Ural Mountains) is 3.5×10^6 m². The territory stretches from north to south for more than 2400 km and is characterized by great landscape diversity. This territory is inhabited by the majority of the Russian population (almost 80 million people).

The description of the natural conditions causing the formation of river runoff is given by a developed GIS which is presented on an open-access geoportals ([Yermolaev et al., 2018](#)). In terms of the lithological structure of rocks there, the sedimentary poorly defined formations (71%) and chemoorganogenic rocks (12.1%) are found as the most widespread types. Absolute heights there range from -28 to 830 m. Almost 37% of the territory is located at an altitude of 120–180 m. The average height of the relief is 140 m. Most of the drainage basins have slopes from 0.5 to 1.5° (about 60%), 14% of slopes in the territory are less than 0.5°, and steep slopes can be found for less than 0.5% of the territory. The dominant slopes are those with average lengths of 500–700 m (69%). Slopes with an average length of over 700 m represent about 17% of the territory. The zonal distributions of climatic elements (as well as disproportionate heat and moisture in different parts) are considered to be the main characteristics of the climate. Atlantic air masses are transformed as they move inland, and climatic conditions change significantly from west to east, resulting in long-term landscape differentiation. The distributions of temperature characteristics can be seen. The average annual temperature changes from -8 °C in the northeastern part of the ETR to 12 – 14 °C on the Black Sea coast and the Caspian lowland. Annual precipitation has a maximum value in the western part of the ETR and is about 600 mm, with a trend to decrease in

the north and, in the southeast (Fig. 1), an extremely heterogeneous distribution of precipitation is typical for different natural zones (Table 1).

Clayey and heavy loamy soils dominate in the granulometric soil texture distribution (61.3%), with light and medium loamy soils representing 19.7% and 14.2%, respectively. Swamp and forest cover for drainage basins decreases from north to south, following a latitudinal zoning principle. A map of swampland and forest cover in drainage basins is presented in the Supplementary Materials (Supplementary figure 3).

Highly fertile soils are located on the southern part of ETR, which is an area that was intensively plowed for a long period (Fig. 2).

The following landscape zones were distinguished from north to south on the study area: subarctic tundra and forest-tundra, boreal

forest, broad-leaved forest, forest-steppe, steppe, semi-arid, and arid (Table in the Supplementary Materials).

3. Methods and sources of data

The current study is based on the data accumulated in the previously created GIS geodatabase “River basins of the European territory of Russia” (Ermolaev et al., 2017; Yermolaev et al., 2018). The data include observations of river runoff at ETR river gauging stations and landscape and geographical conditions, as well as anthropogenic pressure in catchment areas related to the gauging stations.

The drainage basin approach was used for studying the river runoff. Two grids for drainage basin boundaries were developed in a vector format using GIS technologies: the boundaries of the drainage basins and

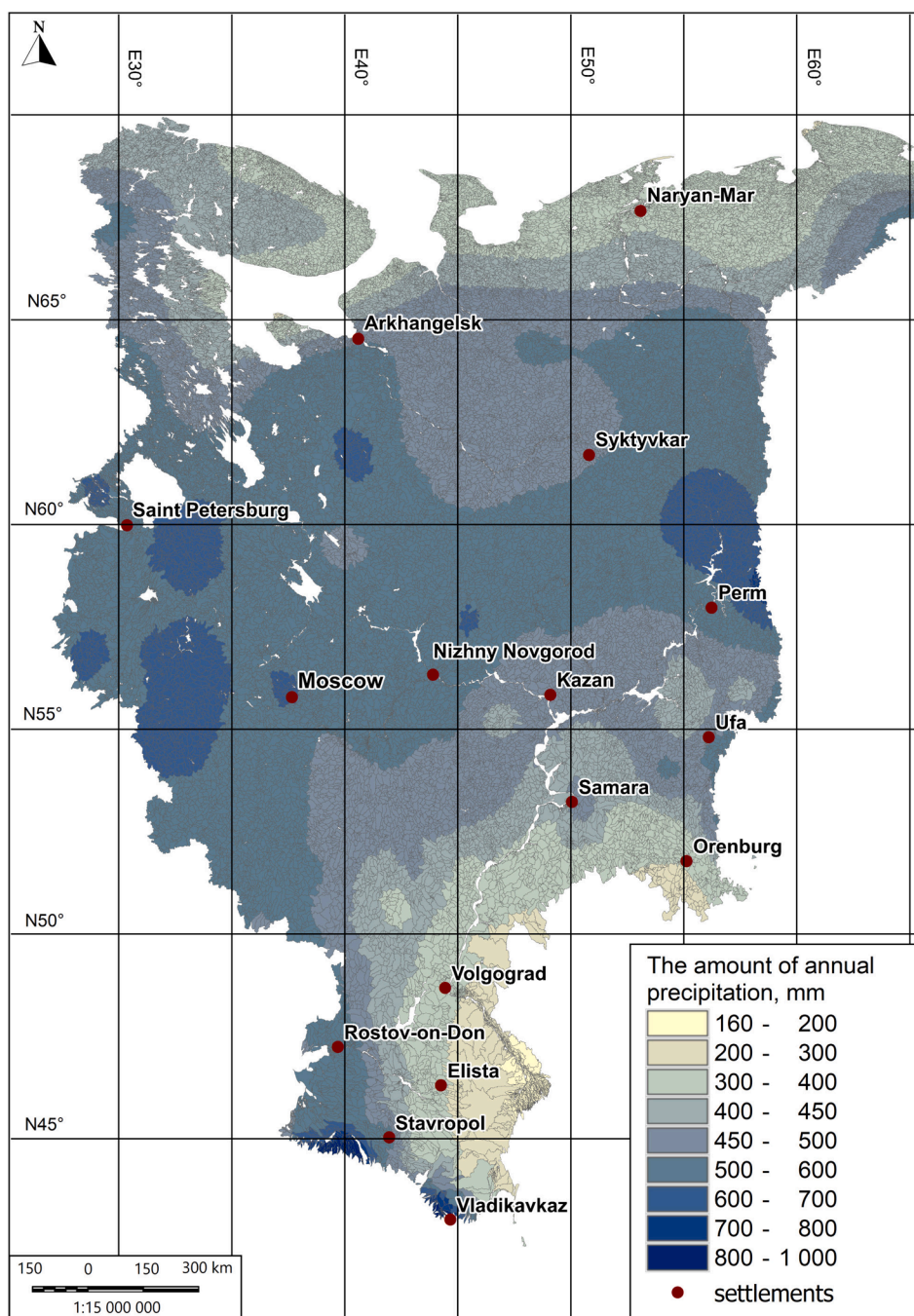


Fig. 1. Map of annual precipitation distribution in the ETR plain zone.

Table 1

The distribution of precipitation according to ETR landscape zones (1960–2018).

Landscape Zone	Rmean, mm	Rcold, mm	Rwarm, mm
Subarctic tundra zone and forest-tundra zone	376	157	219
Boreal forest zone and broad-leaved forest zone	525	161	364
Forest-steppe zone	474	136	338
Steppe zone	420	124	296
Semi-arid zone	298	78	220
Arid zone	204	43	161

Note: Rmean – The amount of annual precipitation; Rcold – snow (solid precipitation); Rwarm – rainfall. The approach to differentiating precipitation was as follows: The year is divided monthly into two periods: warm (when the average monthly air temperature > 0) and cold (when the average monthly air temperature ≤ 0). Accordingly, the average rainfall for the warm season is Rwarm; the average rainfall for the cold season is Rcold.

the boundaries of the gauging station drainage basins (Fig. 3).

The Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) with a spatial resolution of 250 m (Danielson and Gesch, 2011) and the raster model of the hydrographic network were used in constructing the boundaries of the drainage basins for the entire area of study, and in generating a corresponding electronic vector map as well. The boundaries were delineated with the automatic mode using the algorithm implemented in the Whitebox Geospatial Analysis Tools software product (Lindsay, 2014). The technique of delineating drainage basins is presented in detail in the research of (Ermolaev et al., 2017; Yermolaev et al., 2018). The total number of drainage basins in the ETR is 53,865, including 49,516 basins in plains regions. The average area for the allocated drainage basins is 74.48 km². 3.1. Model inputs

The study uses multidimensional samples to statistically analyze the river runoff patterns. The first sample includes elements such as gauging station drainage basins with river runoff data. For this sample, dependent variable Y represents the discharge per unit area in a gauging station drainage basin. Values for the dependent variable were obtained as the average long-term annual river runoff recorded in the gauging station for the whole period of observation (equal to or more than 10 years) as per the area assigned to the gauging station drainage basin.

Table 2 shows the data sources for the river runoff characteristics obtained at the ETR river gauging stations.

The hydrological observation data provide information for the river runoff characteristics across 1440 gauging stations located within plains in the ETR. The unified database (Ermolaev et al., 2017, Yermolaev et al., 2018) created by the systematization and generalization of these materials contain data for the geographical position for each of the 1440 gauging stations, along with the period for the river runoff observations, the long-term average value of river runoff (m³ s⁻¹), and the area of surface catchment related to the gauging station (km²).

The natural zones of the study area are characterized by different degrees of hydrological observation availability (Fig. 4). The highest gauging station densities exist in south boreal and broad-leaved forests, as well as in forest-steppe zones (average of 5–6 stations per 10,000 km²). The density of stations in the northern uninhabited areas (forest-tundra and tundra zones) is much lower (1–3 stations per 10,000 km²).

Table 3 presents the data for the number of gauging stations with catchment areas of different sizes and different observation series durations. The data are also extremely uneven in terms of the observation series duration. The period for runoff observations is considered to be from 1859 to 2017 here. The observation periods at stations vary from 1 to 134 years. It can be assumed that due to some of the stations having a very short period of observation, they cannot sufficiently represent reliable statistical results. Notably, 85% of gauging stations feature more than 10 years of observation.

For about 1% of stations, the surface catchment areas of the station

exceed 50,000 km². It can be noted that such catchment areas, as a rule, are allocated in several natural zones. This reduces runoff zonal characteristics and does not allow the description of the conditions of formation, making observational data at such stations poorly suitable for the current study. Most of the data can be found for gauging stations with catchment areas ranging from 1000 to 5000 km². Notably, 94% of such gauging stations have an observation series duration of more than 10 years.

The independent variables {X} in the sample are quantitative and qualitative characteristics that describe the runoff formation conditions. Table 4 shows the data sources for the natural and anthropogenic conditions for river runoff formation. The statistical environment R and PostgreSQL were used to process the source data (climatic, soil, geological, landscape characteristics, and structure of land use). The morphometric relief characteristics were calculated based on the GMTED2010 model using the ArcGIS software package and WhiteBox GAT.

3.1. Methods of statistical data analysis and model development

The analysis of statistical dependencies was carried out at both the whole ETR plain territory level and particular landscape zones related to the given gauging station location, such as steppe, forest-steppe, broad-leaved forest, or boreal forest. It can be noted that the data quantity required for reliable results for statistical analysis in the tundra, forest-tundra, semi-desert, and desert zones of ETR is relatively low.

The study analyzes the distribution and the explanatory variable structures, along with their multicollinearity, individual correlations, rank correlations, etc. In particular, the study analyzes the so-called variance inflation factor (Fox and Monette, 1992), which is used to determine the degree of correlation between a set of variables as strong multicollinearity significantly increases the error in estimating the parameters.

The study uses a generalized linear model (GLM) and generalized additive model (GAM) (Hastie and Tibshiran, 1990, Dobson, 2002, Wood, 2006, Zuur et al., 2012, 2014) as the main statistical methods to model the dependence of the discharge per unit area on external variables. These methods allow the distribution of a dependent variable that is different from the normal to be considered. The nonlinear GAM method for each predictor evaluates the function of an undefined species by applying optimal prediction splines. Many coefficients in such functions do not allow writing the regression equation explicitly. The model is described by a graphical representation of individual dependencies which are defined by the functions for each of the significant predictors (Hastie and Tibshiran, 1990). The method assumes that individual dependencies are smooth and distorted by the presence of random errors.

Several experiments were conducted and an optimal subset of predictors was selected based on the analysis of their collinearity, i.e., the statistical significance of the predictor's contribution to the model. As a result of the experiments, the best models describing the dependencies of the discharge per unit area on external factors were selected. As a method of model comparison, the AIC was used based not only on the model adjustment to the data but also on the resources used (Sakamoto et al., 1986). To assess the model quality, the determination coefficient (percentage of data variability explained by the model) was adjusted to the number of model regressors used (Seber, 1980). Moreover, statistical data were used based on residual analysis (mean residuals, mean absolute error, standard estimation error, etc.). To develop a model, standardized predictor values were considered, which simplified the linear model interpretation by comparing the contribution of variables into the runoff value.

In the process of model development, the stability of model estimates for sub-samples with various limitations (by the number of years of observations, by catchment area, by river order) was checked. The representativeness of corresponding gauging stations in the studied territory was controlled during this process. Statistical evaluations,

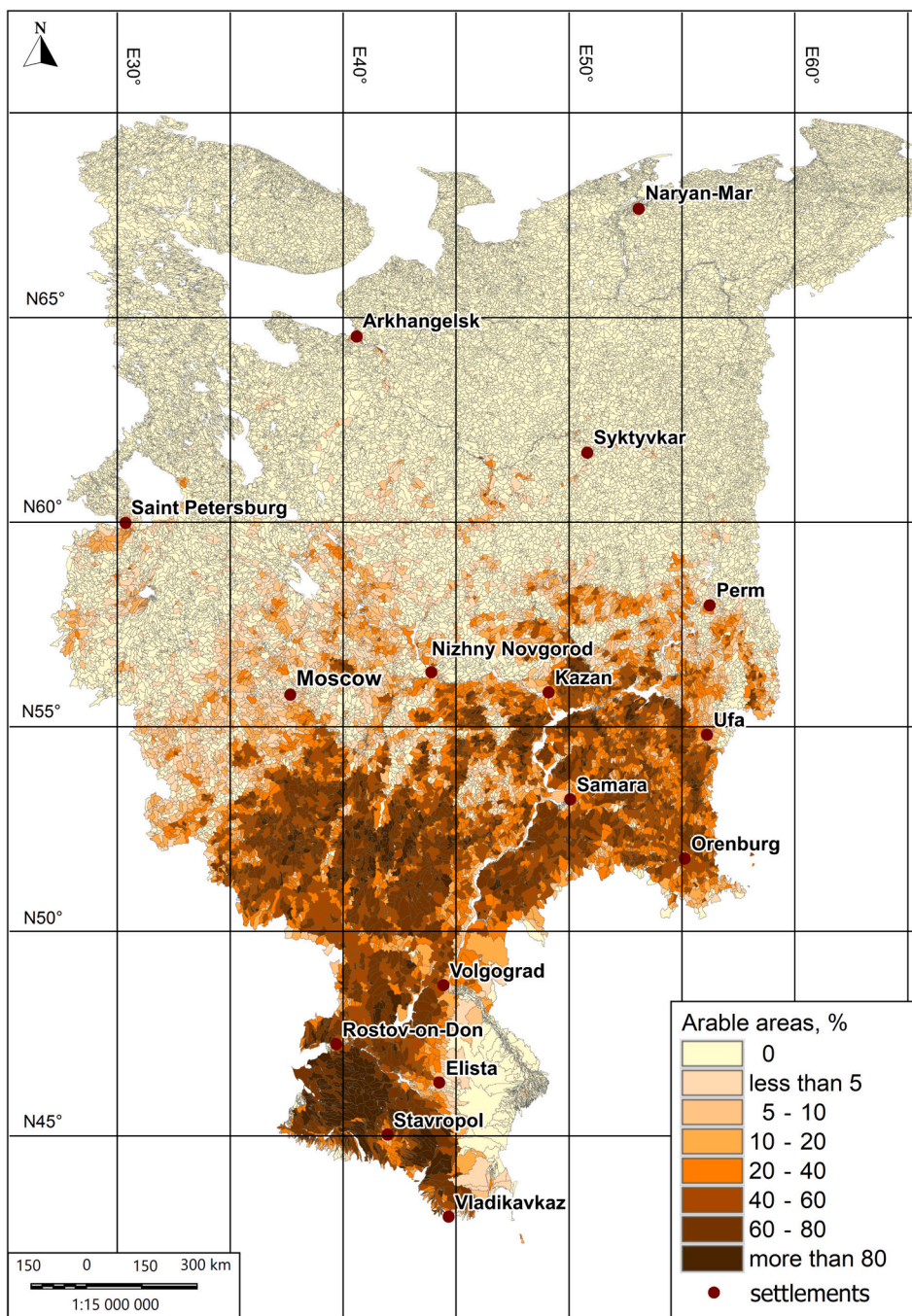


Fig. 2. Map of arable area percentage in the drainage basins of the ETR plain zone.

modeling, and calculations were performed using the mgcv, nlme, and geoR software packages (Wood, 2006; Pinheiro et al., 2014) in the statistical environment R (R Core Team, 2014, <http://www.R-project.org/>).

4. Results

4.1. Modeling of river runoff

The GLM of the discharge per unit area in the ETR plain zone explains more than 85% of the data variability. The determination coefficient values indicate that the model reflects the general patterns of the river runoff. The models include the most significant predictors, such as the sum of active temperatures (sum of average daily temperatures for

days when the temperature is above 10 °C), annual precipitation, average slope gradient, percentage of forested area (Table 5). The advantage of the statistical model developed is its good interpretability in terms of the water balance equation. This includes the positive contribution of precipitation amount, the negative contribution of active temperatures, which characterizes the evaporation capacity of the climate.

According to the model, the temperature gradient (sum of active temperatures) mainly contributes to the variability of the discharge per unit area at a regional level. It creates the zonal trend for the river runoff, i.e., a decrease in the discharge per unit area from north to south. The negative impact on the discharge per unit area active temperatures sum can be explained by the water losses through evapotranspiration. The boreal forest zone and steppe zone significantly contribute to this

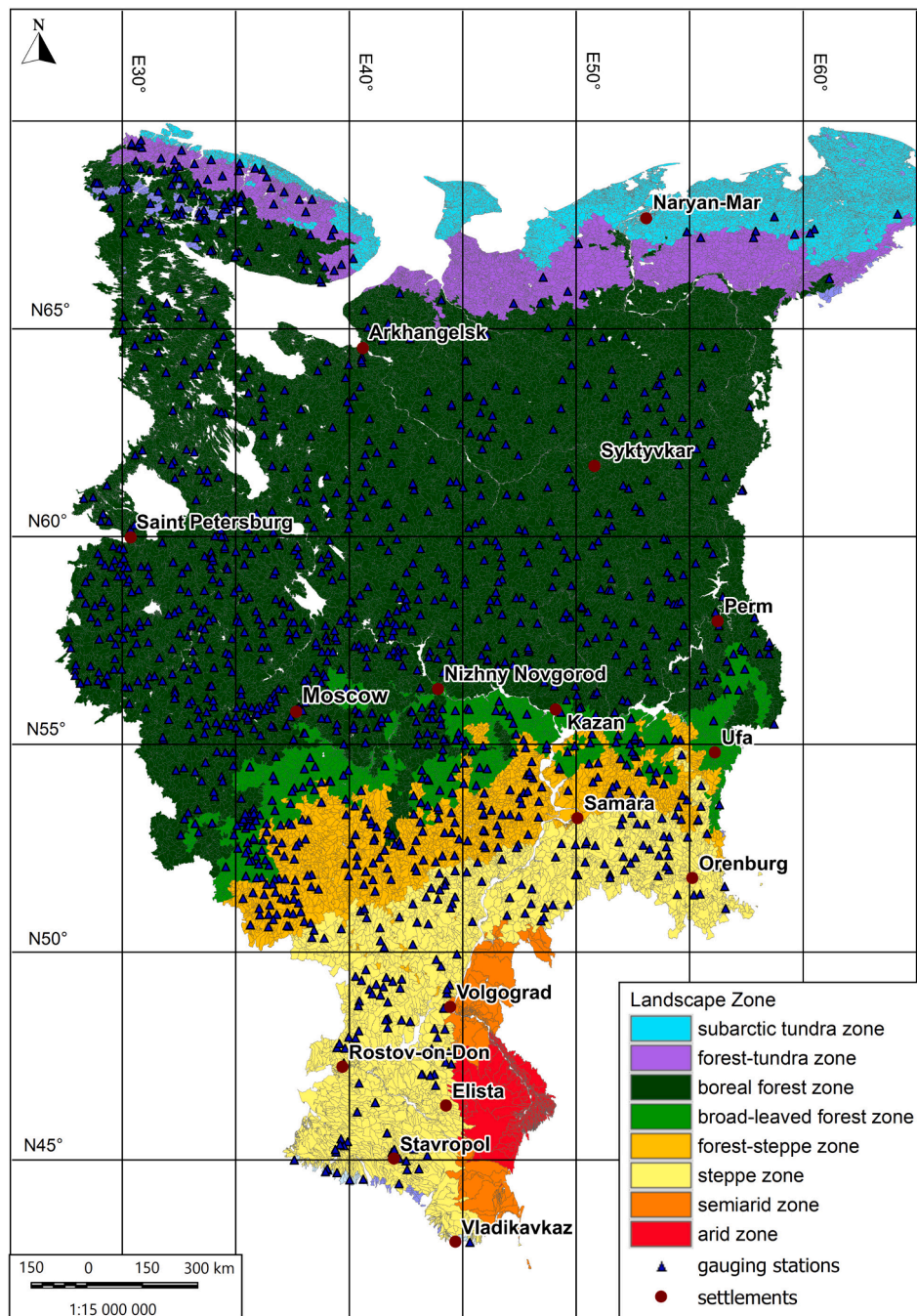


Fig. 3. Map fragment of the drainage basin boundaries of the associated with gauging stations of ETR rivers.

factor, which has broadly extended from the north to the south in the study region. In forest-steppe regions, particularly in broad-leaved forest zones, the gradients of thermoclimaticity are not as significant, therefore other factors contribute more significantly to the runoff variability within these zones.

The model describes the positive contribution of precipitation in the boreal forest zone, forest-steppe zone, and steppe. This factor is the least significant for the broad-leaved forest zone, where the forest cover and relief conditions are the main ones, i.e., the gradients of slopes and slope lengths. The positive contribution of the slope gradient is explained by its inverse effect on evaporation, i.e., the steeper the slopes in a drainage basin are, the shorter time it takes for slope surface runoff to reach riverbeds with less water evaporating in the process. The increase of flow length in plain terrain conditions results in higher losses of water

for evaporation, infiltration, and its transfer to underground runoff, as well as the loss of runoff volumes due to lagging in the terrain while traversing long flattened slopes. At the same time, for low relief in the boreal forest zone, the influences of relief factors (average slope gradients and slope lengths) are significantly less.

The positive contribution to the forest cover near the drainage basin can be explained by the fact that the amount of water reserved in forest soils is higher than in non-forested areas due to their better water retention properties. This water is delivered to the river network in the form of groundwater feed during the annual low streamflow period, thus increasing the annual discharge per unit area. It can be noted that for the boreal forest zone, the forest cover is an important factor in the river runoff model. The high extent of forest cover at drainage basins improves the conditions of water infiltration and groundwater reserve

Table 2
Sources of data on river runoff of ETR rivers.

Data Source	Location of gauging stations	Period of observation	The number of gauging stations
Publication of long-term regime observations at gauging stations – «Surface Water Resources»	European territory of Russia	From the beginning of observations until 1975	1866
Caspian Sea Laboratory of the Institute of Water Problems of the Russian Academy of Sciences (http://caspi.ru/)	the Volga basin and the Kama basin	Data update until 1985	404
Global Runoff Data Centre Open Data (http://www.bafg.de/GRDC/EN/Home/homepage_node.html)	European territory of Russia	From the beginning of observations to 2011	370
Data from FSBU «VNIIGMI-ICDC» (World Data Center)	the Upper and Lower Volga basin, Kama basin	Data update until 2013	823
Automated Information System for State Monitoring of Water Objects of the Federal Water Resources Agency (https://gmvo.skniivh.ru/)	European territory of Russia	2008–2017	966

recharging, which feeds the rivers during the intermodal period.

In comparison to the linear methods, the added complexity of modeling methods does not significantly affect model quality. The model developed for the plain areas of the ETR using the GAM method explains up to 88% of data variability here. It includes the same independent variables, without losing the direction of influence on the

discharge per unit area or its interpretation. Fig. 5 shows non-parametric functions of individual dependencies included in the nonlinear model for plain territories (these dependencies are very close to linear). The results were slightly improved by the refinement of individual dependencies within certain landscape zones. For example, Fig. 6 shows non-parametric functions of the annual amount of precipitation for forest-steppe zones and the slopes lengths for forest zones.

Table 6 shows the statistical indicators of correlation between the data obtained in the model with the initial data for the two modeling methods (linear and nonlinear) based on the residual analysis (ME—mean error, MD—median error, MAE—mean absolute error, MSE—mean square error, SE—standard error of estimation, which is used for estimation of forecast confidence limits). Figs. 7 and 8 illustrate the qualities of the models developed using the GLM and GAM methods, respectively. These figures present diagrams of the obtained and predicted logarithm values of the discharge per unit area (model values are displayed against observed values), along with histograms of the residual frequency values and diagrams showing the dependences of the residuals on the model values and predictors.

It can be noted with 95% probability that the accuracy of the forecast given by the GLM for the logarithm of the river runoff unit is approximately ± 0.49 , and for the GAM it is ± 0.46 . In comparison, a sample value range is about 4.5. Thus, the developed statistical models provide a representation of the actual variability of the discharge per unit area for the plains part of the ETR with a forecast error of about 10%.

To analyze the results in more detail, gauging stations, where the observed runoff values were significantly different from the predicted ones (both underestimated and overestimated), were considered. Fig. 7a and 8a are related to the points that are most distant from the cloud. To understand the characteristics of the territory not included in the model and the cause of such misalignment, the local conditions for river runoff formation in these drainage basins were considered. Firstly, it was found that the misalignment was caused by the soil characteristics in the drainage basins, the variability of which was not sufficiently detailed in the analyzed multidimensional samples. Significantly lower runoff was

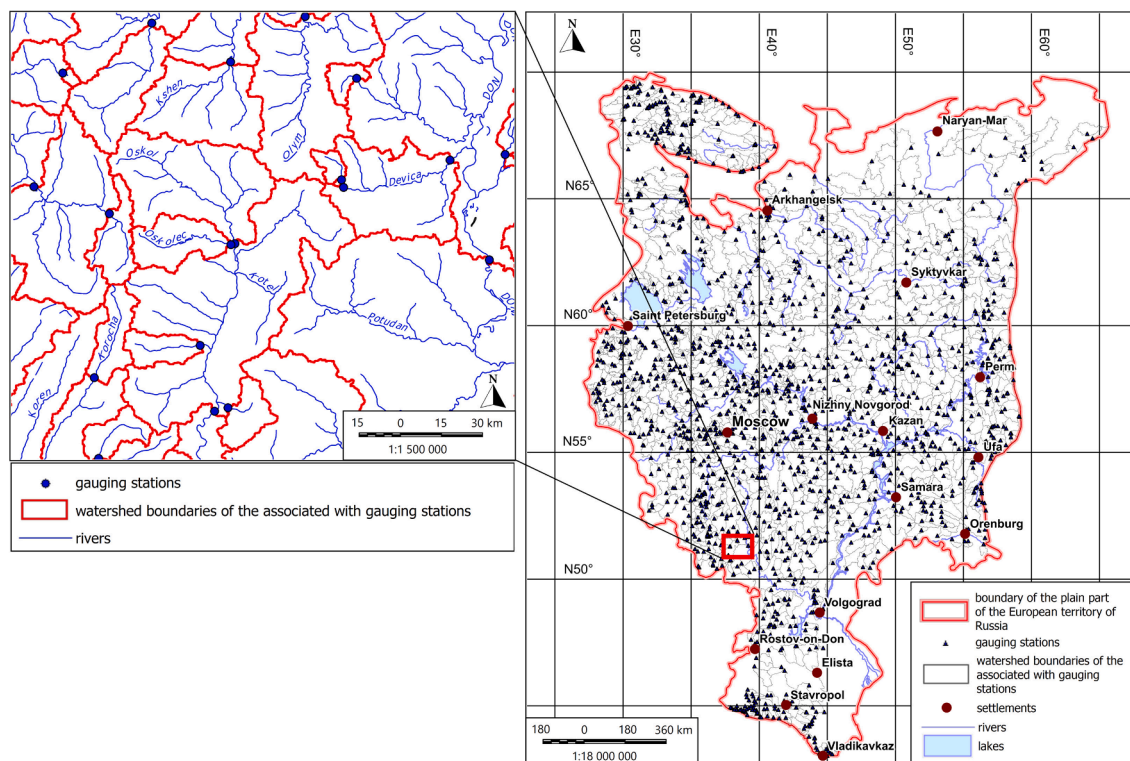


Fig. 4. Map of gauging stations within the landscape zones the ETR plain zone.

Table 3
Distribution of gauging stations by drainage basin area and duration of river runoff observations.

Basin area, sq. km.	Duration of observation						Total	%
	1 year	2 to 5 years	6 to 10 years	11 to 25 years	25 to 40 years	over 40 years		
Under 500	12	43	55	167	73	65	415	28.8
500–1 000	3	13	24	79	58	61	238	16.5
1 000–5000	1	11	20	137	159	185	513	35.6
5 000–10 000	0	5	5	22	32	54	118	8.2
10 000–50 000	1	2	14	25	19	70	131	9.1
50 000 – 100 000	1	0	2	1	2	9	15	1.0
Over 100 000	0	0	1	0	3	6	10	0.7
Total	18	74	121	431	346	450	1440	100
%	1.3	5.1	8.4	29.9	24	31.3	100	

observed in Unzha River (Verkhounzha village – 55° 27' 16" N; 41° 30' 43" E) gauging station when compared to its predicted value. This was due to the spread of sandy soils in the catchment of this gauging station. For the Bir River gauging station catchment (Malosuhoyazovo village – 55° 23' 20" N; 55° 51' 54" E), where the observed runoff was significantly higher than predicted, this was caused by the karst influence, specifically by the intensive groundwaters drainage by river areas formed in karst areas (Resources of surface water of USSR, 1973). Another possible reason can be found while studying the Birla River gauging station (Vishenka village – 53° 55' 09" N; 49° 33' 36" E), where the observed runoff was lower than predicted due to the high regulation of the river and the subsequent runoff discharge for irreversible evaporation.

4.2. Spatial extrapolation of the river runoff modeling results for ungauged river basins

The forecasted river runoff values for the small drainage basin were calculated based on the GLM logarithm model of the discharge per unit area. In other words, the extrapolation of the river runoff values to the ungauged areas of the ETR plain zone was performed.

Forecasted values were calculated using multiple sample in the statistical environment R. The small river drainage basins that cover the study area were considered as the elements in this sample. For each element in the sample, a set of independent variables {X} including quantitative and qualitative characteristics describing runoff formation was defined.

The calculated discharge per unit area values and annual runoff layers can be presented both in table and function forms for the drainage basin vector layer attributes. The mapping of the obtained river runoff values to the vector layer allows the creation of thematic maps in various GIS applications (MapInfo, ArcGIS, QGIS). For example, the cartographic representation of the obtained results could take the form of a map of the discharge per unit area and the annual runoff layer (Fig. 9).

The obtained results have been verified using the National Atlas of Russia maps ("River Runoff") (National Atlas of Russia, 2007) and the runoff layer maps obtained from 605 gauging stations for both before and 1978 (Frolova et al., 2020). The data from the gauging stations for which the mean annual discharge was determined were used to create the map. The values of the discharge per unit area are related to the centers of the gauging station drainage basins. The isolines between the centers of gauging station basins were developed. In mountainous areas, interpolation was carried out to take terrain changes into account (Zaitseva and Koronkevich, 2019). Both maps are presented in the Supplementary Materials (Supplementary figure 1 and Supplementary figure 2). It can be noted that the runoff layer map that has been generated from the simulation results corresponds to the actual runoff values.

The global runoff model was also used to verify the obtained model and predicted runoff indicators (Barbarossa et al., 2017). The model was developed based on runoff data that were averaged over 30 years. The model quality was estimated to be 89% for the explained data

variability. The data from 1885 gauging stations that were used to develop global models do not cover the ETR. It can be concluded that while developing the runoff model, the peculiarities of the ETR runoff formation were not taken into account; however, a model of the river runoff module based on data from the ETR gauging stations was developed using global model predictors (average annual air temperature, average annual precipitation, average slope gradient, and average catchment height). The obtained model explained 73% of the data variability. Despite the comparable qualities of the initial global model and the river runoff model obtained in this study, the global model turned out to be worse in terms of describing the conditions of the ETR runoff formation than that obtained in the current study.

Verification was carried out by comparing the values of the runoff layer in the reference points, i.e., cities represented on both maps. A general visual comparison has been also conducted.

The values of the runoff layer in the reference points located in different landscape zones are generally the same. The size of the runoff layer on the National Atlas map is 300 mm and the forecasted value is 290 mm (Table 7). It can be noticed that the current study results match with the materials above, both in terms of the runoff values and their geographical distributions.

The zonal distribution of the river runoff values across the ETR territory shows a decrease in the discharge per unit area values and the annual river runoff layer from north to south. The discharge per unit area ranges from $0.016 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (690 mm respectively) in the northern part of the tundra zone to $0.0002 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (6 mm respectively) in the southeastern part of the ETR. The total river runoff in the rivers of the plains of the ETR is $0.009 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (296 mm in terms of the annual runoff layer).

For tundra zones, the average value of the discharge per unit area is $0.012 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (296 mm in terms of the annual runoff layer). As for the south of the tundra, in the boreal forest zone, the average river runoff value was found to be $0.009 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$. The drainage basins of the Pechora, Izhma, Tsilma, and Usa are characterized by high values, i.e., about $0.015 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$. High values of the river runoff are also observed in the headwaters of the Onega River, where karst rocks are widespread.

In the broad-leaved forest zones, river runoff values drop to $0.004 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (148 mm respectively). This occurs due to the increased evaporation losses through the higher temperatures and the change of boreal forests to broad-leaved ones, which retain and subsequently evaporate more water. The discharge per unit area ranges from 0.003 to $0.009 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (88–295 mm, respectively).

To the south of the broad-leaved forest zone, in the forest-steppe zone, river runoff values decrease and range from 0.001 to $0.006 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (41–222 mm, respectively). The average value of the river runoff there is $0.003 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (112 mm). The decrease in river runoff values is associated with changes in climatic characteristics, such as a decrease in annual precipitation and an increase in average annual temperature.

The steppe zone of the plains part of the ETR is characterized by river runoff values in the range of 0.0005 to $0.007 \text{ m}^3 \text{ s}^{-1} (\text{km}^2)^{-1}$ (15 to 221

Table 4
Sources of data on natural and anthropogenic conditions of river runoff formation.

Characteristics	Explaining variables {X}	Data Source	Data format
Climatic characteristics	<ol style="list-style-type: none"> 1. Long-term average annual precipitation; 2. Average annual precipitation in May-August (heavy rain season); 3. Average precipitation for cold and warm periods of the year; 4. Coefficient of variation of annual precipitation; 5. Annual average air temperature; 6. Average air temperature in January and July; 7. Sum of active temperatures (sum of average daily temperatures for days when the temperature is above 10 °C); 8. Average long-term highs and lows of annual temperature; 9. Average long-term amplitude of annual temperature; 10. Mean square deviation of annual temperature; 11. Coefficient of annual precipitation variation; 12. Hydrothermic coefficient 	The long-term data from the daily temperature and precipitation observations at weather stations in Russia and the former USSR (FGBU «VNIIGMI-ICDC», http://meteo.ru , Buligina et al.)	ASCII
The terrain conditions	<ol style="list-style-type: none"> 1. Average height; 2. Gradients and exposures of slopes; 3. Profile and planned curvatures; 4. Slope length; 5. Erosion potential 	The Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) with spatial resolution 250 m (Danielson and Gesch, 2011)	Raster data
The geological conditions	Predominant class of pre-Quaternary deposits	The “State geological map of the USSR of pre-Quaternary deposits” at a 1:1,000,000 scale	Raster and vector data
The soil conditions	<ol style="list-style-type: none"> 1. Predominant type of soil; 2. Predominant type of soil-forming rock 	The Unified State Register of Soil Resources of Russia (http://egrpr.esoil.ru/)	Vector data
The structures of land use and anthropogenic developments	<ol style="list-style-type: none"> 1. Percentage of forest cover; 2. Percentage of grassland cover; 3. Percentage of brushwood cover; 4. Percentage of swamp cover; 5. Percentage of arable land 	TerraNorte RLC map of the Russian Terrestrial Ecosystems (the Institute of Space Research of the Russian Academy of Sciences (version 2015) _	Raster data
Landscape zones			

Table 4 (continued)

Characteristics	Explaining variables {X}	Data Source	Data format
	<ol style="list-style-type: none"> 1. Predominant landscape type; 2. Predominant landscape subtype 	The “USSR Landscape Map” at a scale of 1:2,500,000.	Vector data

Table 5
Explaining variables and coefficients of logarithm model (GLM method) for different landscape areas.

Explaining variables	Conversion	Landscape area	Linear model coefficients
Model constants		Boreal forest zone	−5.105
		Broad-leaved forest zone	−5.241
		Forest-steppe zone	−5.411
		Steppe zone	−5.649
Sum of active temperatures (deg. C)		Boreal forest zone	−0.357
		Broad-leaved forest zone	−0.037
		Forest-steppe zone	−0.219
		Steppe zone	−0.300
Annual precipitation (mm)	log	Boreal forest zone	0.098
		Forest-steppe zone	0.290
		Steppe zone	0.085
Percentage of forest cover (%)		Boreal forest zone	0.114
		Broad-leaved forest zone	0.193
Mean slope (deg)	log	Boreal forest zone	0.033
		Broad-leaved forest zone	0.124
		Steppe zone	0.122
Slope lengths (m)	log	Broad-leaved forest zone	−0.141

mm, respectively). The average river runoff value is 0.0017 m³ s^{−1} (km²)^{−1} (54 mm). The decrease in the river runoff values indicates the deterioration of formation conditions. In comparison to the forest-steppe zone, the amount of precipitation in the steppe zone during the warm season would increase the average annual temperature, which in turn would increase the amount of water being evaporated. Furthermore, a significant amount of water is transported across the dense grass cover.

The semi-desert zone of the European territory of Russia is associated with the Caspian lowland. The conditions of runoff formation in this zone can be characterized as extremely unfavorable. The following conditions result in runoff evaporation and filtration, which play a significant role in the water balance: low annual precipitation, high average annual temperatures, plains relief, and water-permeable soils. The average values of the river runoff in this zone are less than 0,001 m³ s^{−1} (km²)^{−1}. The river runoff values decrease to almost zero in the south of this zone.

5. Discussion and perspectives

The modeling and forecasting of the river runoff values in the study were performed for an area of almost 3.5 × 10⁶ m², which includes the entire range of landscapes in the Earth’s humid zone of the northern hemisphere. The obtained models for the river runoff formation, which could be considered statistically reliable, allowed the forecasting of the values of the river runoff for the hydrologically ungauged drainage basins across the studied territory. The forecasted result verification indicates satisfactory qualities for the obtained spatial models, which reflects the zonal nature of river runoff distribution across the territory. These results are similar to the values found in the existing cartographic

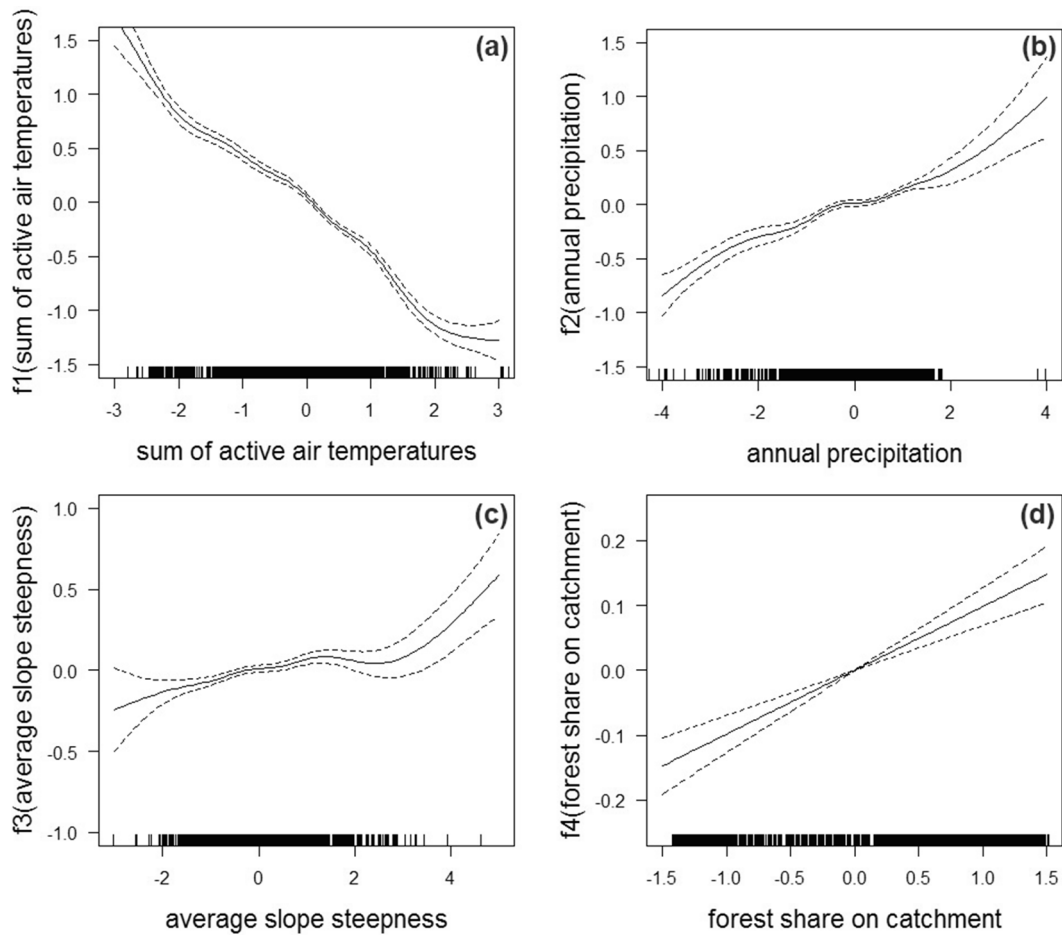


Fig. 5. Individual dependencies of the discharge per unit area on the territory characteristics for the ETR plain zone. (a) - on the sum of active temperatures, (b) - on the annual precipitation, (c) - on the mean slope, (d) - on the percentage of forest cover. Horizontal - standardized value of the predictor, vertical - its contribution to the logarithm model of the river runoff unit (GAM method).

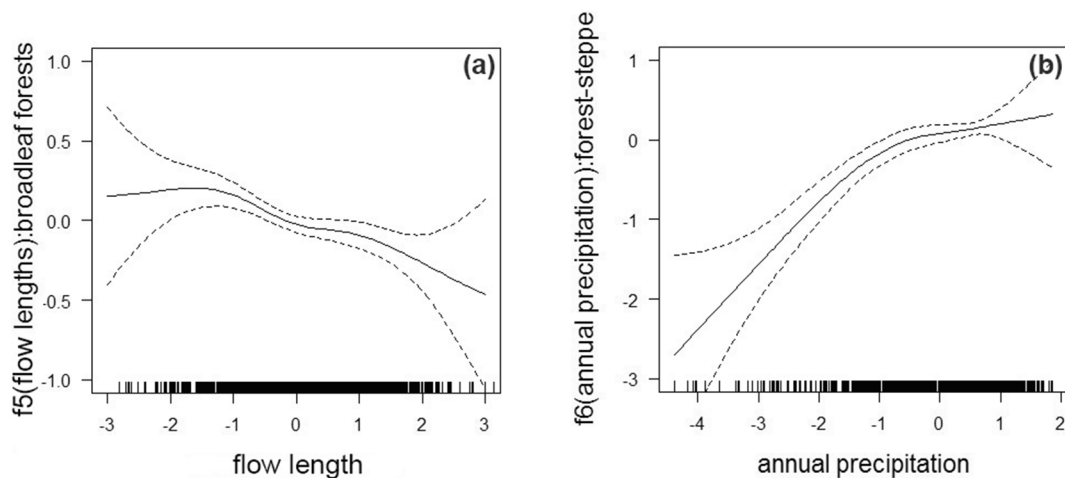


Fig. 6. Examples of individual dependencies of the discharge per unit area on the territory characteristics for certain landscape zones in the ETR plain zone: (a) - from the flow length for the broad-leaved forest zone, (b) - from the annual precipitation for the forest-steppe zone. Horizontal - standardized predictor values, vertical - its contribution to the logarithm model of the river runoff unit (GAM method).

sources. It can be suggested that both the methods used and the results obtained could be used in similar studies for other lowland regions across the world (although primarily for humid plains).

The river runoff formation models have been developed using a multiple regression method. The use of this method imposes statistical

restrictions on the initial data and does not require a mathematical description of the main physical processes of runoff formation in terms of their spatial heterogeneity. A multiple regression method has been applied in this study due to the simplicity of defining parameters and for clear interpretation of the results. Despite this, the method allowed us to

Table 6
Residual statistics of logarithm model (GLM method) for ETR plain zone.

Methods	ME	MD	MAE	MSE	SE
GLM	0	0	0.170	0.063	0.251
GAM	0	0	0.153	0.055	0.235

obtain reliable statistical models that fully reflect the main conditions for the formation of runoff across the study area.

The resulting accuracy remains a debatable topic. River runoff modeling has been carried out according to average long-term data without taking into account the changes that have occurred and the cumulative effects of parameters as climate and land-use.

Therefore, it can be noted that the current study is not aimed at assessing the impact of climate change and land-use patterns in drainage basins on the river runoff. Although these aspects cannot be ignored,

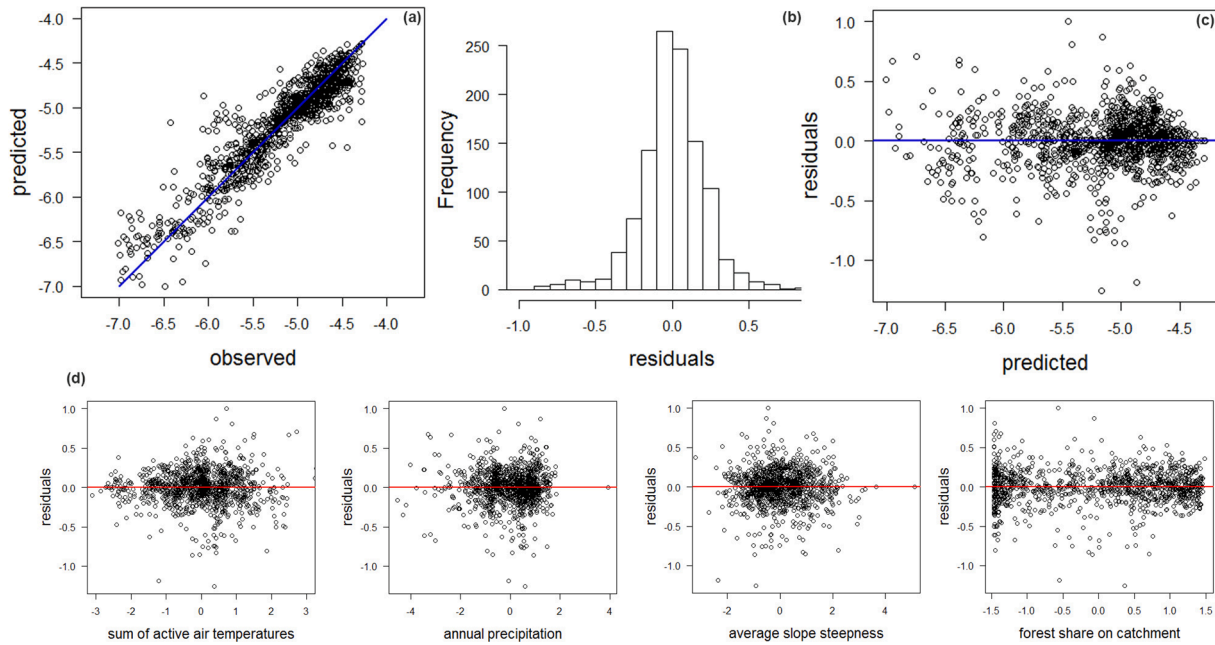


Fig. 7. Diagrams illustrating the quality of constructed models of the discharge per unit for the plain territory of European Russia (GLM method): (a) - horizontally - observed values of the logarithm of the discharge per unit area, vertically - model values (predicted); (b) - histogram of the frequency values of residuals; (c) - horizontally - model values (predicted) of the logarithm of the discharge per unit area, vertically - residuals; (d) - horizontally - model predictors, vertically - residuals.

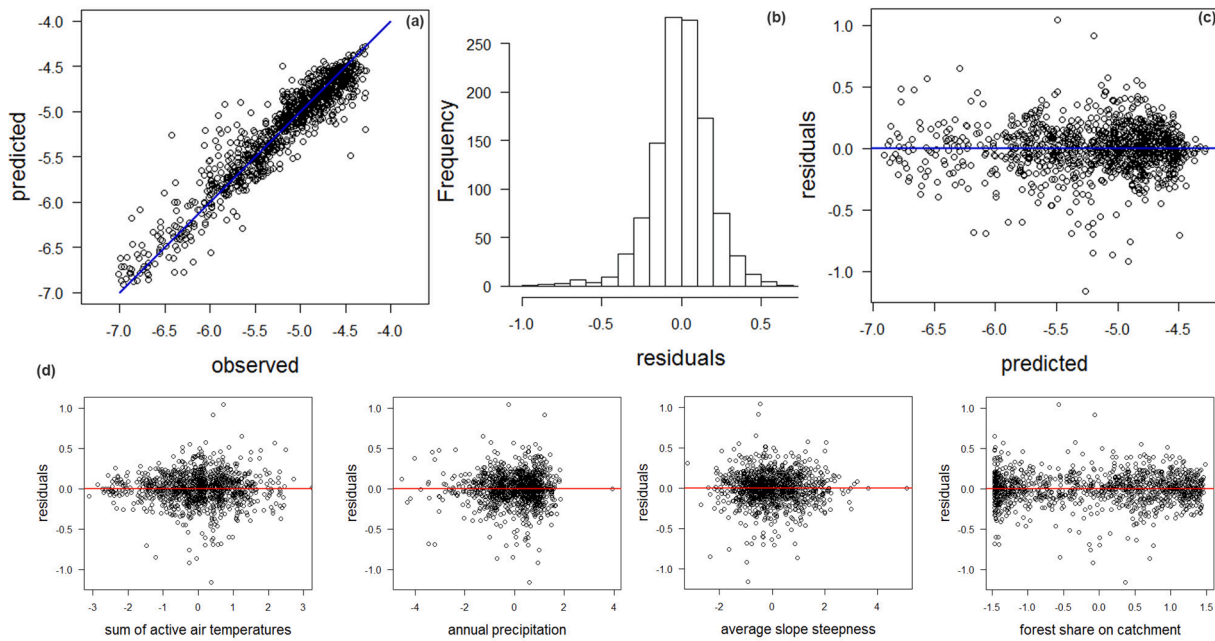


Fig. 8. Diagrams illustrating the quality of constructed models for the ETR plain zone (GAM method): (a) - horizontally - observed values of the logarithm of the discharge per unit area, vertically - model values (predicted); (b) - histogram of the frequency values of residuals; (c) - horizontally - model values (predicted) of the logarithm of the discharge per unit area, vertically - residuals; (d) - horizontally - model predictors, vertically - residuals.

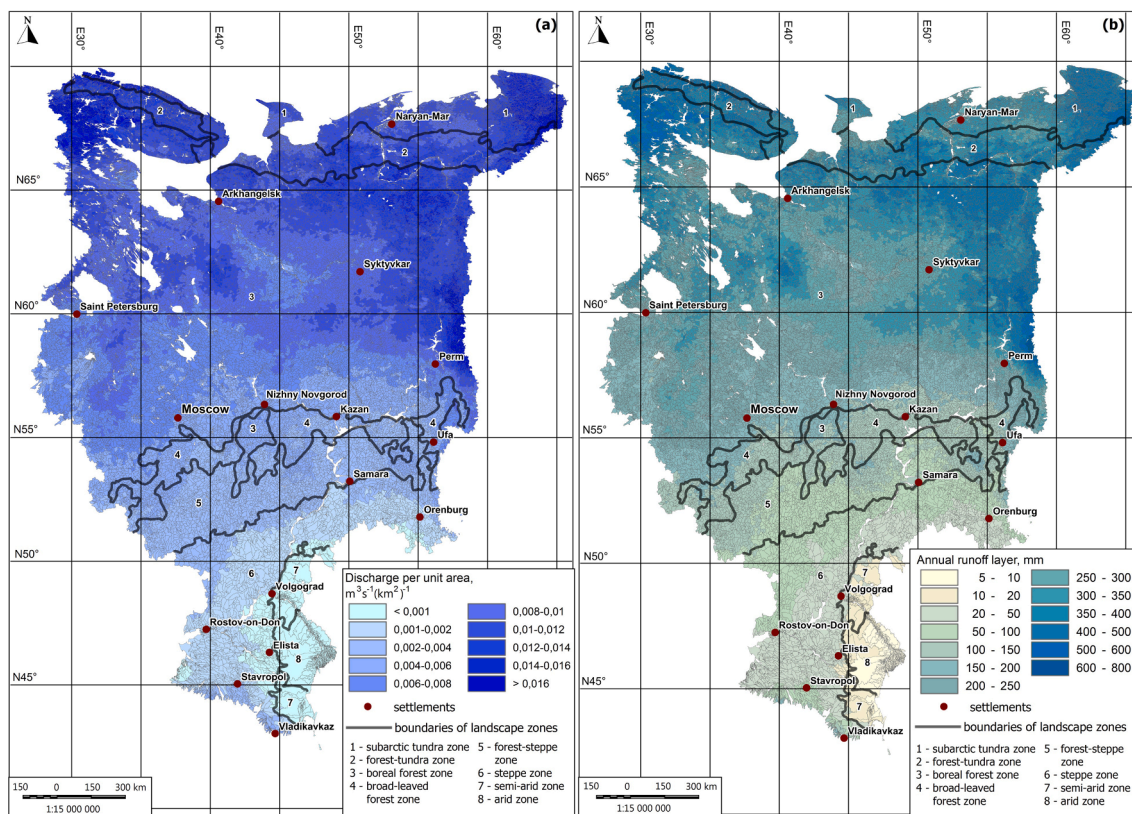


Fig. 9. Cartographic representation of forecast values obtained on the basis of constructed models for ETR drainage basins: (a) - discharge per unit area ($m^3 s^{-1} (km^2)^{-1}$); (b) - annual runoff layer (mm).

Table 7

Comparison of forecast values of the runoff layer and National Atlas of Russia map.

Reference point	Landscape area	Annual runoff layer (Forecasted values)	Annual runoff layer (National Atlas of Russia)
Arhangelsk	Boreal forest zone	290	300
Syktvykar	Boreal forest zone	270	275
Moscow	Boreal forest zone	180	190
Kazan	Broad-leaved forest zone	120	130
Samara	Forest-steppe zone	75	80
Volgograd	Steppe zone	25	30

extensive research is required using other methods of mathematical modeling to assess such changes independently. For such a purpose, conceptual, physical, and-mathematical models with distributed parameters are usually used (Kuchment et al., 2013). Moreover, the ability to solve this type of problem depends on the availability of high-quality information for spatial distribution data about climates and land uses. For example, for the ETR territory, there are no maps of arable land of the same scale for different periods within the 20th century. There are small-scale maps of arable land distributions within administrative units, e.g., data for the USSR and Russian regions can be found; however, this information can be considered as incomparable to the extent of the details used in the current study.

The information around changes in the arable land area within the ETR and the impacts of these changes on river runoff is also very contradictory. The plowing of the ETR across the majority of areas with

fertile soils peaked at the turn of the 19th and 20th centuries. Sharp reductions in the areas of arable land were observed during the First World War, the Russian Revolution (1914–1921), and during the Second World War (1941–1945). During these periods, up to half of the previously plowed land was not cultivated (Ivanova et al., 2019). Data for the river runoff for most gauging stations have been taken after these periods.

Major changes in land use have taken place after the breakup of the Soviet Union. According to the official statistics, the area of arable land has reduced by 11.2% within the period from 1991 to 2014. The largest reductions can be observed across the southern boreal forest zone (19.6%), and, to a lesser extent, across the forest-steppe (11.8%) and steppe zones (6.3%) (Litvin et al., 2017). According to other data sources, the reduction in the arable land area has been 55.6% from 1980 to 2012 in the forest zone, with 28% in the forest-steppe zone and 27.6% in the steppe zone. As a result of economic transformation, 50 million hectares of arable land have been abandoned during this period (Lyuri et al., 2010).

Studies have found that a decrease arable land area has an impact on the amount of surface runoff (Golosoov and Ermolaev, 2019), while the amount of annual runoff varies insignificantly (Kireeva et al., 2017). There have also been changes in the climatic system parameters that determine the runoff, precipitation, and temperature. Several studies have been devoted to the formation of river runoff within the ETR territory about climate changes (Koronkevich, 1990, Koronkevich et al., 2007; Alekseevskij et al., 2013, Frolova et al., 2014; Kireeva et al., 2019).

Currently, it is undoubtedly true that climatic changes have led to the transformation of the annual regime for the majority of rivers across the ETR, i.e., the leveling of the annual runoff hydrographs (Kireeva et al., 2019). An increase in groundwater reserves has occurred due to an increase in the number, duration, and depths of thaws, as well as an increase in meltwater losses in winter. As a result, there is an increase in

the role of low-water periods in the annual runoff, while the high-water periods have decreased. According to the results from two time periods (before 1978 and after 1978), it was found that the annual runoff layer changed slightly, and it was the intra-annual runoff regime that has been most significantly transformed (Frolova et al., 2020).

Based on the studies mentioned above, for the current study, the use of average annual values for the river runoff without taking into account the changes in climate and the nature of land use could be considered justified.

The interpretations of the obtained models show that the formation of river runoff within the study area is mainly influenced by natural factors, i.e., climate indicators that characterize the input and output of the water balance, and the topography and the percentage of forest cover in the drainage basins. As a result of the statistical analysis of the initial data and further modeling, no direct relationship between runoff characteristics and anthropogenic factors was found, in particular, the plowing of river basins and the regulation of rivers (the presence of ponds and reservoirs in them). The influence of anthropogenic factors can be traced in the analysis of the forecasted runoff values, which differ significantly from the observed values at the gauging stations. For one of these gauging stations (Birlya River, Vishenka village – 53° 55' 09" N; 49° 33' 36" E), the value of the flow was overestimated by the model. A detailed examination of the local conditions for the runoff formation in the catchment area of this gauging station revealed a high degree of river regulation and a subsequent flow rate for irrevocable evaporation.

The analysis of foreign studies on river runoff formation indicates the predominant role of natural factors. There are models of the river runoff formation that have been developed for the entire United States and individual regions that could be considered as appropriate examples. In the study (Vogel et al., 1999), regional models of the river runoff have been developed by analyzing the data from 1553 gauging stations and data for the climatic and morphometric characteristics of catchments. High-quality models were obtained using only the catchment area as a predictor; however, it could be noted that the inclusion of climatic and morphometric characteristics in the model could have improved its quality. This study suggests that more accurate models may be obtained for areas with a humid climate. When the climate aridity increases, the model quality deteriorates.

For certain United States regions (West Virginia, the Upper Mississippi Basin), runoff formation models have also been developed based on natural factor data. Data for river runoff and climatic parameters have been averaged for 30 years. For each catchment area, morphometric and soil data were obtained using GIS technologies. As a result of the modeling, the key factors of runoff formation were identified, i.e., the catchment area, the maximum annual temperature, and the gradients of slopes. The quality of the model is estimated at 95–99% in terms of the explained data variability. The influence of the catchment area is not associated with the use of the runoff module as a modulated element, but the water discharge at a given gauging station instead (Tran et al., 2015).

To construct the global runoff model described above (Barbarossa et al., 2017), natural factors of runoff formation were also used, such as the average air temperature, average annual precipitation, average slope gradient, and average catchment height. As a result, a high-quality model was obtained (89% explained data variability).

The studies above indicate the possibility of obtaining territory models within the region and globally that would be based solely on natural factors of flow formation. The use of such models (Vogel et al., 1999) for solving water management problems confirms their high quality.

Nevertheless, the assessment of anthropogenic changes in river runoff data, in particular, the creation of ponds and small reservoirs along watercourses is a primary objective. Currently, there is a vector layer with linear hydraulic structures, i.e., bridges and dams on the ETR. In future studies, it is planned to assess the impact of artificial regulation of watercourses on runoff using remote sensing data.

6. Conclusions

This study has analyzed the patterns of river runoff formation for ETR plain zone. The independent variables included relief morphometric characteristics, climatic indicators reflecting average values, scales, seasonal variations, extreme temperature, and precipitation values, percentages of forest and swamp cover, plowing, percentages of meadows, assessments of the anthropogenic impact on the drainage basin, geographical coordinates of centroids, prevailing soil types, types of soil-forming rock, and classes of pre-Quaternary deposits. These variables were used for modeling. Based on this, statistical models of the runoff dependence on landscape and geographical conditions were obtained and used for analysis. Models of the discharge per unit area for the ETR plain zone have been developed using both linear (GLM) and non-linear methods (GAM) of multiple regression and they explain 85–88% of the data variability. The application of a multiple regression method imposes statistical restrictions on the initial data and does not require a mathematical description of the main physical processes of runoff formation in terms of their spatial heterogeneity. At the same time, this modeling method is distinguished by the simplicity of the parameter determination and the clear interpretation of the results. Sets of significant model factors represent runoff formation conditions in the main landscape zones of the studied areas (boreal forest zone, broad-leaved forest zone, forest-steppe zone, and steppe zone). The obtained trends in the relationships with the dependent variable (discharge per unit area) were interpreted. It was found that the most significant predictors in the model were the annual precipitation, the sum of the active temperatures (characterizing runoff losses on evaporation), average slope gradient, and the forest cover of the catchment. The study shows that the developed statistical models reflect the basic patterns of the river runoff within the studied area. It can be noted that during the statistical analysis of the initial data and modeling, no reliable relationship between the runoff characteristics and the anthropogenic factors (particularly plowed drainage basins and river regulation) was found. Unfortunately, there were no quantitative characteristics for the presence of small ponds and reservoirs in drainage basins. This will require additional study to assess this parameter in the context of river runoff (especially in small river basins) for test basins in each landscape zone using high-resolution images provided by remote sensing. Based on the logarithm model of the discharge per unit area for the ETR plain zone, which was obtained using the GLM method, forecasted values of river runoff indicators for small rivers drainage basins have been calculated. The developed model and its cartographic representations of small river basins accurately reflect the river runoff spatial distribution patterns.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work was performed with the support of the Russian Science Foundation (20-67-46017).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.catena.2021.105327>.

References

- Alekseevskij, N.I., Frolova, N.L., Antonova, M.M., Igonina, M.I., 2013. Assessment of climate change impacts on water regime and river run-off in the Volga river basin. *Water: chemistry and ecology*. 4, 3–12.
- Barbarossa, V., Huijbregts, M., Hendriks, A.J., Beusen, A., Clavreul, J., King, H., Schipper, A.M., 2017. Developing and testing a global-scale regression model to quantify mean annual streamflow. *J. Hydrol.* 544, 479–487.
- Belchikov, V.A., Polunin, A.Y., Simonov, Y.A., Khristoforov, A.V., 2013. Assessment of possible climatic changes in the runoff of the rivers of the Northern Dvina basin in the 21st century. *Meteorol. Hydrol.* 2, 84–92.
- Buligina, O., Razuvayev, V., Aleksandrova, T. Daily Temperature and Precipitation Data for Russia and U.S.S.R. Stations (TTTR). Patent № 2014620942.
- Burgers, (R.) H.E., Schipper, A.M., Hendriks, A.J., 2014. Size relationships of water discharge with catchment area, main-stem length and precipitation. *Hydrolog. Processes*, 28, 5769–5775.
- Di, Chong-Li, Yang, Xiao-Hua, Xia, Xing-Hui, Chen, Xiao-Juan, Li, Jian-Qiang, 2014. Multi-scale modeling of the response of runoff to climate change. *Therm. Sci.* 18 (5), 1511–1516.
- Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey Open-File Report. 26, 2011–1073.
- Dobson, A.J., 2002. *An introduction to generalized linear models*. Chapman & Hall, CRC, New York.
- Dzhamalov, R.G., Safronova, T.I., Telegin, A.A., 2014. Distribution and use of modern water resources of European Russia. *Water resources management in Russia. Legislative regulation and prospects*. 99–111.
- Dzhamalov, R.G., Frolova, N.L., Kireeva, M.B., 2015a. Modern resources of underground and surface waters of the European part of Russia: Formation, distribution, use. GEOS Moscow. 315.
- Dzhamalov, R.G., Frolova, N.L., Kireev, M.B., 2015b. Modern resources of underground and surface waters of the European part of Russia: Formation, distribution, use. GEOS Moscow. 315.
- Limin, Duan, Tingxi, Liu, Xixi, Wang, Luo Yanyun, Wu Long, 2010. Development of a regional regression model for estimating annual runoff in the Hailar river basin of China. *J. Water Resour. Prot.* 2 (11), 934–943.
- Ermolaev, O.P., Maltsev, K.A., Mukharamova, S.S., Kharchenko, S.V., Vedeneeva, E.A., 2017. Cartographic model of river basins of European Russia. *Geogr. Nat. Resour.* 38 (2), 131–138.
- Ivanova, N.N., Golosov V.N., Sidorchuk A.Y., 2019. Dynamics of agricultural development of the southern half of the East European Plain in the XV–XX centuries. Spatio-temporal regularities in the development of modern natural-anthropogenic erosive processes in the Russian Plain. *AN RT, Kazan*, pp. 372.
- Fox, J., Monette, G., 1992. Generalized collinearity diagnostics. *J. Am. Stat. Assoc.* 87, 178–183.
- Frolova, N.L., Agafonova, S.A., Kireeva, M.B., Molchanova, T.G., Povalishnikova, E.S., 2014. Modern features of the intra-annual distribution of river run-off in the Volga basin. *Faculty Geogr., Moscow State University*. 10 (2), 28–39.
- Frolova, N.L., Kireeva, M.B., Kharlamov, M.A., Samsonov, T.E., Entin, A.L., Lurie, I.K., 2020. Mapping the current state and transformation of the water regime of rivers in the European territory of Russia. *Geodesy Cartogr.* <https://doi.org/10.22389/0016-7126-2020-961-7-14-26>.
- Gordeeva, S.M., Malinin, V.N., 2018. On predicting annual runoff of large rivers of European Russia based on decision trees method. *Scientific notes of Russian State Hydrometeorological University*. 50, 53–65.
- Grigoriev, V.Yu., Dzhamalov, R.G., Frolova, N.L., 2018. River run-off of the Oka and Don basins - its dynamics and causes of change. *Hydrological Changes. Questions Geogr.* 145, 194–205.
- Golosov, V., Ermolaev, O., 2019. Spatio-temporal regularities in the development of modern natural-anthropogenic erosive processes in the Russian Plain. *Kazan, AN RT*, p. 372.
- Hastie, T.J., Tibshirani, R.J., 1990. *Generalized Additive Models*. Chapman and Hall/CRC.
- Kireeva, M.B., Ilich, V.P., Frolova, N.L., Kharlamov, M.A., Sazonov, A.A., Mikhailuykova, P.G., 2017. Contribution of climatic and anthropogenic factors to the formation of the low-water period in the Don river basin 2007–2015. *Georisk, Geomarketing publishing center 4* (2017), 10–21.
- Kireeva, M., Frolova, N., Rets, E., Samsonov, T., Entin, A., Kharlamov, M., Telegina, E., Povalishnikova, E., 2019. Evaluating climate and water regime transformation in the European part of Russia using observation and reanalysis data for the 1945–2015 period. *Int. J. River Basin Manage.* <https://doi.org/10.1080/15715124.2019.1695258>.
- Koronkevich, N.I., 1990. Water balance of the Russian Plain and its anthropogenic changes. *Nauka, Moscow*, p. 205.
- Koronkevich, N.I., Barabanova, E.A., Zaitseva, I.S., 2007. The effect of changes in annual values of air temperature and precipitation on the river run-off of the Russian Plain. *Bulletin of the Russian Academy of Sciences. Geograph. Ser.* 5, 64–70.
- Kocherin, D.I., 1927. Average long-term annual and monthly runoff in the European part of the USSR. *Russian University of transport (MIIT)*. 4, 55–94.
- Kuchent, L.S., Gelfan, A.N., Kondratiev, S.A., Lavrov, S.A., 2013. Improving the scientific and methodological base of calculations and forecasts of river run-off based on physical and mathematical models of its formation. VII All-Russian Congress.
- Lindsay, J.B., 2014. The Whitebox Geospatial Analysis Tools project and open-access GIS. The University of Glasgow, UK.
- Litvin, L.F., Kiryukhina, Z.P., Krasnov, S.F., Dobrovolskaya, N.G., 2017. Dynamics of agricultural soil erosion in European Russia. *Eurasian Soil Science*. 50 (11), 1343–1352.
- Lyuri, D.I., Goryachkin, S.V., Karavaeva, N.A., Denisenko, E.A., Nefedova, T.G., 2010. Dynamics of agricultural lands in Russia in the XX century and post-agrogenic restoration of vegetation and soil. *Moscow GEOS*. 2010, 416.
- National Atlas of Russia. Vol. 2 «Nature. Ecology». Moscow: PKO «Cartography», 2007. 495.
- Ning, D., Zhang, M., Ren, S., Hou, Y., Yu, L., Meng, Z., 2017. Predicting hydrological response to forest changes by simple statistical models: the selection of the best indicator of forest changes with a hydrological perspective. *IOP Conf. Series: Earth and Environmental Science* 52, 1–10.
- Overview of hydrometeorological observations, data processing and preparation of information products in 2016, 2017. RFSBO «State Hydrological Institute» (SHI).
- Patel, S., Hardaha, M.K., Seetpal, M.K., Madankar, K.K., 2016. Multiple linear regression model for stream flow estimation of Wainganga river. *Am. J. Water Sci. Eng.* 2 (1), 1–5.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar D., R Core Team, 2014. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1, 117.
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014.
- Reimers, W., 1990. Estimating hydrological parameters from basin characteristics for large semiarid catchments. *Regionalization in Hydrology. IAHS Publ* 191, 187–195.
- Resources of surface waters of the USSR, 1973. T. 11 / Edited by senior research officer of State Hydrological Institute, candidate of geographic sciences N.M. Alyushinskaya. Leningrad, Gidrometeoizdat.
- Sakamoto, Y., Ishiguro, M., Kitagawa, G., 1986. Akaike Information Criterion Statistics. D. Reidel Publishing Company.
- Seber, D., 1980. *Linear Regression Analysis*. Moscow: Mir, 456. (In Russian).
- Terskij, P.N., Zhabkov, K.K., Miheeva, A.I., 2017. The correlation between morphometric characteristics, landscape drivers of flow generation and the characteristics of maximum and mean annual river flow in the Avacha river catchment (Kamchatka). *Researches Aquatic Biolog. Resour. Kamchatka north-west part Pacific Ocean*. 46, 51–65.
- Tran, L.T., O'Neill, R.V., Bruins, R.J.F., Smith, E.R., Harden, C., 2015. Linking land use/land cover with climatic and geomorphologic factors in regional mean annual streamflow models with geospatial regression approach. *Prog. Phys. Geogr.* 39 (2), 258–274.
- Vogel, R.M., Wilson, I., Daly, C., 1999. Regional regression models of annual streamflow for the United States. *J. Irrig. Drain. Eng.* 125 (3), 148–157.
- Voskresensky, K.P., 1962. Norm and variability of the annual flow of rivers of the USSR. *Hydrometeoizdat, Leningrad*, pp. 1–552.
- Wood, S.N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall, London.
- Yermolaev, O.P., Mukharamova, S.S., Maltsev, K.A., Ivanov, M.A., Ermolaeva, P.O., Gayazov, A.I., Mozherin, V.V., Kharchenko, S.V., Marinina, O.A., Lisetskii, F.N., 2018. Geographic Information System and Geoport River basins of the European Russia. *IOP Conference Series: Earth and Environmental Science*. 107(1), Art. № 012108.
- Zaitseva, I.S., Koronkevich, N.I., 2019. Surface water. Big Russian encyclopedia. The electronic version <https://bigenc.ru/geography/text/5564134>.
- Zaykov, B.D., Belinkov, S.Yu., 1937. Average long-term water runoff of the USSR. *State Hydrological Institute* 2.
- Zaykov, B.D., 1946. Average runoff and its distribution in the year on the territory of the USSR. *Federal Serv. Russia Hydrometeorol. Monitor. Environ.* 4 (24), 148.
- Zengin, H., Özcan, M., Değermenci, A.S., Çiğçiz, T., 2017. Effects of some watershed characteristics on water yield in the West Black Sea Region of northern Turkey. *Bosque*. 38 (3), 479–486.
- Zhu, Y., Day, R.L., 2009. Regression modeling of streamflow, baseflow, and runoff using geographic information systems. *J. Environ. Manage.* 90, 946–953.
- Zuur, A.F., Saveliev, A.A., Ieno, E.N., 2012. *Zero Inflated Models and Generalized Linear Mixed Models with R*. Highland Statistics Ltd.
- Zuur, A.F., Saveliev, A.A., Ieno, E.N., 2014. *A Beginner's Guide to Generalized Additive Mixed Models with R*. Highland Statistics Ltd.
- Živković, N., Dragičević, S., Ristić, R., Novković, I., Djurdjić, S., Luković, J., Živković, L., Jovanović, S., 2015. Effects of vegetation on runoff in small river basins in Serbia. *Presenius Environ. Bull.* 24 (6), 1511–1516.