# People and plants - close relationships at the crossroads of the Silk Roads: the case of Tajikistan

Marcin Kotowski, Sebastian Świerszcz, Marcin Nobis, Murodbek Laldjebaev, Barfiya Palavonshanbieva, Arkadiusz Nowak

**Correspondence**

**Marcin Kotowski[1,2], Sebastian Świerszcz[1,3], Marcin Nobis[4], Murodbek Laldjebaev[5], Barfiya Palavonshanbieva[6], Arkadiusz Nowak[1,7]**

[1]Botanical Garden Center for Biological Diversity Conservation in Powsin, Polish Academy of Sciences, Warsaw, Poland
[2]The Franciszek Górski Institute of Plant Physiology, Polish Academy of Sciences, Kraków, Poland
[3]Institute of Agroecology and Plant Production, Wrocław University of Environmental and Life Sciences, Wrocław, Poland
[4]Institute of Botany, Faculty of Biology, Jagiellonian University, Kraków, Poland
[5]Graduate School of Public Policy, Nazarbayev University, Astana, Kazakhstan
[6]Department of Earth and Environmental Sciences, University of Central Asia, Khorog, GBAO, Tajikistan
[7]Botanical Garden of the Wrocław University, Wrocław, Poland

*Corresponding Author: marcin.andrzej.kotowski@gmail.com

**Ethnobotany Research and Applications 27:2 (2024)** - http://dx.doi.org/10.32859/era.27.2.1-19
Manuscript received: 22/01/20xx – Revised manuscript received: 26/01/2024 - Published: 27/01/2024

## Research

## Abstract

*Background:* This study examines the spatial relationship between human populations, livestock and wild useful plants in Tajikistan, a key area along the ancient Silk Roads. It aims to understand how the distribution of these plants correlates with the presence of humans and livestock.

*Methods:* The study uses statistical analyses, including the LSVM model, to assess the distribution of 4269 plant species, of which 1823 are identified as useful. Various factors such as bioclimatic conditions and plant use categories are taken into account.

*Results:* The results indicate a significant correlation between the distribution of useful plants and human population, especially in urbanized areas, which cover 7.4% of Tajikistan. In particular, flora functionality significantly influences human population distribution.

*Conclusions:* The research highlights the importance of spatial relationships between humans and useful flora in population distribution. It suggests that these relationships should be included in models predicting human settlement patterns based on environmental factors.

*Keywords:* useful plants, human population distribution, Tajikistan, Middle Asia, spatial relationship, ethnobotany, support vector machines, supervised learning models.

## Background

From the ancient times, people and plants moved around the world through trade and immigration networks. The most famous and largest of these, the so-called Silk Roads, connected the furthest reaches of East Asia with Europe and Africa, between 2000 BCE and middle of XV century AD coinciding with the fall of the Mongol Empire (Christian 2000). Among other benefits, this connection enabled the trade of previously unobtainable natural products between Asia, Europe and Africa. Almost exactly in the middle of this route, a new culture was born, the descendants of which are still living in the area of present-day Tajikistan (Frye 1963). From the beginning, the culture of Tajiks revolved around local flora which they efficiently utilized to obtain not only the means to survive in very often harsh local conditions, but also an income from selling the products of their local knowledge (Zerjal *et al.* 2002). This knowledge was generally passed on to traders in the form of medicines, dishes, spices or ornaments. They also enabled the spread of plants through the sale or exchange of seeds (Stevens *et al.* 2016).

The written history related to the wild plants used in the area goes as far back as VII - VI centuries BC. The Zoroastrian holy book "Avesta" gives more than 100 names of plant species used for medicinal purposes (Sharopov & Setzer 2018). However, the zenith of early Persian-Tajik herbal medicine was reached during the Samanid Empire (875-999) (Khodadoust *et al.* 2013). Among many scholars, active around this time, there was one whose work had a particular influence on the so-called Western medicine centuries later. Abu Ali ibn Sina, known in the Occident by his Latin name as Avicenna, completed his monumental work in 1025 - the *Canon of Medicine*, in which he described 811 medicines, 612 of which were of plant origin. Among the ingredients used to prepare these medicines were 150 plant species found in the territory of present-day Tajikistan (Rasulova & Junusov 1980).

While the Pamir-Alai's richness and diversity have attracted researchers, the study of its wild useful plants gained attention only after the establishment of the Soviet Union. The first scholar to focus closely on the genetic resources of useful plants was Nikolai Vavilov (1887-1943). As a founder of modern biological sciences in the former Soviet Union, he repeatedly visited Tajikistan conducting scientific expeditions, which allowed him to publish his work *Cultural Flora of Tajikistan in its Past and Future*. Vavilov's investigations and studies led him to classify Tajikistan as a crucial source of crop plant development, a region now often acknowledged as a main area of diversity (Khoury *et al.* 2016, Vavilov 1926).

Of special importance for the recent knowledge of wild species of useful plants of Tajikistan and for the present publication is the ten-volume work "Flora of the Tajik SSR" (Ovchinnikov *et al.* 1957-1991). In total, this work provides data on the local use of 1226 plant species found in Tajikistan. Nevertheless, the number of English-language publications on the ethnobotany of the region is still insufficient. Only a few publications on Tajik ethnobotany have been published in recent years, generally focusing on small areas or ethnic groups (ex. Ali 2021, Ali & Akobirshoeva 2013, Kassam *et al.* 2010, Sõukand *et al.* 2021). In 2020, a group of Polish scientists combined the information provided by Russian scientists in the Flora of Tajikistan with their own phytosociological research, which resulted in the publication of a work entitled Illustrated Flora of Tajikistan (Nowak *et al.* 2020a). The data contained in this work on the occurrence and distribution of Tajik species allowed for an in-depth analysis of the resources of useful plant species found in Tajikistan. Moreover, recent research on the resources of crop wild relatives in Tajikistan has revealed an unprecedented abundance of these species in such a small area (Kotowski *et al.* 2022).

To date, there have been a number of publications addressing the spatial relationship between distribution of human population and biodiversity richness (ex. Luck 2007, Pautasso 2007). However, none of them analyzed these relationships in the context of the richness of useful plants, which is a factor representing the relationship between humans and a particular part of nature. Considering these ancient traditions, this paper aims to determine the spatial relationship between current human and livestock populations and the distribution of useful flora.

Due to the sparse urbanization of Tajikistan, we consider the current distribution of people as a proportional echo of the former population density. In addition, due to the analysis of the past publications on the flora of Tajikistan and current research completed by the staff of the Botanical Garden of the Polish Academy of Sciences, data on the distribution and richness of each individual plant species are derived from an approximately 100-year period of observation. This made it possible to assess the correlation between population density and richness of useful plants with respect to the total richness of the flora of a given sub-region. At the same time, the country's relatively low urbanization and sound species distribution data obtained by long-lasting surveys will allow to avoid drawbacks associated with human activity and anthropogenic changes. Moreover, the obtained biotic and abiotic spatial data will enable creation of statistical models that allow to define the most important determinants of the primary distribution of human populations and associated livestock. Empirically determining the importance of individual predictor will allow to find out whether wild useful plants may have had any influence on the preferences of human settlement location in the past.

## Materials and Methods

**Study area**

***Environmental characteristics***

Tajikistan is located in floristic Tethyan Subkingdom of the Holarctic Kingdom, in its eastern, most continental and arid part - Irano-Tuarnian province. The country covers 143,100 $km^2$ and is a landlocked, mountainous country, with more than 50% of the area located above 3,000 meters above sea level. Only its south-western part is characterized by open lowland areas. Following the bioclimatic classification, the study area can be categorized as a Mediterranean-type macrobioclimate. However, recent studies on the bioclimate of southwestern and central Asia suggest that the Irano-Turanian bioclimatic zone should be distinguished by lower precipitation, longer dry season, higher continentalism and lower winter temperature minima. The annual rainfall in the Western Pamir-Alai exhibits significant variation, with quantities approximately 350 mm in the Zeravshan Mountains, extending up to 600 mm within the Hissar range. Notably, certain locations can experience precipitation levels as high as 2000 mm (Djamali *et al.* 2012). The climate of the Eastern Pamirs is very severe, extremely arid and with strong drying winds. At an altitude of 3,600-4,200 m a.s.l., there are about 30 frost-free days on average per year (Zapryagaeva 1964). In regions at lower altitudes, the typical temperatures span from 23 to 30°C in July, and -1 to 3°C in January. Conversely, in the eastern Pamirs, average temperatures typically drop to a range of 5 to 10°C in July and plummet further to between -15 and -20°C in January. Within the western region of the country, persistent snowfall typically occurs at elevations from 3,500 to 3,600 meters above sea level. while in the eastern regions it is at 5800 m a.s.l. The peak of rainfall in the west of the country is in spring, while in the east it occurs during summer (Latipova 1968, Narzikulov & Stanyukovich 1968, Safarov 2003).

Numerous ridges create an extremely rugged mountain and highland terrain. The exception is the southern part of the country, where relatively gentle slopes, subdued watersheds, wide valleys and foothill plains dominate. This results in a clear vertical zonation of climate, soil and vegetation. Forest and shrubland vegetation is the richest in Central Tajikistan. Extremely rugged terrain, high mountain ranges and deep gorges create a great variety of local microclimate, which in the foothills of the ranges, like the climate of southern Tajikistan, is characterized by long hot and dry summers, and above - cold and very long winters. The lower belts are dominated by sulphurous soils, which above are replaced by brown mountain soils. The highlands of the Eastern Pamirs consists of a system of the highest ridges, wide valleys and drainless depressions. Botanically and geographically, this region has much in common with Tibet (Liu *et al.* 2022, Searle 2013).

***Demographic characteristics***

Tajikistan is the least urbanized of all the former Soviet republics, which means that its distribution is mostly controlled by environmental conditions and external factors rather than infrastructural considerations (UN.ESCAP 2021).Ethnically, Tajiks trace their history directly back to the great Persian empires and civilizations. The Iranian Samanid dynasty-controlled Transoxiana from around 900 AD, with Bukhara as its capital. Together with Samarqand, these great cities were the cultural centres of the empire. Today Samarqand and Bukhara are located in the territory of Uzbekistan. As a distinct indigenous cultural group in the region, Tajiks make up about 84% of the country's population (Agency on Statistics under President of the Republic of Tajikistan 2012). Significant numbers of Tajiks also live in Afghanistan, Uzbekistan, Kyrgyzstan and Xinjiang. The medium-lowland Tajiks are Sunni Muslims living both in small villages practicing subsistence agriculture and urbanized areas such as Dushanbe, Kulob, Bokhtar and Khujand. The Tajiks of the Pamir live in isolated, highly elevated valleys, speak several different Eastern Iranian languages and majority practice Ismailism, a branch of Shia Islam. Most Tajiks still rely on their individual gardening plots to provide food for their people.

***Data collection***

The database used for further analyses in this study was built on data extracted from the 10-volume work "Flora of Tajik SSR". This work was compiled by a multi-author team of scientists (Ovchinnikov *et al.* 1957-1991). The work presents the complete list of plant species occurring within the borders of present-day Tajikistan, which later has been supplemented by Nowak *et al.* (2020a). In order to extract data on reports of local plant use, the authors of this publication carefully analyzed nearly 6,000 pages of text comprising the monumental work of Soviet botanists. According to the information collected, this work contains information on the use of 1225 wild plant species among the Tajik people. To obtain additional information on the useful potential of the flora of Tajikistan, we used databases on the beneficial properties of wild plants such as “Plants For a Future” (PFAF 2022), “Useful Temperate Plants” (Fern 2022) and "KEW Medicinal Plant Names Services" (MPNS 2022) and the most recent list of species contained in the work " Red List of vascular plants of Tajikistan " (Nowak *et al. 2020b)*.

Each of the 4,269 plant species on this list was checked for its useful properties based on the previously mentioned databases. After analyzing the above databases, we obtained information on the potential useful properties of 601 plant species not mentioned in the work "Flora of Tajikistan". Once the species with useful properties were selected, they were assigned to primary use groups:

- food plants (plants used for food purposes both as a meal, as a condiment and for making beverages. All species that contribute to human food needs);
- medicinal plants (all plant species used to improve health and prevent disease);
- forage plants (species harvested and used for livestock feed);
- ornamental plants (plants collected for their aesthetic value, used to decorate interiors and surroundings of living areas);
- industrial plants (plants harvested for commercial purposes, sold to units for subsequent processing and trade);
- domestic use plants (plant species used in households for purposes other than those previously mentioned, e.g. source of energy, tool making, parasite deterrence, room scenting).
-

All taxonomic names adopted were verified with the Taxonomic Name Resolution Service (Boyle et al. 2021).

In order to statistically analyse the relationships between the distribution of useful plants and the distribution of human population living in Tajikistan, we used spatial data provided by CIESIN (2016), which currently has the most accurate databases on population distribution in the studied region. Detailed data on the distribution of livestock during the vegetation period in Tajikistan were obtained from the database provided by Gilbert *et al.* (2018).

In order to use abiotic predictors in developing a model to predict population distribution in Tajikistan, besides data on elevation we also used spatial data on bioclimatic variables. These data were obtained from the CHELSA dataset, with a resolution of approximately 1 km × 1 km (Karger *et al. 2017)*.

***Spatial and correlation analysis***

In order to acquire information on the chorological pattern of wild useful plants belonging to the flora of Tajikistan, we used a system of operational geographical units (OGU) defined by phytogeography and altitude. According to the phytogeographic classification introduced by Grubov (2010), each unit corresponds to a polygon within a specific subregion, (Fig. 1.) and a 100-meter altitude band. This resulted in 808 OGUs with relatively homogeneous environmental parameters within each unit (Raduła *et al.* 2021).

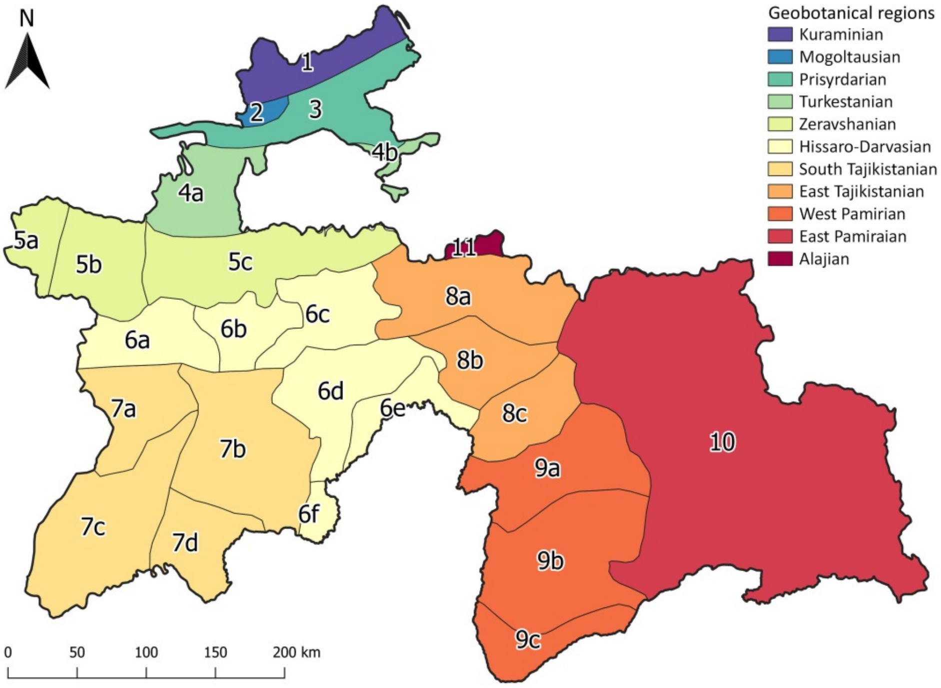


Figure 1. Phytogeographical division of Tajikistan

Using the available literature, based on occurrence/absence data, we compiled all assessed ranges of wild useful plants for each phytogeographic subregion in Tajikistan. We then calculated the potential richness of useful plants for each OGU. Altitudes above 5,100 m a.s.l. in Eastern Tajikistanian A, B and C, Alayan and Zeravshanian C and above 5,700 m a.s.l. in Western Pamirian A, B and C, Eastern Pamirian were omitted because vascular plants do not occur there.

Maps of altitudinal ranges in the phytogeographic sub-regions were derived from a digital elevation model (DEM) with a resolution of approximately 100 m (Jarvis *et al.* 2008). For each OGU, we assigned information about the richness of useful plants and their conservation status based on available data.

Once the distribution of the individual species of useful plants had been determined, it was possible to represent them spatially in terms of the previously determined useful groups (Fig. 1). The data collected also allowed for a visual representation of the distribution of useful vegetation in relation to population distribution (Fig. 2).

During the analyses of the spatial data relationships, we focused on the numerical data corresponding to each polygon delineated by a phytogeographic region and a 100-m elevation band. We also reduced the numerical data by those corresponding to the altitudinal bands located above 5000 m a.s.l. This is due to the occurrence of permanent snow cover above this altitude, which at the same time excludes both the presence of vegetation as well as population and livestock. Accordingly, the data analysis took place on 770 OGUs located at or below 5000 m a.s.l.

The spatial distribution maps, the database for correlation analyses and the spatial means of gravity of the occurrence were created using the software Qgis version 3.20.1.

The correlations occurring between individual spatial data were defined by applying Spearman correlation analysis for non-parametric sets in R version 4.0.5 (R Core Team 2022).

***Predictor importance analysis***

In order to test the importance of the predictors applied in developing a model to determine the distribution of human populations and livestock during vegetation period in Tajikistan, we used a linear SVM (Support Vector Machines) model. This is a model using machine learning based on classification and regression analysis. SVMs work by mapping data into a multidimensional feature space so that data points can be categorized even if the data is not linearly separable. Supervised learning models use a training set to teach the models to produce the desired results, which include inputs and correct outputs. The algorithm measures its accuracy using a loss function and adjusts until the error is sufficiently minimized. The model is trained using labelled sets of data and adjusts its weights until it is set up appropriately. This is done through a process of cross-validation. This helps to identify which predictors have the greatest impact on the model output. The predictor importance plot shows the relative importance of each predictor in model estimation (IBM 2022).

The model was created using IBM SPSS Modeler v. 18.3. Before creating the model, the predictors were screened using the selection tool included in this software. This tool pre-selects the most relevant predictors for predicting a particular outcome while removing those with no significant relationship to changes in the outcome value. This allows for a more efficient model that uses fewer predictors. The predictors used in the model were selected based on the following specifications:

- Maximum percentage of missing values = 70
- Maximum percentage of records in a single category = 90
- Maximum number of categories as a percentage of records = 95
- Minimum coefficient of variation = 0.1
- Minimum standard deviation = 0
-

The regularization type used in the model was L2 which adds the “squared magnitude” of the coefficient as the penalty term to the loss function. The penality parameter (Lambda) was 0.1 and the regression precission (Epsilon) was 0.1.

The spatial compositional data used to create the models were:

Targets:
- Human population distribution size;
- Livestock distribution size.

Predictors:
- All plant species richness;
- All useful plant species richness;
- Altitude;
- Annual Mean Temperature;
- Annual Precipitation;
- Food plant species richness;
- Forage plant species richness;
- Domestic use plant species richness;
- Human population distribution size (only in model predicting livestock distribution size);
- Industrial plant species richness;
- Isothermality;
- Max Temperature of Warmest Month;
- Mean Diurnal Range;
- Mean Temperature of Coldest Quarter;
- Mean Temperature of Driest Quarter;
- Mean Temperature of Warmest Quarter;
- Mean Temperature of Wettest Quarter;
- Medicinal plant species richness;
- Min Temperature of Coldest Month;
- Ornamental plant species richness;
- Precipitation of Coldest Quarter;
- Precipitation of Driest Month;
- Precipitation of Driest Quarter;
- Precipitation of Warmest Quarter;
- Precipitation of Wettest Month;
- Precipitation of Wettest Quarter;
- Precipitation Seasonality;
- Temperature Annual Range;
- Temperature Seasonality.

In summary, data describing two targets and 29 predictors were used to run LSVM models after being assigned to the previously selected 770 OGUs.

## Results

**Spatial richness of useful plants**

After analyzing available databases, of the 4,269 plant species found in Tajikistan, 1,823 were determined to have useful properties. 519 plants found in the flora of Tajikistan were listed in the Plants for a Future database, 547 in the Useful Temperate Plants database, 1,040 in the Kew Medicinal Plants database and 1,225 in the collective work of Russian scientists Flora of Tajik SSR (Ovchinnikov et al. 1957-1991). Among the 1,823 identified wild useful plants, 1,137 are species with medicinal properties, 603 are forage species, 558 food species, 329 industrial species, 300 domestic use species and 429 ornamental species. 16 of the analyzed species exhibit all of the above properties. These species are *Amorpha fruticosa, Artemisia absinthium, Arundo donax, Asparagus officinalis, Cichorium intybus, Cyperus longus, Melilotus albus, Morus alba, Ocimum basilicum, Prunus mahaleb, Phragmites australis, Ricinus communis, Salix turanica, Silybum marianum, Sorghum bicolor, Ulmus pumila.*

The map showing the richness of useful plant species within Tajikistan demonstrates an apparent gradient in the abundance of useful plant species going from east (smallest richness) to west (highest richness) of the country (Fig 2.). This gradient applies to plant species belonging to all categories of usage. We identified the highest potential richness of useful plants in the mid-mountain valleys and floodplains of Tajikistan. Potential richness is understood here as the richness of species resulting from their natural ranges and habitat preferences excluding current impacts resulting from human activities and the distribution of urbanized areas which also includes the presence of alien species.The highest number of wild useful plant species (612) is found in the South Tajikistanian A phytogeographic subregion and in the South Tajikistanian B subregion (575 species), while the lowest species richness of useful plants occurs in Alaian phytogeographic subregion (170 species) and West Pamirian A (210 species) (Fig 2.).

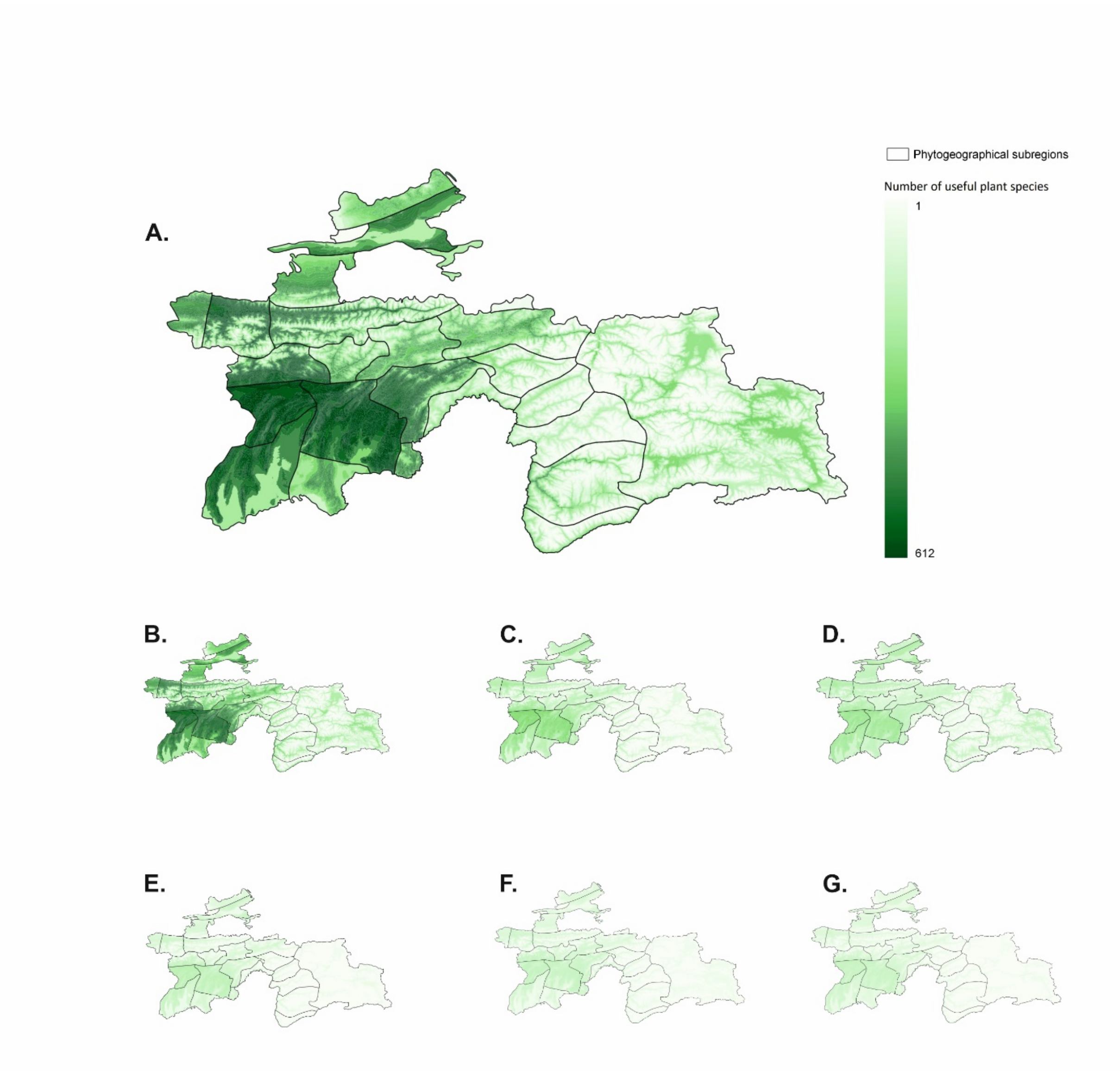


Figure 2**.** The richness of wild species of useful plants with a breakdown by use group. A - all useful plants, B - medicinal plants, C - food plants, D - forage plants, E - industrial plants, F - domestic plants, G - ornamental plants

The spatial means of gravity of the individual variables' occurrence (Fig 3.) show that in relation to the distribution of useful plants, the coordinates of the point representing the mean of gravity of the human population distribution in the area is located closest to the averaged coordinates of the occurrence of the plants used in the household, followed by the group of plants used in industry, food plants, ornamental plants and medicinal plants. The spatial center of gravity of the occurrence of all plant species recorded in Tajikistan is located further than all the aforesaid groups of useful plants. The spatial center of gravity of forage crops is farthest from the spatial center of human population. In close proximity to the spatial center of gravity of the human population are also the centers of gravity of livestock such as cattle, sheep and goats.

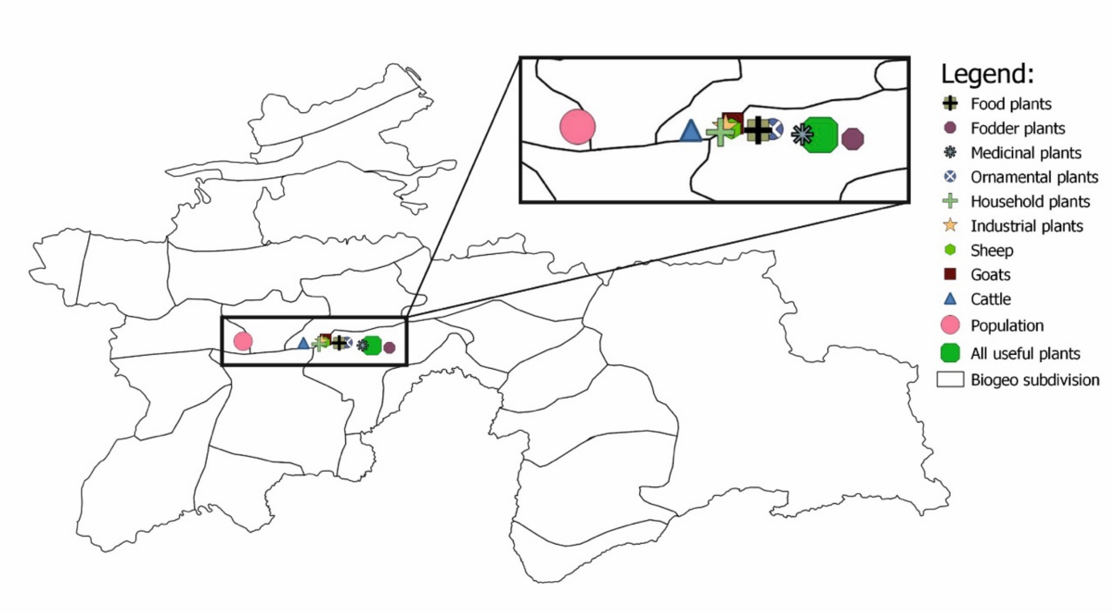


Figure 3. The spatial means of gravity of the occurrence of individual variables

***Human, livestock and plant relationships***

The statistical analysis of the spatial relationship between individual variables showed a strong correlation between all determinants (Fig. 4). The weakest positive correlation has been observed between the distribution of the human population and livestock (r=0.36). Stronger correlations were observed between the richness of useful plant species and the distribution of human population (from r=0.6 for all useful plants to r=0.67 for plants used in household) in comparison to the relationship between the distribution of livestock population and the distribution of useful plants (from r=0.47 for forage species to r=0.51 for all useful species). No significant differences were noted between the degree of correlation regarding various groups of plant use for both human and livestock distribution.

In order to relate the above results to other factors influencing the distribution of the human population, we conducted Spearman correlation analysis for non-parametric sets for spatial data on altitude and bioclimatic variables. The results of the correlation are presented below (Table 1).

Table 1. Spearman's rho between human population distribution and abiotic factors

| | |
|---|---|
| Altitude | -0.67 |
| BIO1 = Annual Mean Temperature | 0.65 |
| BIO6 = Min Temperature of Coldest Month | 0.65 |
| BIO10 = Mean Temperature of Warmest Quarter | 0.65 |
| BIO11 = Mean Temperature of Coldest Quarter | 0.65 |
| BIO5 = Max Temperature of Warmest Month | 0.64 |
| BIO9 = Mean Temperature of Driest Quarter | 0.63 |
| BIO8 = Mean Temperature of Wettest Quarter | 0.60 |
| BIO14 = Precipitation of Driest Month | -0.43 |
| BIO17 = Precipitation of Driest Quarter | -0.42 |
| BIO18 = Precipitation of Warmest Quarter | -0.4 |
| BIO4 = Temperature Seasonality (standard deviation ×100) | 0.31 |
| BIO2 = Mean Diurnal Range (Mean of monthly (max temp - min temp)) | 0.30 |
| BIO7 = Temperature Annual Range (BIO5-BIO6) | 0.27 |
| BIO15 = Precipitation Seasonality (Coefficient of Variation) | 0.25 |
| BIO12 = Annual Precipitation | -0.20 |
| BIO13 = Precipitation of Wettest Month | -0.15 |
| BIO16 = Precipitation of Wettest Quarter | -0.15 |
| BIO3 = Isothermality (BIO2/BIO7) (×100) | 0.14 |
| BIO19 = Precipitation of Coldest Quarter | -0.11 |

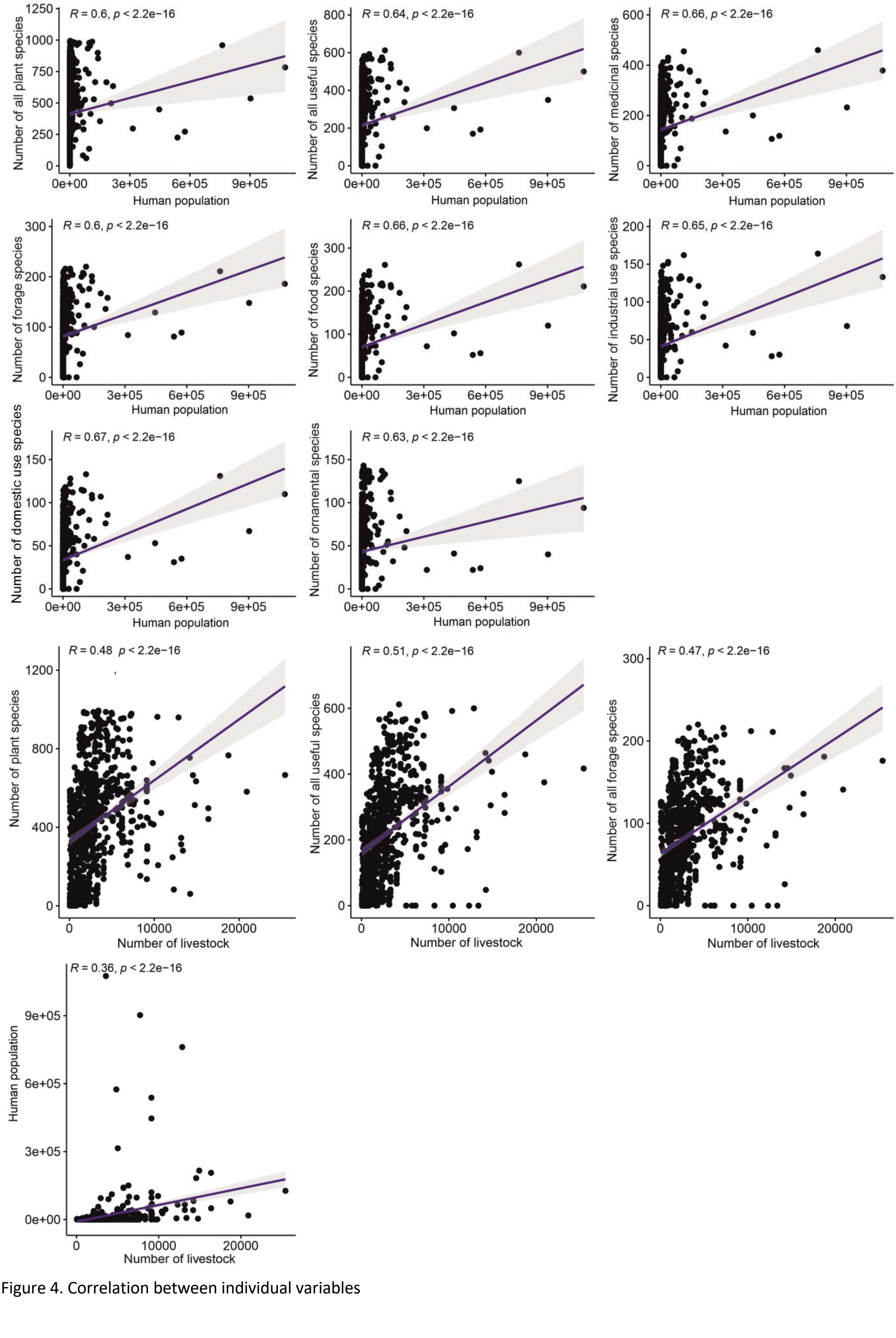


Figure 4. Correlation between individual variables

In this case, the strongest correlations, comparable in value to the correlation between human population distribution and the distribution of useful plants, are found for altitude values (Altitude=-0.67) and temperature values (BIO1=0.65). The values for precipitation are not so meaningful, but it is worth noting that in this case there is a relatively high negative correlation between human population distribution and dry season precipitation (BIO14=-0.43, -0.42).

***Predictors of human population and livestock distribution***

To run the model on the size of the human population distribution, the predictors listed in subsection 2.4 were screened using the feature 'selection tool' in IBM SPSS Modeller based on the target of the human population size distribution. After this step, eight of the 28 predictors that did not show importance for the analyses were removed (Table 2.)

Table 2. Predictor screening results for the human population distribution size model.

| Rank | Field | Measurement | Importance | Value | Screening |
|---|---|---|---|---|---|
| 1 | bio5 | Continuous | important | 1.000 | true |
| 2 | bio10 | Continuous | important | 1.000 | true |
| 3 | bio9 | Continuous | important | 1.000 | true |
| 4 | bio1 | Continuous | important | 1.000 | true |
| 6 | bio6 | Continuous | important | 1.000 | true |
| 7 | bio11 | Continuous | important | 1.000 | true |
| 8 | ALT | Continuous | important | 1.000 | true |
| 9 | bio8 | Continuous | important | 1.000 | true |
| 10 | all_domestic | Continuous | important | 1.000 | true |
| 11 | all_food | Continuous | important | 1.000 | true |
| 12 | all_indust | Continuous | important | 1.000 | true |
| 13 | all_medicinal | Continuous | important | 1.000 | true |
| 14 | all_forage | Continuous | important | 1.000 | true |
| 15 | bio15 | Continuous | important | 1.000 | true |
| 16 | all_use | Continuous | important | 1.000 | true |
| 17 | bio14 | Continuous | important | 1.000 | true |
| 18 | bio17 | Continuous | important | 1.000 | true |
| 19 | bio18 | Continuous | important | 1.000 | true |
| 20 | all_species | Continuous | important | 0.997 | true |
| 21 | all_ornamental | Continuous | important | 0.997 | true |
| 22 | bio12 | Continuous | marginal | 0.941 | false |
| 23 | bio16 | Continuous | uninportant | 0.694 | false |
| 24 | bio13 | Continuous | uninportant | 0.551 | false |
| 25 | bio19 | Continuous | uninportant | 0.319 | false |

| | Measurement | Reason | Screening |
|---|---|---|---|
| bio7 | Continuous | Coefficient of variation below threshold | false |
| bio4 | Continuous | Coefficient of variation below threshold | false |
| bio3 | Continuous | Coefficient of variation below threshold | false |
| bio2 | Continuous | Coefficient of variation below threshold | false |

*all_domestic* - domestic plants, *all_food* - food plants, *all_forage* - forage plants, *all_indust* - industrial plants, *all_medicinal* - medicinal plants, *all_ornamental* - ornamental plants, *all_species* - all plant species, *all_use* - all useful plants, *ALT* - Altitude, *bio1* - Annual Mean Temperature, *bio10* = Mean Temperature Of Warmest Quarter, *bio11* = Mean Temperature Of Coldest Quarter, *bio12* = Annual Precipitation, *bio13* = Precipitation Of Wettest Month, *bio14* = Precipitation Of Driest Month, *bio15* = Precipitation Seasonality (Coefficient Of Variation), *bio16* = Precipitation Of Wettest Quarter, *bio17* = Precipitation Of Driest Quarter, *bio18* = Precipitation Of Warmest Quarter, *bio19* = Precipitation Of Coldest Quarter, *bio2* = Mean Diurnal Range (Mean Of Monthly (Max Temp - Min Temp)), *bio3* = Isothermality (Bio2/Bio7) (×100), *bio4* = Temperature Seasonality (Standard Deviation ×100), *bio5* = Max Temperature Of Warmest Month, *bio6* = Min Temperature Of Coldest Month, *bio7* = Temperature Annual Range (Bio5-Bio6), *bio8* = Mean Temperature Of Wettest Quarter, *bio9* = Mean Temperature Of Driest Quarter.

The predictors screened out as a result of the above analysis are: mean diurnal range, isothermality; temperature seasonality, temperature annual range, mean temperature of wettest quarter, annual precipitation, precipitation of wettest month, precipitation of wettest quarter, precipitation of coldest quarter. These are also the predictors that showed the lowest correlation with the distribution of the human population, shown in Table 2.

The selected predictors were then run through the model to identify the predictors of greatest importance in predicting the size of the human population distribution (Fig. 5, Fig. 6).

According to the LSVM model analysis, the most important predictor of the size of the population distribution is BIO10, which is the average temperature in the warmest quarter of the year. The second most important predictor is all_use, which is the species richness of traditionally used plants. Then the third  is the average temperature in the coldest quarter of the year. The fourth most important predictor relates to the species richness of plants traditionally used as food (all_food) and the fifth to plants used in households (all_househ). Of the five most important predictors, two relate to climatic variables (temperature) and three relate to useful plant richness variables. Other predictors characterized by high correlations with population distribution size such as altitude or overall plant species richness are of marginal importance.

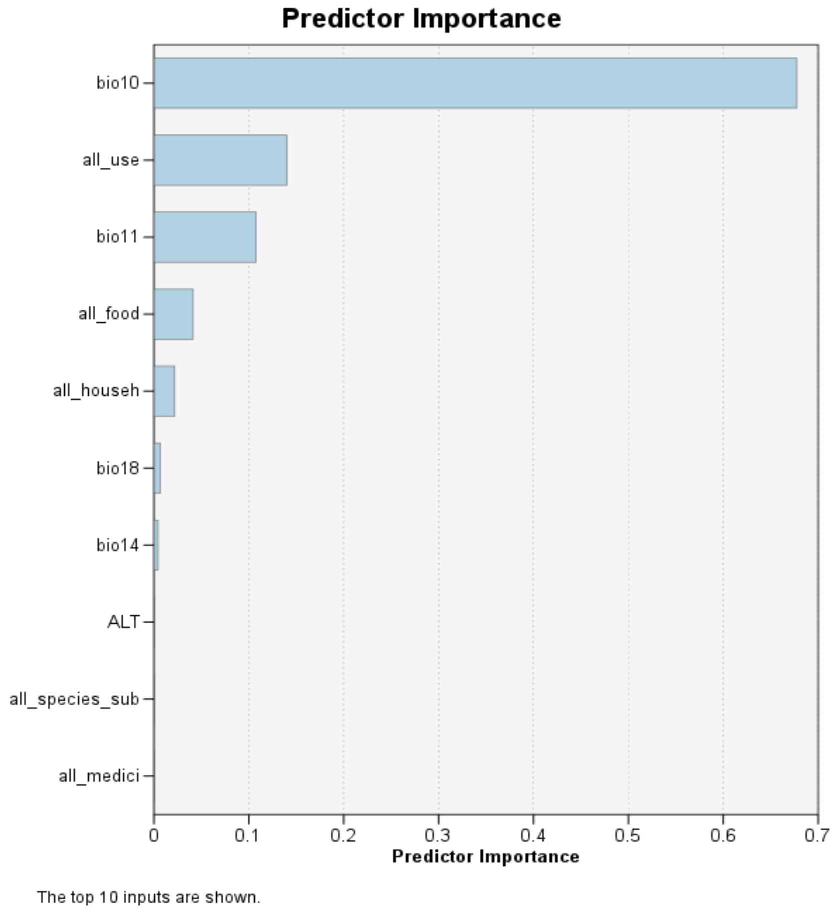


Figure 5. Predictor importance for the human population distribution size model.

According to the 'predicted by observed' model diagram (Fig. 6), the developed model is able to correctly predict the population size for the selected OGU spatial polygons, up to a number of approximately 200000 people. The model is therefore not applicable in highly urbanized areas. For example, this is the case for the polygons corresponding to urban agglomerations such as Dushanbe, where an area of 706 km2 is inhabited by approximately 1075300 citizens (extreme polygon in the graph below). The model therefore fails for nine of the 770 OGU polygons. The total area of such highly urbanized polygons is 11045 km2. This area represents only 7.4% of the total area of the country, which confirms the accuracy of the model.

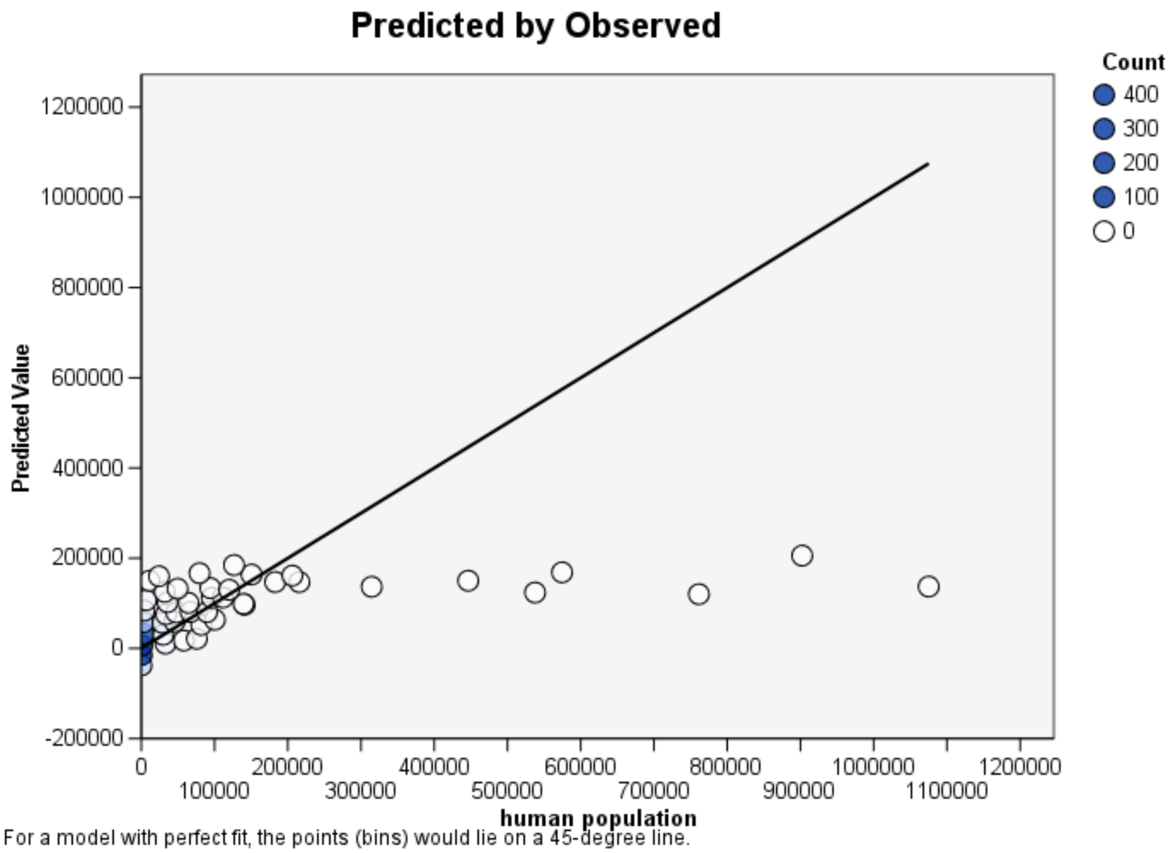


Figure 6. Binned scatterplot of the predicted human population size by the observed values.

Another LSVM model was a model predicting the size of the livestock population distribution during the vegetation period. In this case, we also screened the relevant data. As a target, we set the abundance of livestock assigned to each OGU polygon, and as predictors all the data considered in the previous model as well as the size of the human population distribution. Screening results are available in Table 3.

Table 3. Predictor screening results for the livestock population distribution size model during the vegetation period.

| Rank | Field | Measurement | Importance | Value | Screening |
|---|---|---|---|---|---|
| 1 | bio6 | Continuous | 0 | 1,000 | true |
| 2 | bio9 | Continuous | 0 | 1,000 | true |
| 3 | bio11 | Continuous | 0 | 1,000 | true |
| 4 | bio5 | Continuous | 0 | 1,000 | true |
| 5 | bio1 | Continuous | 0 | 1,000 | true |
| 6 | bio10 | Continuous | 0 | 1,000 | true |
| 7 | bio8 | Continuous | 0 | 1,000 | true |
| 8 | all_househ | Continuous | 0 | 1,000 | true |
| 9 | all_indust | Continuous | 0 | 1,000 | true |
| 10 | all_food | Continuous | 0 | 1,000 | true |
| 11 | all_medici | Continuous | 0 | 1,000 | true |
| 12 | all_use | Continuous | 0 | 1,000 | true |
| 13 | all_forage | Continuous | 0 | 1,000 | true |
| 14 | bio15 | Continuous | 0 | 1,000 | true |
| 15 | all_orname | Continuous | 0 | 1,000 | true |
| 16 | all_species_sub | Continuous | 0 | 1,000 | true |
| 17 | human population | Continuous | 0 | 1,000 | true |
| 19 | ALT | Continuous | 0 | 1,000 | true |
| 20 | bio18 | Continuous | 0 | 1,000 | true |
| 21 | bio14 | Continuous | 0 | 1,000 | true |

| | | | | | |
|---|---|---|---|---|---|
| 22 | bio17 | Continuous | 0 | 1,000 | true |
| 23 | bio16 | Continuous | 0 | 1,000 | true |
| 24 | bio13 | Continuous | 0 | 1,000 | true |
| 25 | bio19 | Continuous | 0 | 1,000 | true |
| 26 | bio12 | Continuous | 0 | 0,952 | true |

| | Measurement | Reason | Screening |
|---|---|---|---|
| bio7 | Continuous | Coefficient of variation below threshold | false |
| bio4 | Continuous | Coefficient of variation below threshold | false |
| bio3 | Continuous | Coefficient of variation below threshold | false |
| bio2 | Continuous | Coefficient of variation below threshold | false |

Accordingly, for this model, only four predictors were screened out. These are temperature annual range, temperature seasonality, isothermality, mean diurnal range.

For the analysis performed for the model (Fig. 7), the most important predictor of livestock presence in the field is BIO6, the min temperature of coldest month. The second most important predictor is the predictor considered most important for a model predicting the size of human population distribution, namely BIO10 - mean temperature of warmest quarter. The third most important predictor is BIO5 - max temperature of warmest month, while the fourth is BIO14 - precipitation of driest month. Subsequently, the fifth most important predictor is the species richness of medicinal plants used by humans. This predictor shows almost identical importance to the sixth predictor determining the species richness of plants used by humans for food. It is worth noting here that the predictor describing the size of the human population distribution does not considerably impact the size of the livestock population distribution during the vegetation period.

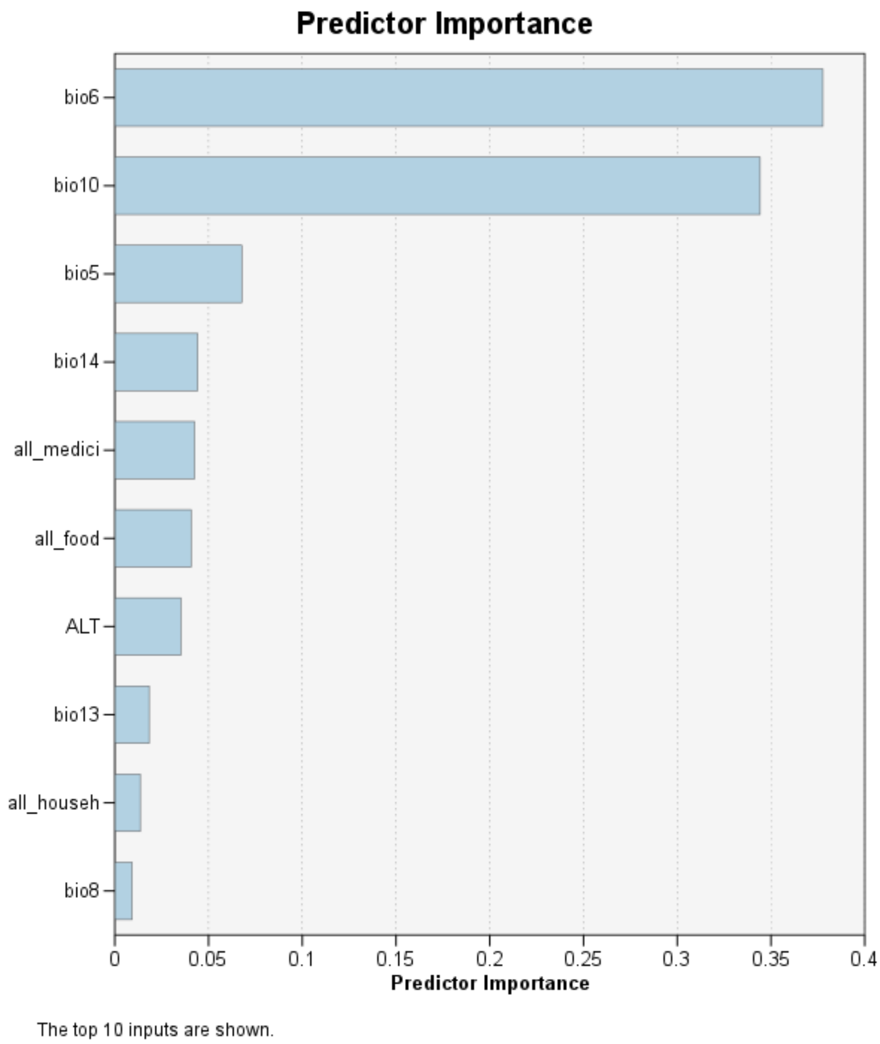


Figure 7. Predictor importance for the livestock distribution size model.

The generated 'predicted by observed' graph describing the present model (Fig. 8) indicates that it is able to correctly predict livestock abundance in the polygons with an assigned abundance that is no greater than about 13,000 individuals. 12 of the 770 polygons that were assigned a higher livestock abundance therefore do not show consistency with the model created (two points on the graph overlap). In total, these 12 polygons represent 4902 km², or 3.4% of the country's area.

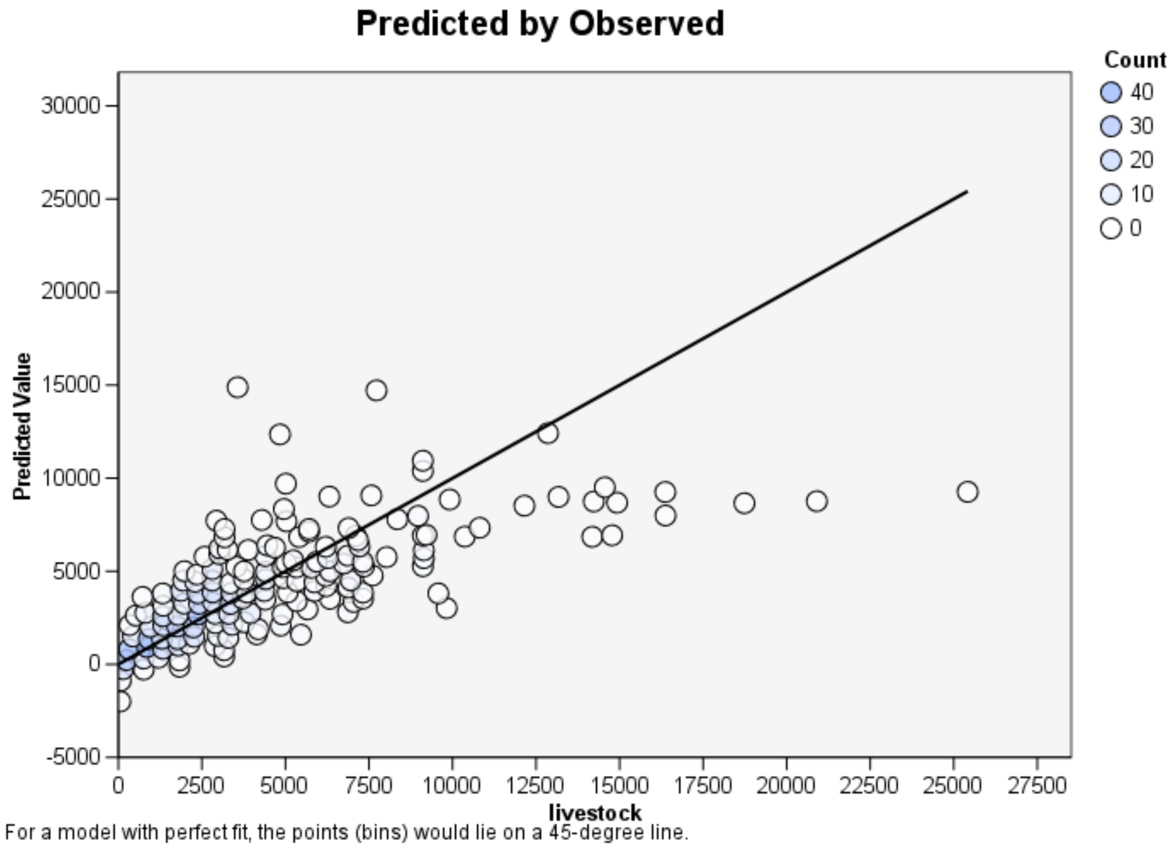


Figure 8. Binned scatterplot of the predicted livestock number by the observed values.

## Discussion

Abundant useful plants in southern Tajikistan's mid-mountain valleys and floodplains match research findings from Tian Shan, Kazakhstan on wild relatives of cultivated plants.

Interpreting the relationship between human and livestock distribution and flora species richness requires considering other key factors influencing their distribution. Here we can refer to argument about people deliberately choosing remote (mountainous) areas to avoid state interference (slavery, conscription, taxes, etc.) (Scott 2010). Undoubtedly, historical and political factors, like forced resettlement during the Soviet period for cotton cultivation, significantly influenced current human and animal distribution. In addition, in the case of livestock farming during the communist regime, there were kolkhozes that focused on seasonal movement to pastures, often using trucks whose location largely depended on the fodder. This was one of the main reasons for the significant decline in the number of animals after the collapse of the Soviet Union, as the supply of fuel and suitable equipment disappeared, and thus the amount of forage produced declined. Post-communism witnessed a return to unsystematic grazing, benefiting our statistical models aiming to minimize excessive human interference (Robinson *et al.* 2010).

The importance of flora, and especially of useful plants for spatial organization of humans is emphasized by the fact that despite the apparent significant relationship between the distribution of human and livestock populations and plant species richness, there is no significant relationship between the distribution of human population and population of grazing livestock during vegetation period (Fig 3.). This is due to the fact that of all the variables analyzed, the only strongly significant factor influencing the distribution both human and livestock populations is the richness of the flora. The results also indicate that human distribution shows a stronger correlation with useful flora presence than overall country's flora. Livestock distribution, however, correlates equally with both total and useful flora richness.

Analysis reveals that not only do useful plants significantly impact human distribution, but also their specific usage groups, like food and medicinal plants. Undoubtedly, the most important observation is that these predictors show far greater importance to the model than the predictors usually typified as most important in defining the location of human settlement (Strano *et al.* 2021, Zhang *et al.* 2014). Examples include the comparatively low importance of altitude, amount of precipitation or total flora richness, despite the fact that there is a significant relationship between the above factors and the size of the population distribution (See Fig. 4., Tab 3.). Significant correlation doesn't always imply predictor importance in models (Lo *et al.* 2015), as multicollinearity—high correlation between independent variables—can cause substantial model interpretation discrepancies (Allen 1997).

In the case of a model designed to predict the spatial distribution of the human population, the total richness of useful plants is a predictor of second importance only to the average temperature in the warmest quarter of the year while showing greater significance than the other 11 temperature-related predictors. The distinctiveness of these groups indicates that they may have individual effects on the distribution of human populations, which is confirmed by the high correlation coefficient.

The results depicting a much higher importance of the richness of useful plants than the total species richness of the flora in predicting the distribution of human populations also disproves the argument of an autocorrelation prevailing between these two predictors. In the case of groups such as industrial plants, ornamental plants, or plants used for fodder, the correlations can only be considered to be due to downsampling of the overall species richness, which at the same time indicates that plants used for this purpose cannot have an individual effect on the distribution of human populations. The positive correlations in this case are only due to the positive correlation shown between population distribution and overall plant species richness.

Another objection to our thesis supporting the importance of useful plants for the distribution of human population distribution may be the causal direction. People inhabiting an area have important reasons to utilize plants, assigning them use value. This argument suggests plants became useful only after population settlement. This argument may be contradicted by data on the percentage of useful plants in relation to altitude. According to the characteristics of the studied factors, human abundance is highest in river valleys and lowlands. The available data shows that the biggest human population in Tajikistan is observed at an altitude of 300 - 1300 meters above sea level (see Supplementary data). According to the argument claiming that human presence 'created' useful plants, the percentage of useful plant species with respect to all species should be highest in areas at these altitudes. However, the least square curve graph showing the percentage of useful plants as a function of altitude (Fig. 9.) shows that these values are highest not only at the lowest altitudes, but also at the highest altitudes where the human population in turn reaches the lowest values.

Analyzing both the model that predicts livestock abundance and the correlation charts, we can see the lack of importance of species richness of plants considered as forage. Temperature-related predictors have the highest significance in this case. Given that the data on the distribution of grazing animals relates only to the vegetation period, that is, when they are not kept in homesteads, the temperature-related predictors may not so much relate to the livestock themselves as to the occurrence of the plants they feed on. It is interesting to note that in creating this model, vegetation recognized by humans as having food and medicinal values was relatively important in predicting the distribution of livestock. This may be due to the fact that plants recognized by humans as forage have favorable characteristics from a human perspective (e.g., ease of access, abundance, rapid weight gain of livestock, etc.), and are not necessarily considered palatable by animals. The lack of increased correlation between the distribution of livestock and the natural occurrence of plant species known by humans as forage species may be also due to outstanding species and spatial richness of plants that benefit livestock, resulting in no apparent relationship between livestock presence and the abundance of plant species known by humans as forage species.The results of this study are therefore able to empirically show that, among other factors, distribution of human population has been related to utility of certain flora species whereas this relationship is not seen in the distribution of livestock populations. While the human population density grows in places with useful plant species within their range, the livestock is grazed in places with vegetation not necessarily useful to humans.

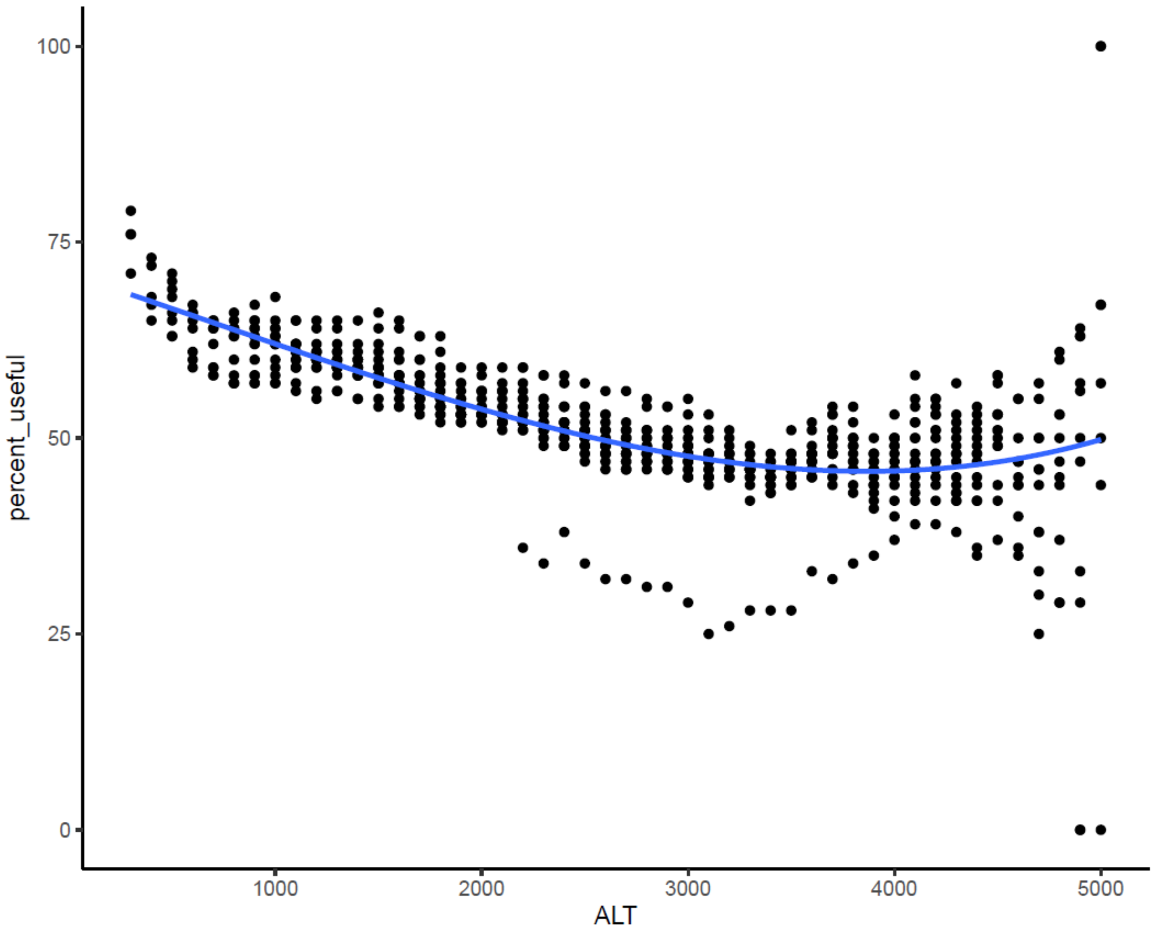


Figure 9. The least square curve graph of the percentage of useful plants along the altitude.

Although the spatial means of gravity of the occurrence concerning human and livestock populations during the vegetation period are located not far apart, no significant spatial correlation was found between the two variables. The reason behind that is the fact that despite these populations are located in close proximity to each other due to mutual coexistence, they rarely occur in the same locations. This is because of the characteristics of the pastoral management according to which, when the human population settles in valleys and areas exhibiting a flat terrain, the animals are grazed in areas with higher elevation and slope (alpine meadows and steppes).

## Conclusion

The spatial analysis presented in this paper proves a statistically significant importance of useful flora as one of the factors determining the primary distribution of human populations. The results of the analyses may provide a useful source of data for creating models predicting the spatial distribution of human populations based on environmental factors alone. On the basis of the analyses presented, we further propose to add to the group of factors already known to determine the distribution of human populations the spatial relationship that prevails between human presence and the occurrence of wild growing useful plants.

## Declarations

**List of Abbreviations**: Not applicable.
**Ethics Approval and Consent to Participate:** Not applicable.
**Consent for Publication**: Not applicable.
**Data Availability:** The datasets generated and/or analyzed during the current study are available as supplementary material and can be provided by the corresponding author upon request.
**Declaration of conflicting interests:** The authors declare that they have no potential conflicts of interest related to the research, writing and publication of this article.
**Funding**: Not applicable.

**Authors' Contributions:** MK contributed to the conception of the study, data collection and analysis, and drafting the initial version of the paper. SŚ was involved in data collection and spatial analysis. MN, ML, BP, and AN assisted in data collection and were actively involved in the research discussions. All authors participated in the creation of the final draft, revisions, and have read and approved the final manuscript for publication.

## Literature cited

Agency on Statistics under President of the Republic of Tajikistan. 2012. Nacional'nyj sostav, vladenie jazykami i graždanstvo naselenija Respubliki Tadžikistan. T. 3. Dushanbe, Tajikistan: Agentstvo po statistike pri Prezidente Respubliki Tadžikistan.

Ali A. 2021. Effect and impact of indigenous knowledge on local biodiversity and social resilience in Pamir region of Tajik and Afghan Badakhshan. Ethnobotany Research and Applications 22:1-26. doi: 10.32859/era.22.03.1-26

Ali A, Akobirshoeva A. 2013. Status and potential use of medicinal and aromatic plants in Pamir Region of Tajik and Afghan Badakhshan. International Research Journal of Plant Science 4(5):34-41.

Allen MP (Ed.). 1997. The problem of multicollinearity. In: Understanding Regression Analysis. Springer US, Boston, MA, USA, pp. 176-180. doi: 10.1007/978-0-585-25657-3_37

Boyle BL et al. 2021. Taxonomic Name Resolution Service, version 5.0. Botanical Information and Ecology Network. Available at: https://tnrs.biendata.org/. Accessed 23 June 2022.

Christian D. 2000. Silk Roads or Steppe Roads? The Silk Roads in World History. Journal of World History 11(1):1-26.

Chukavina AP. 1984. Flora Tadzhikskoi SSR. T. VII. Zontichnye - Verbenovye. Izdatelstvo Nauka, Moscow, Russia.

CIESIN - Columbia University. 2016. Facebook Connectivity Lab and Center for International Earth Science Information Network. High Resolution Settlement Layer (HRSL). Source imagery for HRSL © 2016 DigitalGlobe. Accessed 23 Jun 2022.

Djamali M et al. 2012. Climatic determinism in phytogeographic regionalization: a test from the Irano-Turanian region, SW and Central Asia. Flora - Morphology, Distribution, Functional Ecology of Plants 207:237-249.

Fern K. 2022. Temperate Plants Database. Accessed 23 June 2022. http://temperate.theferns.info/

Frye RN. 1963. The heritage of Persia. World Publishing Company, Cleveland, OH, USA

Gilbert M et al. 2018. Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. Scientific Data 5:180227. doi: 10.1038/sdata.2018.227

Grubov VI. 2010. Schlussbetrachtung zum Florenwerk "Rastenija Central'noj Azii" [Die Pflanzen Zentralasiens] und die Begründung der Eigenständigkeit der mongolischen Flora. Feddes Repertorium 121:7-13.

IBM. 2022. United States. Available at: https://www.ibm.com/us-en. Accessed 23 Feb 2023.

Karger DN et al. 2017. Climatologies at high resolution for the earth's land surface areas. Scientific Data 4:170122. doi: 10.1038/sdata.2017.122

Kassam KA et al. 2010. Medicinal Plant Use and Health Sovereignty: Findings from the Tajik and Afghan Pamirs. Human Ecology 38:817-829. doi: 10.1007/s10745-010-9356-9

Khodadoust K et al. 2013. Dental and oral diseases in Medieval Persia, lessons from Hedayat Akhawayni. Journal of Medical Ethics and History of Medicine 6:9.

Khoury CK et al. 2016. Origins of food crops connect countries worldwide. Proceedings of the Royal Society B: Biological Sciences 283:20160792. doi: 10.1098/rspb.2016.0792

Kinzikaeva GK. 1988. Flora Tadzhikskoi SSR. T. IX. Marenovye - Slozhnotsvetnye. Izdatelstvo Nauka, Moscow, Russia.

Kochkareva TF. 1986. Flora Tadzhikskoi SSR. T. VIII. Kermekovye - Podorozhnikovye. Izdatelstvo Nauka, Moscow, Russia.

Kotowski MA et al. 2022. The primal garden: Tajikistan as a biodiversity hotspot of food crop wild relatives. Agronomy for Sustainable Development 42:112. doi: 10.1007/s13593-022-00846-9

Latipova WA. 1968. Kolichestvo osadkov. In: Narzikulov IK, Stanyukovich KW (eds). Atlas Tajikskoi SSR. Akademia Nauk Tajikskoi SSR, Dushanbe, Tajikistan, pp. 68-69.

Liu J et al. 2022. Name and scale matter: Clarifying the geography of Tibetan Plateau and adjacent mountain regions. Global and Planetary Change 215:103893. doi: 10.1016/j.gloplacha.2022.103893

Lo A, Chernoff H, Zheng T, Lo S-H. 2015. Why significant variables aren't automatically good predictors. Proceedings of the National Academy of Sciences 112:13892-13897. doi: 10.1073/pnas.1518285112

Luck GW. 2007. A review of the relationships between human population density and biodiversity. Biological Reviews 82:607-645. doi: 10.1111/j.1469-185X.2007.00028.x

Medicinal Plant Names Services (MPNS). 2022. Accessed 23 June 2022. https://mpns.science.kew.org/mpns-portal/

Narzikulov IK, Stanyukovich KW. 1968. Atlas Tajikskoi SSR. Akademia Nauk Tajikskoi SSR, Dushanbe, Tajikistan.

Nowak A et al. 2020a. Illustrated flora of Tajikistan and adjacent areas. Polish Academy of Sciences, Botanical Garden Center for Biological Diversity Conservation, Polish Botanical Society, Warsaw; Cracow; Opole, Poland.

Nowak A et al. 2020b. Red List of vascular plants of Tajikistan – the core area of the Mountains of Central Asia global biodiversity hotspot. Scientific Reports 10:6235. doi: 10.1038/s41598-020-63333-9

Ovchinnikov PN. 1957. Flora Tadzhikskoi SSR. T. I, Paprotnikoobraznye - Zlaki. Izdatelstvo Akademii Nauk SSSR, Moscow, Russia.

Ovchinnikov PN. 1963. Flora Tadzhikskoi SSR. T. II, Osokovye - Orkhidnye. Izdatelstvo Akademii Nauk SSSR, Moscow, Russia.

Ovchinnikov PN. 1968. Flora Tadzhikskoi SSR. T. III, Opekhovye - Gvozdichnye. Izdatelstvo Nauka, Moscow, Russia.

Ovchinnikov PN. 1975. Flora Tadzhikskoi SSR. T. IV, Rogolistnikovye - Rozotsvetnye. Izdatelstvo Nauka, Moscow, Russia.

Ovchinnikov PN. 1978. Flora Tadzhikskoi SSR. T. V, Krestotsvetne - Bobovye. Izdatelstvo Nauka, Moscow, Russia.

Ovchinnikov PN. 1981. Flora Tadzhikskoi SSR. T. VI, Bobovye (rod Astragal). Izdatelstvo Nauka, Moscow, Russia.

Pautasso M. 2007. Scale dependence of the correlation between human population presence and vertebrate and plant species richness. Ecology Letters 10:16-24. doi: 10.1111/j.1461-0248.2006.00993.x

Plant for a Future (PFAF). 2022. Database. Accessed 23 June 2022. https://pfaf.org/user/Default.aspx

R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/. Accessed 23 June 2022.

Raduła M et al. 2021. Palaeoclimate has a major effect on the diversity of endemic species in the hotspot of mountain biodiversity in Tajikistan. Scientific Reports 11(1):18684. doi: 10.1038/s41598-021-98027-3

Rasulova MR. 1991. Flora Tadzhikskoi SSR. Slozhnotsvetnye. Izdatelstvo Nauka, Leningrad, Russia.

Robinson S, Safaraliev G, Muzofirshoev N. 2010. Carrying capacity of pasture and fodder resources in the Tajik Pamirs. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy.

Safarov N. 2003. National strategy and action plan on conservation and sustainable use of biodiversity. Governmental Working Group of the Republic of Tajikistan, Dushanbe, Tajikistan.

Scott JC. 2010. The art of not being governed: An anarchist history of upland Southeast Asia. Nus Press, Singapore.

Searle M. 2013. Roof of the World: Tibet, Pamirs. In: Searle M. (Ed.), Colliding Continents: A Geological Exploration of the Himalaya, Karakoram, and Tibet. Oxford University Press, Oxford, UK, pp. 271-301. doi: 10.1093/oso/9780199653003.003.0016

Sharopov F, Setzer WN. 2018. Medicinal Plants of Tajikistan. In: Egamberdieva D, Öztürk M. (Eds.), Vegetation of Central Asia and Environs. Springer International Publishing, Cham, Switzerland, pp. 163-209. doi: 10.1007/978-3-319-99728-5_7

Sõukand R, Prakofjewa J, Pieroni A. 2021. The trauma of no-choice: Wild food ethnobotany in Yaghnobi and Tajik villages, Varzob Valley, Tajikistan. Genetic Resources and Crop Evolution 68:3399-3411. doi: 10.1007/s10722-021-01200-w

Stevens CJ et al. 2016. Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. The Holocene 26:1541-1555. doi: 10.1177/0959683616650268

Strano E et al. 2021. The agglomeration and dispersion dichotomy of human settlements on Earth. Scientific Reports 11:23289. doi: 10.1038/s41598-021-02743-9

UN.ESCAP. 2021. ESCAP population data sheet 2021. United Nations Economic and Social Commission for Asia and the Pacific. Accessed 23 June 2022. https://repository.unescap.org/bitstream/handle/20.500.12870/4098/ESCAP-2021-PB-Population-data-sheet.pdf?sequence=1&isAllowed=y

Vavilov NI. 1926. Centers of origin of cultivated plants. Nauka, Leningrad, Russia.

Zapryagayeva VI. 1964. Dikorastushchiye plodovyye Tadzhikistana. Nauka, Moscow; Leningrad, Russia.

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C. 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. The American Journal of Human Genetics 71:466-482. *doi:* 10.1086/342096

Zhang Z, Xiao R, Shortridge A, Wu J. 2014. Spatial Point Pattern Analysis of Human Settlements and Geographical Associations in Eastern Coastal China — A Case Study. International Journal of Environmental Research and Public Health 11:2818-2833. doi: 10.3390/ijerph110302818