ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ҰЛТТЫҚ ҒЫЛЫМ АКАДЕМИЯСЫ

әл-Фараби атындағы Қазақ ұлттық университетінің

Х А Б А Р Л А Р Ы

ИЗВЕСТИЯ

НАЦИОНАЛЬНОЙ АКАДЕМИИ
НАУК РЕСПУБЛИКИ КАЗАХСТАН
Казахский национальный
университет имени аль-Фараби

N E W S

OF THE ACADEMY OF SCIENCES
OF THE REPUBLIC OF
KAZAKHSTAN
al-Farabi Kazakh National University

PHYSICO-MATHEMATICAL SERIES

1 (345)

JANUARY – MARCH 2023

PUBLISHED SINCE JANUARY 1963

PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

«ҚР ҰҒА Хабарлары. Информатика сериясы».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Меншіктеуші: «Қазақстан Республикасының Ұлттық ғылым академиясы» РҚБ (Алматы қ.). Қазақстан Республикасының Ақпарат және қоғамдық даму министрлігінің Ақпарат комитетінде 14.02.2018 ж. берілген **No 16906-Ж** мерзімдік басылым тіркеуіне қойылу туралы куәлік.
Тақырыптық бағыты: *физика-математикалық сериясы».*
Қазіргі уақытта: *«ақпараттық технологиялар»бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*
Мерзімділігі: *жылына 4 рет.*
Тиражы: *300 дана.*
Редакцияның мекен-жайы: *050010, Алматы қ., Шевченко көш., 28, 218 бөл., тел.: 272-64-39*
http://www.physico-mathematical.kz/index.php/en/

Типографияның мекен-жайы: «Аруна» ЖК, Алматы қ., Мұратбаев көш., 75.

«Известия НАН РК. Серия информатики».
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Собственник: *Республиканское общественное объединение «Национальная академия наук Республики Казахстан» (г. Алматы).*

Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и общественного развития Республики Казахстан **No 16906-Ж** выданное 14.02.2018 г.

Тематическая направленность: *серия физика-математическая.*

В настоящее время: *вошел в список журналов, рекомендованных ККСОН МОН РК по направлению «информационные коммуникационные технологии».*

Периодичность: *4 раз в год.*

Тираж: *300 экземпляров.*

Адрес редакции: *050010, г. Алматы, ул. Шевченко, 28, оф. 218, тел.: 272-64-39*
http://www.physico-mathematical.kz/index.php/en/

Адрес типографии: ИП «Аруна», г. Алматы, ул. Муратбаева, 75.

News of the National Academy of Sciences of the Republic of Kazakhstan.
Series of informatics.
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

Owner: RPA «National Academy of Sciences of the Republic of Kazakhstan» (Almaty). The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Social Development of the Republic of Kazakhstan **No. 16906-Ж,** issued 14.02.2018
Thematic scope: *series physical-mathematical series.*
Currently: *included in the list of journals recommended by the CCSES MES RK in the direction of «information and communication technologies».*
Periodicity: *4 times a year.*
Circulation: *300 copies.*
Editorial address: *28, Shevchenko str., of. 218, Almaty, 050010, tel. 272-64-39*
http://www.physico-mathematical.kz/index.php/en/

Address of printing house: ST «Aruna», 75, Muratbayev str, Almaty.

NEWS OF THE NATIONAL ACADEMY OF SCIENCES
OF THE REPUBLIC OF KAZAKHSTAN
PHYSICO-MATHEMATICAL SERIES
ISSN 1991-346X
Volume 1, Namber 345 (2023), 136-153
https://doi.org/10.32014/2023.2518-1726.174

UDC 004.65, 910.27

© K. Kyrgyzbay*, E. Kakimzhanov, 2023
Al Farabi Kazakh National University, Almaty, Kazakhstan.
E-mail: kyrgyzbay.kudaibergen@gmail.com

CREATION OF A GEODATABASE OF ALMATY REGION BASED ON GIS TECHNOLOGIES

K. Kyrgyzbay — Doctoral student of the Department of Cartography and Geoinformatics. Faculty of Geography and Nature Management. Al Farabi Kazakh National University. 050040. Almaty, Kazakhstan.
E-mail: kyrgyzbay.kudaibergen@gmail.com, https://orcid.org/0000-0002-4279-6436;
E. Kakimzhanov — PhD. Department of Cartography and Geoinformatics. Faculty of Geography and Nature Management. Al Farabi Kazakh National University. 050040. Almaty, Kazakhstan.
E-mail: erkinkakimzhanov@gmail.com, https://orcid.org/0000-0001-6454-681X.

Abstract. This study is devoted to the creation of a geodatabase of the Almaty region for the unification and collection of spatial information. The purpose of the work is to study the methods, data and tools used in determining the architecture and structure of the spatial database. To achieve the objectives, the characteristics of open data were considered, as well as the functionality of software designed for analysis and modeling in a GIS environment. The information in the geodatabase was structured and compared based on the principle of similarity and hierarchical subordination. The analysis of geospatial information was performed in the software of ArcGIS 10.6 and QGIS 3.22. Data analysis and modeling took place moving from the local scale to the country level. MODIS satellite images were used at the country and regional level, multispectral raster data from Landsat 8 for visual investigation of the Almaty region and Sentinel 2A/2B images for analysis at the level of districts and administrative units. Spatial analysis and GIS processing processes are performed at each level of the study. The geodatabase is divided into spatial data classes. These include the administrative division class, hydrographic class, digital terrain model class, agricultural data class and raster data. Administrative-territorial information is based on the division of the Almaty region into districts and cities of regional significance. Man-made objects, including roads, railways and settlements, were imported from the Open Street map (OSM) server, which are generally accessible to the GIS community. Hydrographic objects were analyzed and processed using remote sensing data and digital terrain models. The agricultural class is divided into a layer of agricultural land and soil cover data. Complex climate information consists of raster files and vector multilinear shapefiles. A digital relief model and modifications of this model, in the form of data on the slope steepness and exposure, were created during the processing of SRTM radar topographic images. As a result of the research, a geodatabase of the Almaty region was created with a volume of more than 50 GB of RAM and consisting of classes of spatial, raster and statistical data. Each geospatial object has its own attribute information or pixel values for raster files. The information obtained will be the basis for further research related to the irrigated agricultural lands of the territory.

Key words: GIS technologies, geodatabases, remote sensing, spatial data classes

© Қ.Т. Қырғызбай*, Е.Х. Какимжанов, 2023
Әл-Фараби атындағы Қазақ Ұлттық университеті, Алматы, Қазақстан.
E-mail: kyrgyzbay.kudaibergen@gmail.com

ГАЖ ТЕХНОЛОГИЯЛАРЫ НЕГІЗІНДЕ АЛМАТЫ ОБЛЫСЫНЫҢ ГЕОДЕРЕКТЕР БАЗАСЫН ҚҰРУ

Қырғызбай Қ.Т. — Докторант. Картография және геоинформатика кафедрасы. География және табиғатты пайдалану факультеті. Әл-Фараби атындағы Қазақ Ұлттық университеті. 050040. Алматы, Қазақстан.
E-mail: kyrgyzbay.kudaibergen@gmail.com, https://orcid.org/0000-0002-4279-6436;
Какимжанов Е.Х. — PhD. Картография және геоинформатика кафедрасы. География және табиғатты пайдалану факультеті. Әл-Фараби атындағы Қазақ Ұлттық университеті. 050040. Алматы, Қазақстан.
E-mail: erkinkakimzhanov@gmail.com, https://orcid.org/0000-0001-6454-681X.

Аннотация. Берілген зерттеу жұмысы кеңістіктік ақпаратты біріздендіру және сақтау үшін Алматы облысының геодеректер базасын құруға арналған. Жұмыстың мақсаты-кеңістіктік мәліметтер базасының архитектурасы мен құрылымын анықтауда қолданылатын әдістерді, деректерді және құралдарды зерттеу. Тапсырмалардың толық орындалуына қол жеткізу үшін ашық деректердің сипаттамалары, сондай-ақ ГАЖ ортасында талдау мен модельдеуге арналған бағдарламалық жасақтаманың функционалдығы қарастырылды. Геодеректер базасындағы ақпарат ұқсастық және иерархиялық бағыну принципі негізінде құрылымдалған және салыстырылған. Геокеңістіктік ақпаратты талдау ArcGIS 10.6 және QGIS 3.22 бағдарламалық жасақтамаларында жүргізілді. Деректерді талдау және модельдеу жергілікті масштабтан мемлекет деңгейіне дейінгі аралықта жүзеге асырылды. Мемлекет және облыс деңгейіндегі ақпаратты талдау MODIS ғарыштық суреттері арқылы зерттелді, сонымен қатар Алматы облысын визуалды зерттеу үшін Landsat 8 мультиспектрлі растрлық деректері және аудандар мен әкімшілік бірліктер деңгейінде талдау жүргізу үшін Sentinel 2A/2B суреттері пайдаланылды. Зерттеудің әр деңгейінде кеңістіктік талдау және геоақпараттық өңдеу процестері орындалды. Зерттеу барысында геодеректер базасын кеңістіктік деректер кластарына бөлу жүзеге асырылды. Берілген жіктеу жұмысы әкімшілік-аумақтық бөліну, гидрографиялық, цифрлық рельеф моделі, ауылшаруашылық деректер кластары мен растрлық деректерді қамтыды. Әкімшілік-аумақтық ақпарат Алматы облысын аудандарға және облыстық маңызы бар қалаларға бөлуге негізделген. Антропогендік нысандар, соның ішінде жолдар, теміржолдар және елді мекендер ГАЖ ортасына қолжетімді Open street map (OSM) серверінен импортталды. Гидрографиялық объектілер Жерді қашықтықтан зерделеу деректері мен сандық рельеф моделін қолдану арқылы талданып, өңделді. Ауылшаруашылық класы ауылшаруашылық жерлері мен топырақ жамылғысы туралы мәліметтер қабатына жіктелді. Кешенді климаттық ақпарат растрлық файл және векторлық изосызықтардан тұрады. Жер бедерінің көлбеу және экспозиция туралы мәліметтерін қамтитын жердің сандық моделі мен оның модификациялары SRTM радиолокациялық топографиялық суреттерін өңдеу негізінде құрастырылған. Зерттеу нәтижесінде компьютер жадындағы көлемі 50 ГБ-тан астам кеңістіктік, растрлық және статистикалық деректер кластарынан тұратын Алматы облысының геодеректер базасы құрылды. Әрбір геокеңістіктік объектінің өзіндік атрибуттық ақпараты немесе растрлық файлдар үшін пиксель мәндері қамтылған. Алынған ақпарат аумақтың суармалы егістік жерлерін кейінгі зерттеулерінде негіз болады.

Түйін сөздер: ГАЖ технологиялары, геодеректер базасы, жерді қашықтықтан зерделеу, кеңістіктік деректер кластары.

© Қ.Т. Қырғызбай*, Е.Х. Какимжанов, 2023
Казахский национальный университет имени аль-Фараби,
Алматы, Казахстан.
E-mail: kyrgyzbay.kudaibergen@gmail.com

СОЗДАНИЕ БАЗЫ ГЕОДАННЫХ АЛМАТИНСКОЙ ОБЛАСТИ НА ОСНОВЕ ГИС-ТЕХНОЛОГИЙ

Қырғызбай Қ.Т. — Докторант, кафедра Картографии и геоинформатики, факультет Географии и природопользования. Казахский национальный университет имени аль-Фараби. 050040. Алматы, Казахстан.
E-mail: kyrgyzbay.kudaibergen@gmail.com, https://orcid.org/0000-0002-4279-6436;
Какимжанов Е.Х. — PhD, кафедра Картографии и геоинформатики, факультет Географии и природопользования, Казахский национальный университет имени аль-Фараби. 050040. Алматы, Казахстан.
E-mail: erkinkakimzhanov@gmail.com, https://orcid.org/0000-0001-6454-681X.

Аннотация. Данное исследование посвящено созданию базы геоданных Алматинской области для унифицирования и сбора пространственной информации. Целью работы является исследование методов, данных и инструментов, использующихся при определении архитектуры и структуры базы пространственных данных. Для достижения поставленных задач были рассмотрены характеристики открытых данных, а также функционал программных обеспечений, предназначенных для анализа и моделирования в ГИС-среде. Информация в базе геоданных была структурирована и сопоставлена на основе принципа сходства и иерархической соподчиненности. Анализ геопространтсвенной информации выполнялся в программных обеспечениях ArcGIS 10.6 и QGIS 3.22. Анализ и моделирование данных происходили переходя от локального масштаба до уровня страны. Были использованы космические снимки MODIS на уровне страны и области, мультиспектральные растровые данные Landsat 8 для визуального исследования Алматинской области и снимки Sentinel 2A/2B для проведения анализа на уровне районов и административных единиц. На каждом уровне исследования выполняются процессы пространсвенного анализа и ГИС-обработки. База геоданных разделена на классы пространственных данных. К ним относятся класс административно-территориального деления, гидрографический класс, класс цифровой модели рельефа, класс сельскохозяйственных данных и растровые данные. Административно-территориальная информация основана на разделении Алматинской области на районы и города областного значения. Антропогенные объекты, в том числе дороги, железные дороги и населенные пункты импортированы с сервера Open street map (OSM), которые являются общедоступными для ГИС-сообщества. Гидрографические объекты анализированы и обработаны с применением данных дистанционного зондирования Земли и цифровой модели рельефа. Сельскохозяйственный класс подразделяется на слой сельскохозяйственных угодий и данных о почвенном покрове. Комплексно-климатическая информация состоит из растровых файлов и векторных полилинейных шейп-файлов. Цифровая модель рельефа и модификации данной модели в виде данных о крутизне склона и экспозиции созданы в процессе обработки радарных топографических снимков SRTM. В результате исследования создана база геоданных Алматинской области с объемом более 50 ГБ оперативно памяти и состоящий из классов пространственных, растровых и статистических данных. В каждом геопространственном объекте имеется собственная атрибутивная информация или пиксельные значения для растровых файлов. Полученная информация будет основой для дальнейших исследований, связанных с орошаемыми сельскохозяйственными угодьями территории.

Ключевые слова: ГИС-технологии, базы геоданных (БГД), дистанционное зондирование Земли, классы пространственных данных

Introduction

GIS is a tool for storing, analyzing, and displaying spatial information. Many aspects of life, including social, economic, geographical, computer science, information systems, education, healthcare are closely interrelated with GIS technologies. A special place in the planning of territorial units is occupied by the analysis and modeling of spatial data.

The technology of the geodatabase can be a potentially more effective solution to the ongoing problem of efficient storage and retrieval of spatial data. The geodatabase allows you to integrate various sources of information, and also supports management, storage and analysis. The geodatabase supports spatial and non-spatial data: vector, raster and attribute data, as well as time data and time series (Nur et al., 2018).

For a large amount of spatial and geographical information, a centralized repository is needed, where a convenient and unified environment is created in the form of a geodatabase. Vector objects (geographical objects with vector geometry) are versatile and are frequently used geographical data types, well suited for representing objects with discrete boundaries. A spatial object is an object that stores its geographical representation, usually represented as a point, line, or polygon, as one of the properties (fields) in a row.

A lot of research has been done using GIS and a geodatabase for different purposes. Works of Nur et al and Zhao Ming devoted to research in the field of natural risks using GIS for spatial planning (Nur et al., 2018; Zhao et al., 2017). Vulnerability to natural risk was assessed in a study by Rahman et al. (Rahman et al., 2015).

The use of GIS and geodata in agriculture, including in determining suitable crops for a number of characteristics, was considered in the Senanayake`s study (Senanayake et al., 2017). The creation of a geodatabase based on the ArcGIS program for storing soil data when monitoring water bodies in the United States is considered in the work Luzio et al. (Luzio et al., 2017)

The main purpose of this article is to create a geodatabase of the Almaty region using remote sensing and GIS. To achieve this goal, remote sensing data and vector shapefiles were selected and processed. Research methods and tools have been selected.

Materials and methods

Almaty region (current Almaty and Zhetysu regions, formed on June 8 , 2022) is located in the southeastern part of Kazakhstan, between 42^0 20’ and 47^0 12’ north latitude, 73^0 47’ and 82^0 34’ east longitude. The area of Almaty region is 223,911 km^2. It is divided into 17 districts and 3 cities of regional significance on an administrative-territorial basis (Kyrgyzbay et al., 2022).

The initial stages of creating a geodata database include the definition of the relevant data and the goal of unification. For the Almaty region, the data selected classes of administrative-territorial, hydrographic, agricultural, climatic geospatial information, as well as a digital relief model and its subsequent modifications. The purpose of the unification of spatial information is the subsequent use of ordered data for the zoning of the territory according to the hydromodule feature, for the implementation of which complex and reliable spatial objects of agricultural direction are required.

Open access vector files. Vector formats of spatial data form the basis of the geodatabase of the Almaty region. Open data is imported from the site https://download.geofabrik.de/, which uses Open Street Map (OSM) data. This Web service provides an opportunity to download free shape files of settlements, roads, railways and the administrative-territorial division of the country at different levels.

Remote sensing data. The creation of digital terrain models, the determination of the last contours of polygonal water bodies (lakes and reservoirs) and rivers and the extraction of complex climatic information are carried out on the basis of remote sensing data or combined methods.

Digital terrain models and their modifications (exposure and elevation steepness) are created based on SRTM (Shuttle Radar Topography Mission) images. The images are publicly available on the USGS Earth Explorer website (https://earthexplorer.usgs.gov/). The difference in absolute heights in pixels provides the possibility of spatial determination of river terraces, subsequently automatic vectorization of rivers.

The USGS Earth Explorer Data Portal is a universal resource for obtaining geospatial datasets from extensive collections. Users can navigate using an interactive map or text search to obtain Landsat satellite images, radar data, UAS data, digital line graphs, digital terrain model data, aerial photographs, Sentinel satellite data, some commercial satellite images, including IKONOS and OrbView3, vegetation data, digital cartographic data from the National Map and many more other data sets. Users can search for the exact location using an interactive map or enter specific coordinates to view the available data types.

Since the contours of lakes and reservoirs are variable and seasonal, the use of the last recorded water level is relevant in the study. In this case, it is advisable to use Sentinel-2A/2B multispectral satellite images to determine the contours of water bodies (https://www.sentinel-hub.com/) with high spatial resolution (10 m). The Data Hub service product catalogs provide seamless access to the full set of searchable data using graphical user interfaces, as well as the OData and OpenSearch APIs. Data available for asynchronous access is marked as “offline». An attempt to download these “offline” products will result in their extraction. After their extraction, the requested products will be available for download at the original URL.

Complex climatic data (average precipitation and air temperature indicators) are obtained from the website https://www.worldclim.org/. Where the combined raster files of climatic indicators are located.

Statistical data. Agro-climatic indicators (sum of active temperatures, Selyaninov hydrothermal coefficient, coefficient of cooling) of Almaty region are calculated on the basis of daily climatic data of The POWER Project (https://power.larc.nasa.gov/). The resulting tabular data in CSV format is imported into the GIS environment for further analysis and processing (Kyrgyzbay et al., 2022).

Processing tools in the software of ArcGIS, QGIS. The processing of the elements of the geodata database is carried out using the ArcToolbox (ArcGIS) and Analysis tools (QGIS) groups.

Figure 1 shows the geoprocessing and analysis tools used. With the help of these tools, raster remote sensing data, vector files and statistical information were processed.

ArcGIS. The study used the version of ArcGIS 10.6, which has the following groups of tools and add-ons that were directly involved in creating the geoddatabase of the Almaty region:

1) Analysis tools (Extract, Overlay, Proximity);

2) Cartography tools (Generalization);

3) Connversion tools (Excel, From KML, From Raster, To Raster);

4) Data Management tools (Feautures, File Geodatabase, Generalization, Raster, Topology);

5) Spatial Analyst Tools (Hydrology, Interpolation, Map Algebra, Math, Surface, Zonal);

The Analysis tools group was used for processing vector layers. The process of cutting vector layers according to a certain extent or geometric shape was performed on the basis of Clip. The Overlay group of tools was used in the analysis of internal cutting (Erase), determination of spatial identity (Identity), Spatial Join (Spatial Join), spatial difference (Intersect).

The Cartography tools group is used to change the geometric characteristics of layers. Smooth line and smooth polygon are used to smooth vector layers, simplify line and simplify polygon are designed to simplify vector layers.

The tools of the Hydrology group are used to simulate the flow of water over the surface. The Fill tool fills in the local depressions in the surface raster to remove all small errors and inaccuracies inherent in the data. A local sink is a cell with an uncertain flow direction; there are no cells around it with a height lower than that of this cell. The point of the mouth is the boundary cell with the lowest height for the catchment area of the local depression.

Conversion tools are used to override file formats: excel to convert from tables to Microsoft excel format, From KML to convert to shape file. Spatial definition of vector files from a raster layer was applied using Raster to Point, Raster to Polyline, Raster to Polygon. Similar options converting to raster format are used in the tools Point to Raster, Polygon to Raster, Polyline to Raster.

In the Data Management Tools group, there are the following tools that are used to create components of the Almaty region geodata database: Feautures, File Geodatabase, Generalization, Raster, Topology. The geometric characteristics were changed using the tools Feauture to Line, Feauture to Line, Feauture to Polygon. Interaction with the geodata database and its components is performed on the basis of File Geodatabase (Compact, Recover File Geodatabase). A group of Generalization tools are used to apply the simplification tools and is divided into Dissolve, Eliminate, Eliminate polygon Part. Raster processing is performed using a set of Raster Dataset (Mosaic, Mosaic to New Raster), Raster processing (Clip).

Checking vector layers for compliance with topological rules is performed on the basis of Topology (Add Rule To Topology, Create Topology).

Changing the type and obtaining new spatial information is performed in the Spatial Analyst tools group. The Hydrology tools are designed to analyze water bodies, including Basin and Watershed determines river catchment basins, the Fill tool fills empty cells in the raster, Flow accumulation is designed to analyze the accumulation of data on river basins, Flow distance and Flow Length are designed to calculate the length of the line of river terraces. The input data for all tools and groups are raster files of digital terrain data.

The definition of raster surfaces based on point data is implemented using the Interpolation tools (IDW and Kriging). The Map Algebra (Raster Calculator) tool is used to analyze raster files using mathematical formulas. Math is intended for the application of mathematical constants, formulas and equations for geospatial objects.

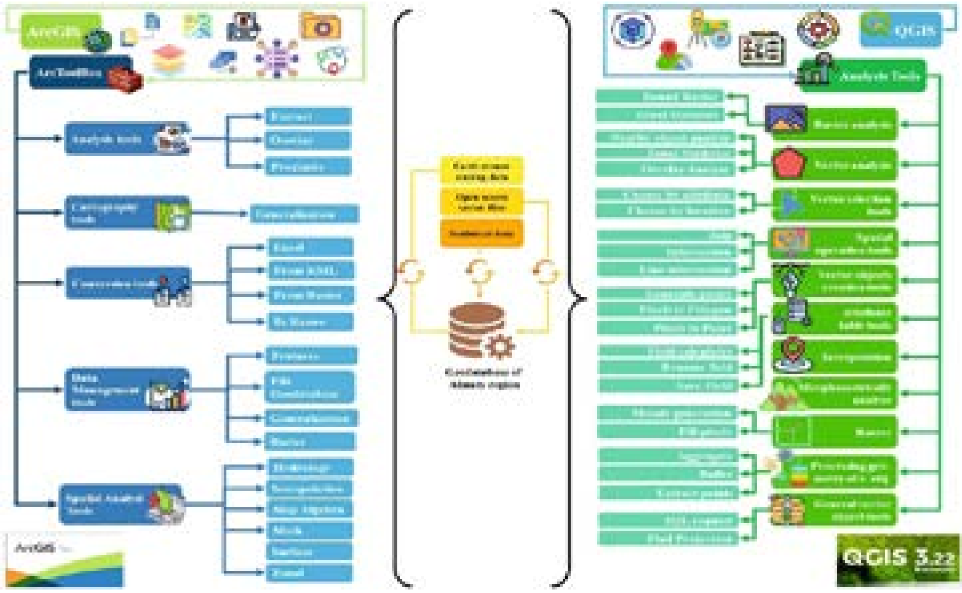


Fig. 1. Analysis and geospatial processing tools used in this study

SRTM data analysis and identification of geomorphological patterns are used in the Surface group of tools (Aspect, Contour, Hillshade, Slope). Aspect is used to determine extended geomorphological forms, relative to the cardinal directions. The Contour tool draws horizontals based on pixel information of absolute height. Hillshade modifies the raster by defining the relief wash. Terrain slope data is calculated using the Slope tool.

The Zonal (Zonal fill) tool works with vector data to create a grid for analysis.

QGIS. Detection and processing of vector and raster data was also performed using GIS software with open source QGIS 3.22 (Figure-1). When creating the geodata database of the Almaty region in QGIS, the following groups of tools were used: Raster analysis (Round Raster, Zonal Statistics), Vector analysis (Nearest Neighbor Analysis, Overlay Analyse, Sum of length, Distance to nearest line/point, Table unique values), Vector selection tools (Choose by attribute, Choose by location, Choose with expression, Extract by attribute, Extract by location, Extract with expression), Spatial operation tools (Extract/ delete by extend, Join, Intersection, Line intersection, Drop by point), Vector objects creation tools (Generate points, Pixels to Polygon, Pixels to Point, Create cells, Points to contour), Attribute table tools (Field calculator, Rename field, Save Field, Delete Field), Interpolation (IDW Interpolation, TIN Interpolation, Line Density, Heat Map), Morphometrical analyse (Hypsometrical curve, Index intersection, Slope, Hillshade, Color relief, Exposition), Processing geometry of vector objects (Agrregate, Buffer, Extract points, Join by values, Rotate, Simplify), General vector object tools (SQL request, Find Projection), Raster (Mosaic generation, Fill pixels).

Raster tools are aimed at processing and analyzing the raster. Round Raster is used to round the pixel values of the raster, Zonal statistics is used to change the raster layer using mathematical formulas and levels.

The analysis of vector data was performed on the basis of the tools Nearest Neighbor Analysis, Overlay Analysis, Sum of length, Distance to nearest line/point, Table unique values. Nearest Neighbor Analysis performs a distance analysis in which the distance to the selected object is determined. Overlay Analysis is used to overlay vector data for data unification. Sum of length calculates the sum of all lengths of polylines. Distance to nearest line/point determines the nearest point or lines relative to the object under study. Work with attribute data is performed on the basis of Table unique values.

Vector data selection is performed based on Vector selection tools. The tools Choose by attribute, Choose by location, Choose with expression are used to select vector objects according to functional characteristics. Extract by attribute, Extract by location, Extract with expression are used for extraction by special characteristics of vector objects. Vector data analysis is performed using the Spatial operation tools tool. Deletion or extraction by the selected extent is performed on the basis of Extract/ delete by extend. Join, Intersection, Line intersection are based on the union of polygonal, point and linear objects, Drop by point is designed to split the line by the selected point.

The main purpose of using and interacting with the two main GIS programs is to compare the results and average the resulting output.

Data analysis using GIS and remote sensing methods. The methods of geospatial analysis in this study are shown in Figure 2.

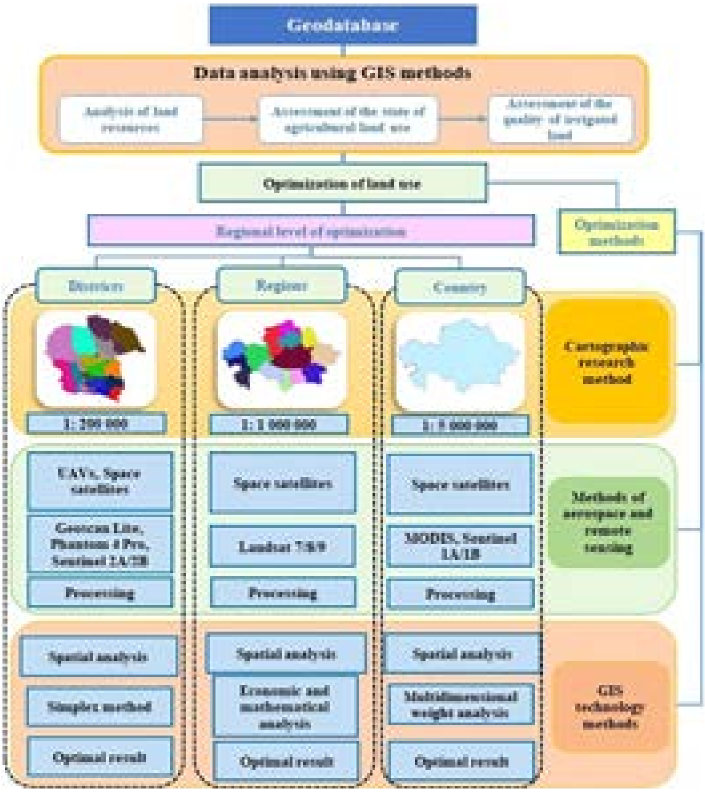


Fig. 2. Processing and analysis of geospatial data to create a geodatabase of the Almaty region by value and scale coverage

The main task of creating a geodatabase of the Almaty region is based on GIS analysis methods at different scales of coverage of the territory. In this study and in future works, the authors selected an object of irrigated agricultural land on the territory of the Almaty region. For this reason, the thematic component of the database is focused on a comprehensive survey of agricultural land.

Data analysis using GIS methods consists of 3 components: analysis of land resources, assessment of the condition of agricultural land and the quality of irrigated land.

The main task of creating a geodata database of the Almaty region is based on GIS analysis methods at different scales of coverage of the territory. In this study and in future works, the authors selected an object of irrigated agricultural land on the territory of the Almaty region. For this reason, the thematic component of the database is focused on a comprehensive survey of agricultural land.

Data analysis using GIS methods consists of 3 components: analysis of land resources, assessment of the condition of agricultural land and the quality of irrigated land.

Land use optimization is a complex group of tools and methods for the rational use of land and resources. Optimization is divided into levels and methods. The methods include the cartographic method, the aerospace and remote sensing method, as well as GIS technology methods. Optimization levels cover from districts to country.

The cartographic research method determines its own scale limits for each optimization level (Kazakhstan from 1: 5,000,000, Almaty region 1: 1,000,000, for districts 1: 200,000) and the level of data generalization.

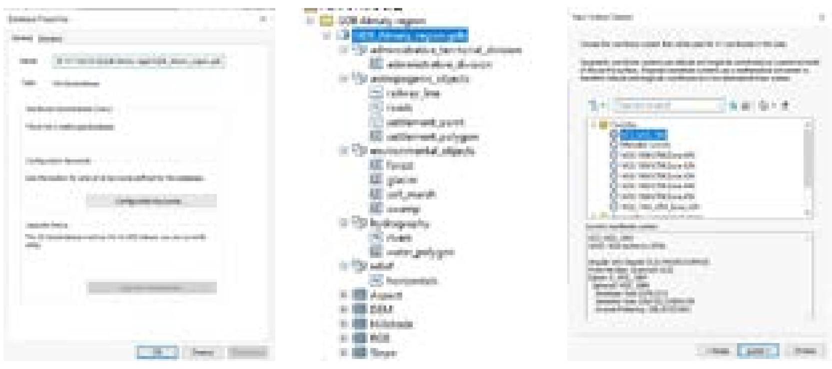
Aerospace and remote sensing methods include sets of raster files from open satellite data such as MODIS, Landsat 7/8/9 Sentinel 2A/2B. MODIS is processed to cover the entire area of the Almaty region with low spatial resolutions to determine contours and field conditions on a small scale. The Landsat data has a 30-meter resolution, thereby allowing the analysis to be carried out on an average scale. At the district level (scale 1:200,000 and larger), Sentinel 2A/2B data were used to accurately determine the contours of agricultural fields.

GIS technology methods include such processing methods as spatial analysis (for all levels), simplex method (district level), economic and mathematical analysis (region level) and multi-dimensional weight analysis (for the territory of the country). After conducting a comprehensive analysis, the most optimal result is selected, taking into account the purpose and specifics of the study.

Results

The created ArcGIS geodatabase is a set of geographical datasets of various types stored in a shared folder of a file system or a multi-user relational database management system (What is a geodatabase?, 2022.).

Figure 3 shows the parameters (a), structure and hierarchy (b), as well as the assignment of the coordinate system (c) for the geodata of the Almaty region. The structure of the layers forming the database consists of separate classes of spatial objects and raster files.



a) b) c)

Fig. 3. Parameters (a), structure (b) and assignment of projection (c) to the geodata database of Almaty region

When combining the spatial data of the Almaty region, the ESRI file database type with the gdb file extension was selected. The file geodatabase of the Almaty region is an extensive information structure where the main sets of object classes and spatial data classes are stored. This type of geodata database has a number of characteristics suitable for the requirements of this study, namely the volume of the database over 1 TB, as well as opening and processing using the ArcGIS program.

The geodatabase contains three main data sets – class of spatial objects, tables and raster sets. The database optimizes the process of managing spatio-temporal data, organizes, systematizes, and stores a wide range of different data. When creating a new feature class, several properties of the feature class were set that define its structure. The creation of a suitable feature class for the data model depended on the properties of the feature class: name, type, geometry properties, coordinate system, configuration keywords, fields, and field properties (Feature class basics, 2022).

When creating a new feature class, several properties of the feature class were set that define its structure. The creation of a suitable feature class for the data model depended on the properties of the feature class: name, type, geometry properties, coordinate system, configuration keywords, fields, and field properties

Figure 4 illustrates the structure of the geodatabase by thematic division. This differentiation was carried out in order to combine similar physical-geographical, administrative-territorial and agricultural characteristics of vector layers and raster images.

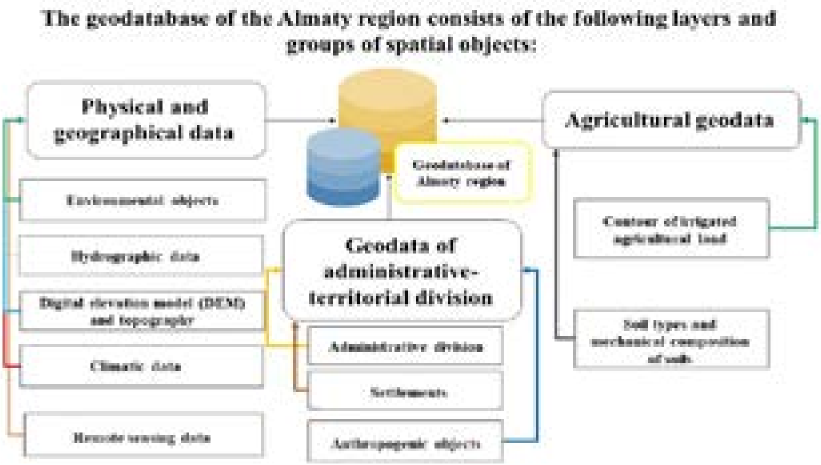


Fig. 4. Division of spatial data into groups on a thematic basis

Physical and geographical geodata are divided into the following groups: environmental objects (glaciers, swamps, salt marshes, sands); hydrography (lakes and reservoirs, rivers, catchment basins); digital elevation model and topography (DEM, slope, exposure, relief washing); climatic geodata (complex climatic information); space images (multispectral channels and their combinations). Geodata of administrative-territorial division consists of groups of administrative divisions (districts and cities of regional subordination), settlements; anthropogenic objects (roads and railways). Agricultural geodata consists of contours of irrigated agricultural lands and types, as well as the mechanical composition of soils.

Figure 5 shows the spatial data classes of the Almaty region. There are administrative-territorial division classes, hydrographic class, DEM data, agricultural class and integrated climate information. Each class is defined and has its own thematic layers in the geodatabase.

Creation and processing of a class of administrative-territorial division and anthropogenic objects. Administrative-territorial and anthropogenic data include the polygonal layer of the Almaty region, the layer of districts and administrative units, data on the division into rural districts, settlements, roads and railways.

The administrative-territorial division layer of the region consists of 17 districts (Aksu, Alakol, Balkhash, Enbekshikazakh, Eskeldi, Zhambyl, Ili, Karatal, Karasay, Kege, Kerbulak, Koksu, Panfilov, Rayymbek, Sarkan, Talgar, Uygur) and 2 cities of regional subordination (the cities of Taldykorgan and Tekeli) (Kyrgyzbay et.al, 2022).

Geospatial processing of the layer was carried out in the ArcMap environment using OSM data (open street map - https://www.openstreetmap.org/). Entering information about the names of administrative units and geometric parameters was carried out in the attribute table using the editor and calculate geometry functions. The geometric parameters included data on the area (area_sqkm) and perimeter (peri_km) of the districts.

Vector layers of roads, railways and settlements were obtained from OSM data. The OSM database consists of streets, local data, and building polygons. Getting access to OSM data in GIS format is integrated into QGIS. This guide shows the process of searching, downloading and using OSM data in QGIS.

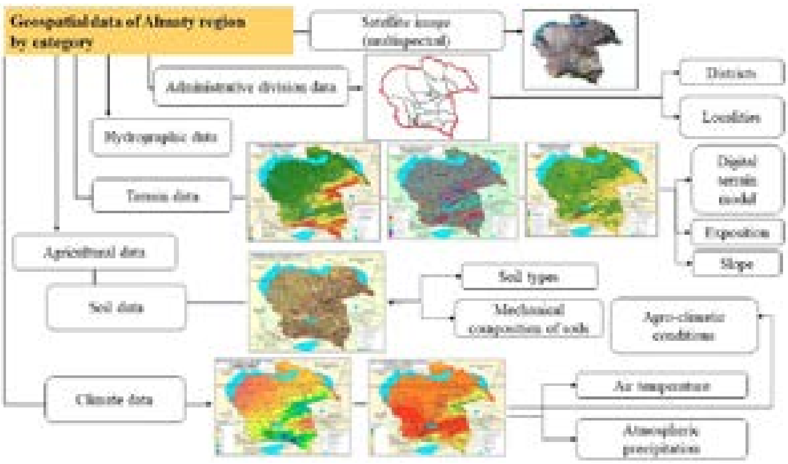


Fig. 5. Components of the geodatabase of the Almaty region

Creating and processing a class of hydrographic data. The class of hydrographic objects consists of lakes, reservoirs, rivers and catchment basins. The spatial layer of lakes and reservoirs was extracted from multispectral satellite images using mathematical formulas, and also identified using medium-scale topographic maps (1: 200 000).

Polygonal water bodies were determined using the NDWI index. Calculates the normalized water index (NDWI) from a multi-channel raster object and returns a raster object with index values. The Normalized Difference Water Index (NDVI) method is an index for determining and monitoring changes in the content of surface waters. NDWI (1) is calculated using the near infrared (NIR) and green (Green) channel:

$$NDWI = (NIR - SWIR) / (NIR + SWIR) \quad (1)$$

The rivers were digitized and identified using satellite images and topographical maps. For vectorization of rivers, DEMS in the form of SRTM images with a spatial resolution of 30 m were used. Data processing took place in the environment of the Spatial Analyst Tools — Hydrology (Fill, Flow Direction, Flow accumulation), Conditional (Con) and Conversion Tools (From Raster- Raster to Polyline).

Processing and analysis of remote sensing data. SRTM raster images were obtained in several stages: 1) registration of a new or existing user into the USGS system by filling in a login and password; 2) selection of the area of interest (in our case, it is the territory of Almaty region); 3) selection of the period and cloud coverage in Search criteria; 4) selection of the type of product in Data sets — Digital Elevation — SRTM — SRTM 1 Arc-Second Global; 5). Next, it was necessary to choose the appropriate raster coatings for the Almaty region

Mosaic creation from raster height data was performed in the ArcGIS-ArcMap environment using the Mosaic to New Raster tool. This tool has combined several raster data sets into a new raster data set. The input data has the same number of channels and bit depth. The mosaic tool has more options available when combining raster datasets into an existing raster. When storing a set of raster data in a JPEG file, a JPEG 2000 file or a geodata database, the compression type and compression quality were specified in the environment parameters (Mosaic To New Raster, 2022).

Creating a digital terrain model, slope steepness data and exposure. The digital relief model (DEM) contains information about the height of only the true relief, excluding vegetation, buildings and other anthropogenic objects. A digital terrain model (DEM) is necessary to obtain the most detailed information about the terrain on any territory.

The digital model of the Almaty region was created on the basis of 30-meter resolution images downloaded and processed in the SRTM GIS environment. The raster surface of the region under study is pixel information with data on the absolute height. The surface is modeled using raster data sets. (Zhang and Gao, 2020).

After processing the DEM, data on slope and slope exposure were obtained. For each cell, the Slope tool calculates the maximum degree of change in the z value between a particular cell and its neighboring cells. In fact, the maximum degree of change in height values per unit distance between a cell and eight adjacent cells determines the steepest descent downhill from the cell. Conceptually, the Slope tool selects a plane for z-values from a neighborhood of 3x3 cells around the processed or central cell.

The slope value of the plane data is calculated using the averaged maximum technique. The direction of the flat faces is the exposure of the processed cell. The lower the slope value, the flatter the earth's surface is; the higher the slope value, the steeper the slopes are located on the surface (Slope, 2022).

The exposure of the slopes was determined using the Aspect tool. This tool extracts the slope exposure from the raster surface. The exposure determines the direction of the slope of the maximum rate of change of values from each cell to its neighboring cells. This characteristic of the relief can be considered as the direction of the slope. The values of the output raster represent the compass directions of the exposure. It is measured clockwise in degrees from 0 (north) to 360 (north again), passing a full circle. Flat areas that do not have a downward direction are given a value of -1 (Lesson: Terrain Analysis, 2022).

Processing of complex climate information. Complex climate information includes data on the average annual amount of precipitation, average annual air temperature and wind speed. Compilation and analysis of complex climate maps of the studied region were carried out on the basis of climate information obtained from the website www.worldclim.org in the format of raster files.

Processing of climatic raster data was carried out in the software ArcMap (ArcGIS) and QGIS, using a GIS tool. Raster Calculator is designed to perform an algebraic expression of a single line using multiple tools and operators using a simple tool interface. In this study, the tool was used to determine the arithmetic mean of monthly climatic indicators (Raster Calculator, 2022).

Processing of agricultural data. The agricultural class of the Almaty region is divided into a layer of agricultural land and data on the characteristics of the soil cover. The vector layer of land shows attribute information and contours of irrigated arable land. The contours of the fields were digitized manually based on Sentinel 2A/2B with a 10-meter spatial resolution.

The soil cover layer includes data on the types and mechanical composition of soils. Data on soil characteristics were used from the FAO website (https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/en /).

Topology check. The Error Inspector allows to view errors in a table that shows which rules were violated, in which classes of objects errors occurred, shows the geometry of errors, the IDs of objects that created errors, as well as errors marked as exceptions. It is possible to sort errors by any fields in the table so that you can work with all errors of a certain type. You can also choose to display only errors of a certain type in the table, or only errors in the visible extent of the map, or only errors marked as exceptions.

In addition to the ability to view and sort errors, the Error Inspector allows you to select errors, navigate and approach selected errors, as well as apply error corrections for their various types.

The ArcMap application allows to check the topology for a piece of data during an editing session. This check can take a long time, especially for large and complex datasets or with a large number of topology rules.

As a result of the above steps and data processing, a geodatabase of the Almaty region was created with 5 spatial classes and 8 raster files.

Discussion

The result of the creation of the geodatabase of the Almaty region can be interpreted as a basis for spatial planning of the territory for various purposes. In this study, for optimal use of spatial data, it was decided to divide the data into parts corresponding to the administrative units of the region (Figure-6).

Each administrative unit was assigned data on thematic coverage. The data is a hierarchical subordination and nesting of layers.

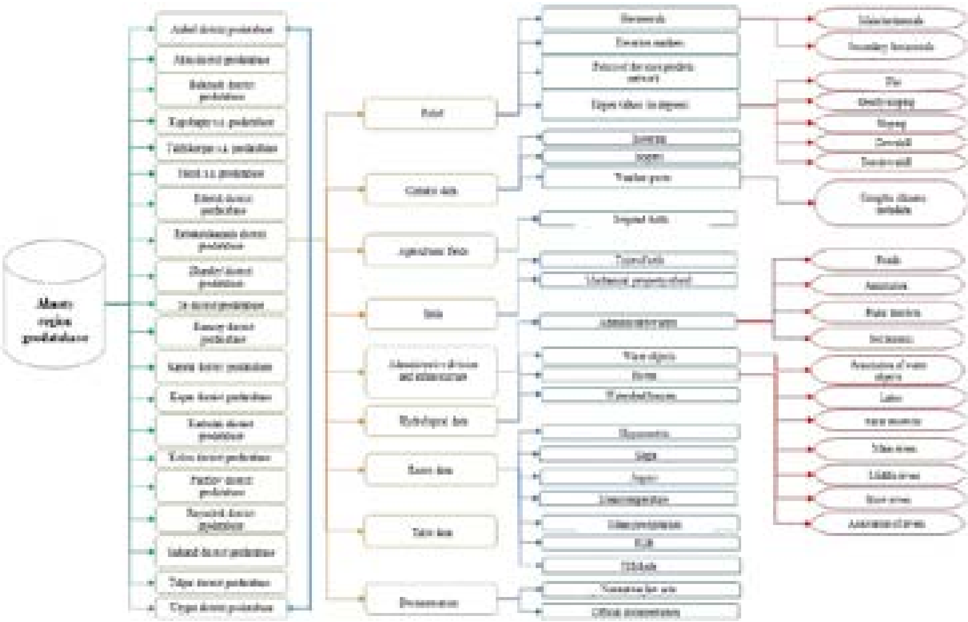


Fig. 6. Hierarchical division of a set of object classes and spatial data classes of the Almaty region on the basis of administrative division

20 administrative units are divided into relief, climatic information, agricultural fields, soils, administrative division, hydrographic data, raster files, tabular data and technical documentation. Each spatial class has its own subclass:

- The relief is divided horizontally (main and auxiliary), data on peaks, slope steepness (smooth, gentle, steep);
- Climate data is divided into isogietes, isotherms and meteorological data (climate metadata);
- The agricultural class has contours of fields and grounds;
- Hydrographic information is divided into lakes, reservoirs, rivers and catchment basins of rivers;
- Raster information has a wide range of data. These include DEM, slope, exposure, average annual air temperature, average annual precipitation, composition of multispectral data;
- Tabular data consists of statistical information;
- Technical documentation includes information about projection data, geometry of layers, etc.

Conclusion

This article presents a study on the creation of a geodata database of the Almaty region. Raster and open vector data were used during the study. Data processing and analysis were carried out in the ArcGIS and QGIS software.

As a result of the research, a geodata database of the Almaty region was created consisting of 5 classes of spatial data and raster files. Each class is divided according to the thematic component and the scope of information. The structure of the geodata database is unified and has one geographical projection.

The model proposed in the article can be supplemented and modified using other data, methods and programs. Our future research plans include a more comprehensive analysis of the relationship of thematic data in relation to the zoning of agricultural irrigated lands, as well as the biological characteristics of individual crops.

REFERENCES

Feature class basics. — Text: electronic//ArcGIS Desktop: [site]. — URL: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/feature-class-basics.htm (date of application: 13.11.2022).

Kyrgyzbay K., 2022 — *Kyrgyzbay K.* Agro-climatic zoning of Almaty region using GIS technologies. *Reports of NAS RK. Computer Science Series*, (2), 76–91. https://doi.org/10.32014/2022.2518-1726.131.

Lesson: Terrain Analysis. — Text: electronic//QGIS documentation: [site]. — URL: https://docs.qgis.org/3.22/en/docs/training_manual/rasters/terrain_analysis.html (date of application: 11.11.2022).

Luzio M., White M., Arnold J., Williams J., Kiniry J., 2017. — *Luzio M., White M., Arnold J., Williams J., Kiniry J.* A Large Scale GIS Geodatabase of Soil Parameters Supporting the Modeling of Conservation Practice Alternatives in the United States. Journal of Geographic Information System, 9. 267–278, doi: 10.4236/jgis.2017.93016.

Mosaic To New Raster. — Text: electronic//ArcGIS for Desktop: [site]. — URL: https://desktop.arcgis.com/en/arcmap/10.3/tools/data-management-toolbox/mosaic-to-new-raster.htm (date of application: 01.11.2022).

Nur W.H., Kumoro Y., Susilowati Y., 2018 — *Nur W.H., Kumoro Y., Susilowati Y.* GIS and Geodatabase Disaster Risk for Spatial Planning. 2018 IOP Conf. Ser.: Earth Environ. Sci. 118 012046. DOI :10.1088/1755-1315/118/1/012046.

Rahman N., Ansary M.A., Islam I., 2015. — *Rahman N., Ansary M.A., Islam I.* Manuscript GIS based Mapping of Vulnerability to earthquake and fire hazard in Dhaka city, Bangladesh Int. J. Disaster Risk Reduct.

Raster Calculator. — Text: electronic//ArcGIS for Desktop: [site]. — URL: https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/raster-calculator.htm date of application: 03.11.2022).

Senanayake S.S., 2017 — Senanayake S.S. Munasinghe and M.A.K., Wickramasinghe W.M.A.D.B. Development of geo-database to recommend suitable crops at village level in Sri Lanka. Annals of Sri Lanka Department of Agriculture 2017. 19: 209 – 228.

Slope. — Text : electronic // ArcGIS for Desktop : [site]. — URL: https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/slope.htm (date of application: 10.11.2022).

What is a geodatabase?. — Text: electronic//ArcGIS Desktop: [site]. — URL: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geodatabases/what-is-a-geodatabase.htm (date of application: 10.11.2022).

Zhang S., Gao H., 2020 — *Zhang S., Gao H.* Using the Digital Elevation Model (DEM) to Improve the Spatial Coverage of the MODIS Based Reservoir Monitoring Network in South Asia. Remote Sens. 2020, 12, 745. https://doi.org/10.3390/rs12050745.

Zhao M. and Liu X., 2017 — *Zhao M. and Liu X.* Reprint of: Regional risk assessment for urban major hazards based on GIS geoprocessing to improve public safety q Saf. Sci. 97. 112–9.

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

Publication Ethics and Publication Malpracticein the journals of the National Academy of Sciences of the Republic of Kazakhstan

For information on Ethics in publishing and Ethicalguidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the National Academy of Sciences of the Republic of Kazakhstan implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any otherlanguage, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The National Academy of Sciences of the Republic of Kazakhstan follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the National Academy of Sciences of the Republic of Kazakhstan.

The Editorial Board of the National Academy of Sciences of the Republic of Kazakhstan will monitor and safeguard publishing ethics.

Правила оформления статьи для публикации в журнале смотреть на сайтах:

www:nauka-nanrk.kz

http://physics-mathematics.kz/index.php/en/archive

ISSN2518-1726 (Online),

ISSN 1991-346X (Print)

Заместитель директор отдела издания научных журналов НАН РК *Р. Жәлиқызы*

Редакторы: *М.С. Ахметова, Д.С. Аленов*

Верстка на компьютере *Г.Д. Жадыранова*

Подписано в печать 30.03.2023.

Формат 60х881/8. Бумага офсетная. Печать – ризограф.

15,5 п.л. Тираж 300. Заказ 1.

Национальная академия наук РК

050010, Алматы, ул. Шевченко, 28, т. 272-64-39