

Robert Oleniacz*, Mateusz Rzeszutek*

Determination of Optimal Spatial Databases for the Area of Poland to the Calculation of Air Pollutant Dispersion Using the CALMET / CALPUFF Model**

1. Introduction

In Poland, as well as in the whole world, advanced pollution dispersion models are playing an increasingly important role, as they take into account the diversity of terrain relief and land use in calculations in the adopted computational grid. One of them is the CALPUFF modeling system, which is a multi-layered air pollution dispersion model simulating the impact of time-varying meteorological parameters on the transport of pollutants. U.S. Environmental Protection Agency (U.S. EPA), in their guidelines on air quality modeling, prefer using this model to assess the transport of pollutants over long distances [1]. This model is supported by a group of geophysical preprocessors preparing terrain information and meteorological preprocessors, as well as a meteorological processor CALMET [2, 3]. CALMET / CALPUFF software is being systematically developed and improved, as evidenced by the approval of the upgrade from version 5.8 to version 5.8.4 by the EPA on the 4th December 2013. The entity responsible for the development of the model is the Atmospheric Studies Group (ASG), acting under TRC Environmental Corp. [4].

CALMET is a meteorological model which generates two-and three-dimensional meteorological fields of the selected parameters. In the calculation of the three-dimensional meteorological grid of the wind speed and direction, it takes into account: the kinematic terrain effects, according to the approximation by Liu and Yocke [5], terrain blocking effects to air mass flow parameterized by the local Froude number [6] and slope flows from the hills determined from empirical formulas based on the Mahrt's work [7]. This model requires an introduction of a number of meteorological parameters derived from surface stations, upper air stations, as well as

* AGH University of Science and Technology, Faculty of Mining Surveying and Environmental Engineering, Department of Environmental Management and Protection, Krakow, Poland

** This paper has been prepared within the scope of the AGH UST statutory research no. 11.11.150.008

information about the precipitation amount. There is also a possibility of introducing additional or alternative meteorological data from the forecasting models such as MM5/MM4, WRF, RAMS, RUC, or others [3, 8].

A group of the so-called geophysical preprocessors is composed, inter alia, of the programs such as TERREL, CTGPROC and MAKEGEO. A group of these programs converts the existing digital elevation and land use data to a modeling grid defined by the user. TERREL preprocessor converts digital elevation models and CTGPROC digital land use data. The last preprocessor for the preparation of spatial data is MAKEGEO, which reads the fractional data processed by the TERREL and CTGPROC preprocessors, assigning factors dependent on the category of land use and creating output data file GEO.DAT [8, 9].

A group of CALMET / CALPUFF models is widely used in the world to calculate the dispersion of pollutants in the air in a terrain which is spatially variable, and to determine the impact of fragrances on air quality [10–15]. In Poland, this model is used, among others, for developing regional air protection programs, as well as for advanced forecasting of the levels of substances in the ambient air [16, 17]. Hence, it is important to determine spatial data sets which truly reflect the actual relief and land use for the Polish territory. The aim of this study is to identify widely-available optimal spatial data that can be used in the modeling process of pollutant dispersion in the air, conducted using the CALMET / CALPUFF model.

2. Spatial Data Sources

Among the global spatial data on terrain relief, GTOPO30 and SRTM data deserve special attention, as they can be successfully used in the CALMET / CALPUFF modeling system of pollution dispersion. The above-mentioned digital elevation models are made available free of charge by the U.S. Geological Survey's.

GTOPO30 is a global digital elevation model (DEM) with a 30 arc-second grid spacing (about 1.0 km) completed in 1996. It was created through the cooperation of the U.S. Geological Survey, Earth Resources Observation and Science (EROS) Center, and many other state institutions around the world. GTOPO30 data are based on raster and vector sources of topographic information, and its scope covers the entire globe [18].

In the year 2000, the National Aeronautics and Space Administration (NASA) in cooperation with the National Geospatial-Intelligence Agency (NGA), Jet Propulsion Laboratory (JPL), National Aeronautics and Space Research Centre of the Federal Republic of Germany (DLR – Deutsches Zentrum für Luft- und Raumfahrt e.V) and the Italian Space Agency (ASI – Agenzia Spaziale Italiana) sent the space shuttle Endeavour on an 11-day mission. Its aim was the mapping of the terrain relief of almost the entire surface of the earth using a single-pass interferometry. As a result of the Shuttle Radar Topography Mission (SRTM), spatial data for 80% of the land lying between 60° north and 56° south latitude were obtained. High resolution data

of 1 arc-second (approximately 30 m) were available for the U.S. territory, while for the rest of the studied surface – data at a 3 arc-second resolution (approximately 90 m) [19]. After calibration of the measurement system, the mean elevation error for the territory of Eurasia was 6.2 m [20], and for the territory of Poland the mean error was at the level of 2.7 m in the hills, while on the plains it was about 1.0 [21].

The U.S. Geological Survey also provides spatial data sets containing information on land use, which can be effectively exploited in the process of preparation of spatial data for air pollution dispersion modeling. Land use data with a high 3 arc-second resolution are available to the U.S. territory, exclusively. GLCC data (Global Land Cover Characterization) are sets of information about the land cover for all the continents, which are based primarily on unsupervised classification. These data are characterized by a 30 arc-second resolution (approximately 1.0 km) and are available to all continents, however, they come from the period of April 1992 – March 1993. Land cover map was developed through the cooperation of the U.S. Geological Survey (USGS), the National Center for Earth Resources Observation and Science (EROS), University of Nebraska-Lincoln (UNL) and the Joint Research Centre (JRC) [22].

The European Environment Agency (EEA) within the Corine project provides, inter alia, land cover data sets – Corine Land Cover – for the European Union countries. CLC 2006 land cover data are available in the vector and raster forms with a resolution of 100 and 250 meters. This database contains a uniform classification of terrain, three hierarchical levels were established. The first level includes five main types of the globe cover, the second level is distinguished by 15 forms possible to be presented on the maps at 1:500,000 and 1:1000,000 scales, while the third level (featuring the greatest detail) distinguishes 44 land use classes. Information on the land cover is available to 38 countries. The classification system is uniform for all the countries participating in the project [23, 24]. In Poland, the unit responsible for the CLC 2006 project is the Chief Inspectorate of Environmental Protection, acting as the National Focal Point for cooperation with the EEA. A direct contractor was the Institute of Geodesy and Cartography. The funds for the national CLC 2006 project came from the sources of the European Environment Agency and the National Fund for Environmental Protection and Water Management [25].

3. Methodology for Preparing Spatial Data – the Example of the City of Krakow

Preparing the assumed computational grid containing spatial information required for the purposes of an air pollutant dispersion modeling process, conducted using the CALMET / CALPUFF model, consists of three stages (Fig. 1):

1. acquiring and processing data of digital elevation model,
2. creating a reclassified grid of land use category,
3. combining the processed data into a single output file of the MAKEGEO pre-processor including the assignment of geophysical factors.

Below, an exemplary methodology for the preparation of the above spatial data for the city of Krakow area has been presented.

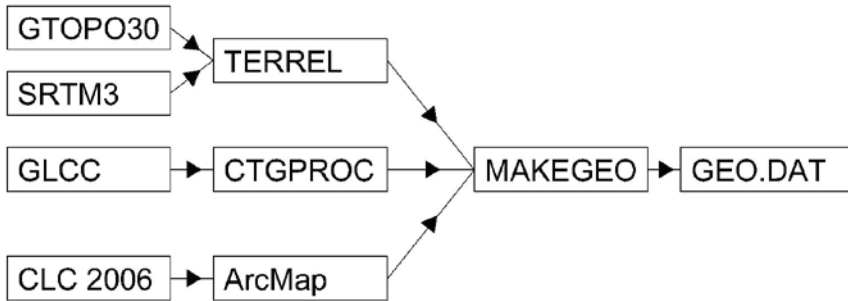


Fig. 1. Schematic methodology for the preparation of spatial data for the air pollutant dispersion modeling using the CALMET / CALPUFF model

3.1. Digital Elevation Model

The first stage of this complex process of preparing spatial data is data acquisition of digital elevation model. Elevation data can be extracted from the GTOPO30 and SRTM3 databases [18, 19]. Data downloaded from the above-mentioned spatial databases are input information to the TERREL preprocessor. This program accepts horizontal elevation values, cuts out a selected range of data, then calculates the coordinate system and creates a grid of mean value cells or values used in various modeling systems. TERREL creates a terrain grid as a file of elevation data in discrete receptors for the MAKEGEO processor. Depending on the data type, resolution of the computational grid can be defined, assuming that it must be equal to, or less than, the resolution of the input data. In the case of creating a grid with lower resolution compared to the input data, the program averages the data within the adopted discrete receptor. The TERREL preprocessor also allows us to enter coastline data if seas or oceans are included in the computational grid. It is possible to prepare only one computational grid of digital elevation model in the program.

Using this tool, four computational grids for a part of the area of the city of Krakow and the surrounding areas were prepared, which differed in resolution or origin of the input data. The three-dimensional visualization of the prepared elevation data has been shown in Figures 2–5. Table 1 contains the characteristics of digital elevation model computational grids. These grids have a rectangular coordinate system with an area of 26 × 26 km in the WGS 84 reference system (World Geodetic System '84) with the UTM mapping (Universal Transverse Mercator). UTM zone 34 was adopted, and the coordinates of the lower left corner of the computational grid (x, y) are $x = 420,000$; $y = 5532,000$.

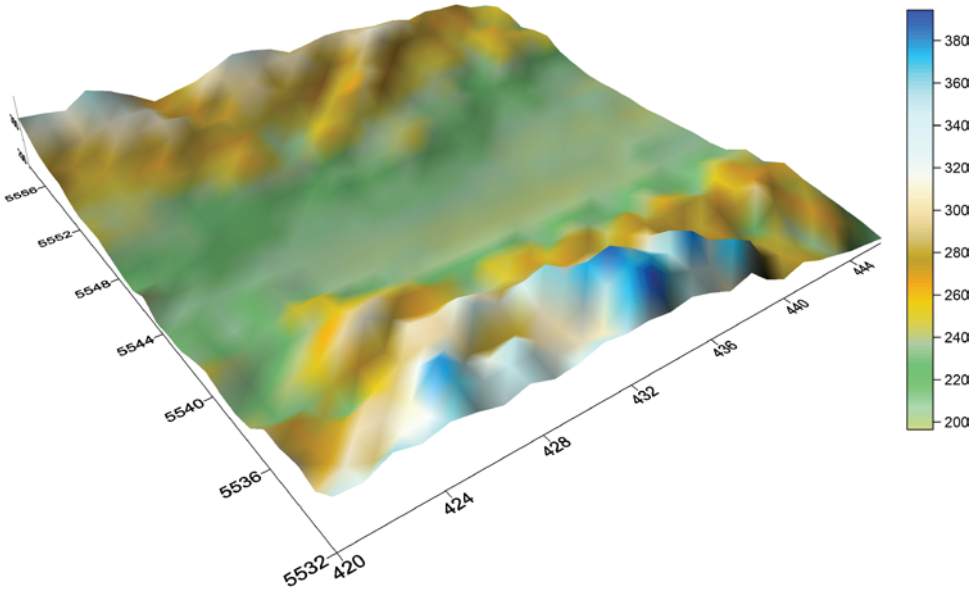


Fig. 2. Three-dimensional visualization of the GTOPO30 digital elevation model for a computational grid with a resolution of 1.0 km (the Krakow region)

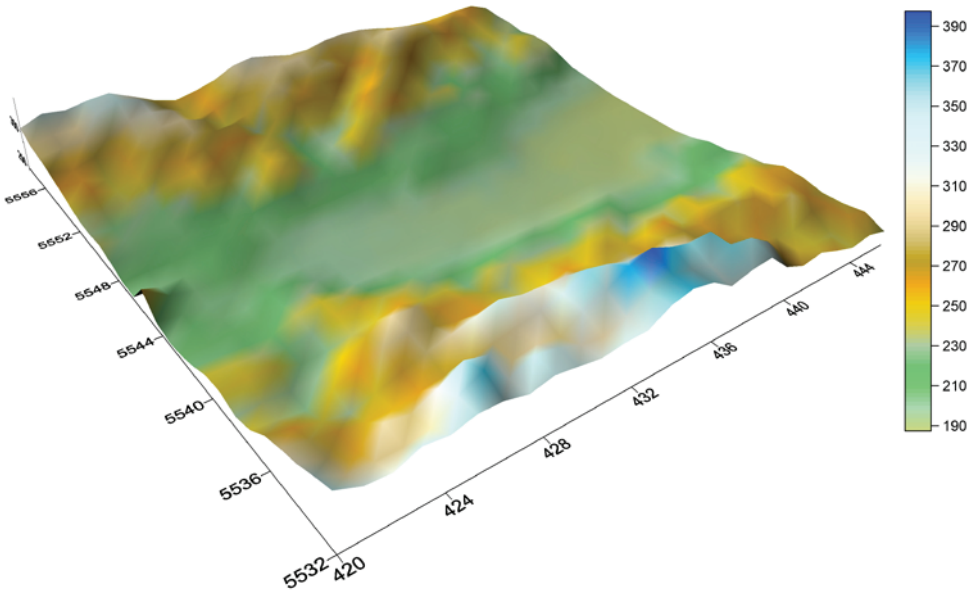


Fig. 3. Three-dimensional visualization of the SRTM digital elevation model for a computational grid with a resolution of 1.0 km (the Krakow region)

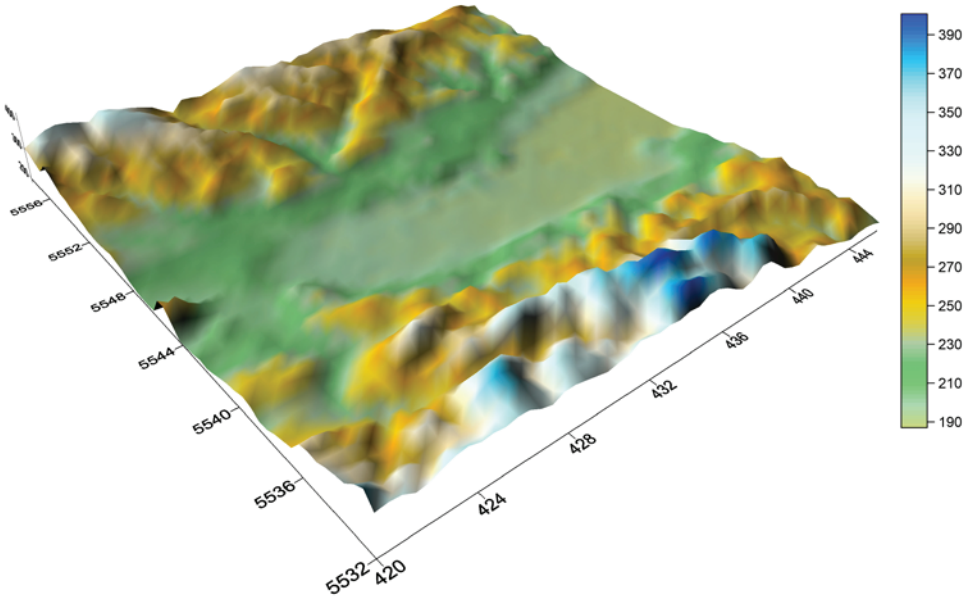


Fig. 4. Three-dimensional visualization of the SRTM digital elevation model for a computational grid with a resolution of 0.5 km (the Krakow region)

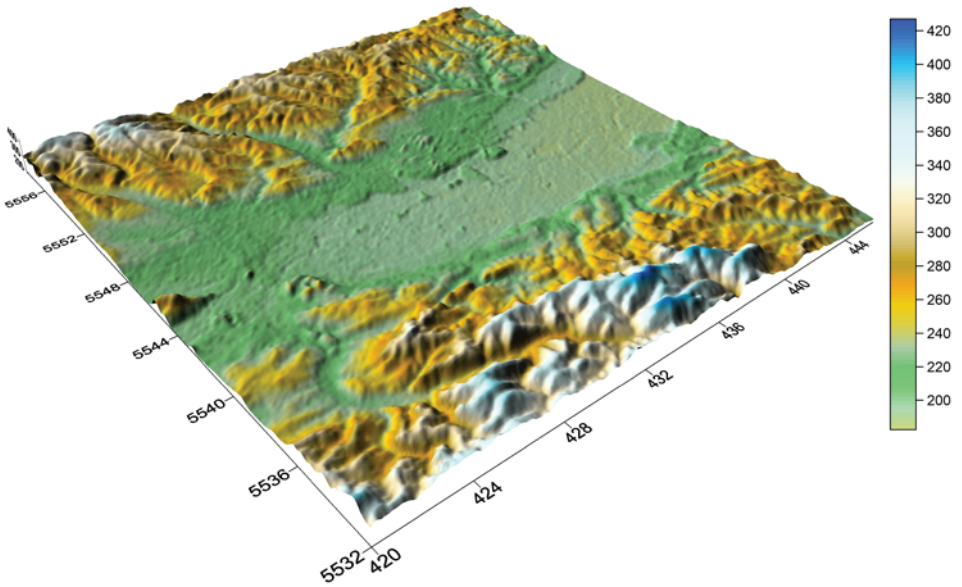


Fig. 5. Three-dimensional visualization of the SRTM digital elevation model for a computational grid with a resolution of 100 m (the Krakow region)

Table 1. Characteristics of computational grids of a digital elevation model (DEM)

Name of grid (DEM)	Resolution grid [m]	Database
TERREL_00	1000 × 1000	GTOPO30
TERREL_01	1000 × 1000	SRTM3
TERREL_02	500 × 500	SRTM3
TERREL_03	100 × 100	SRTM3

3.2. Land Use

The second stage in the preparation of spatial data is creating a computational grid of the land use category. To prepare computational grids, the data from The Global Land Cover Characterization (GLCC) and Corine Land Cover 2006 (CLC 2006) [22, 25] were used. GLCC data is input information for the preprocessor CTGPROC. This program, just like TERREL, cuts out a defined range of data and calculates the coordinate system, producing in effect a computational grid containing digital information the land use in the adopted discrete receptors for the preprocessor MAKEGEO. In this program, it is possible to define the computational grid resolution, however, it is advisable to use a resolution not lower than it is apparent from the characteristics of the source data. In the case of non-compliance with this condition, the processed computational grid will be incomplete.

As a result of the conducted activities, a computational grid of land use category of a single discrete receptor dimension of 1 × 1 km was prepared. The length of the grid sides was assumed to be 26 km each, and as a reference point – lower left corner of the grid was identified in the WGS 84 reference system (World Geodetic System '84) with the UTM mapping (Universal Transverse Mercator) for the zone 34 with the coordinates: $x = 420,000$; $y = 5532,000$.

The Corine Land Cover 2006 data are not compatible with the CTGPROC preprocessor. The main differences are the format of data storage, the reference system and the land use classification category. The GLCC database contains 38 categories of land use, in the Corine Land Cover data set at the third detail level, there are 44 classes, while the meteorological processor CALMET distinguishes up to 14 categories of land use. In order to implement the CLC 2006 data to the MAKEGEO processor, ArcMap software was used. In this program, procedures specific to the CTGPROC program were conducted (i.e. conversion of the coordinate system, cutting out the defined data region and reclassification of land use categories), and the data set was exported to a file in the ASCII format. The reclassification of the CLC data for the classification of land use compatible with the CALMET processor was carried out in accordance with the assignment contained in Table 2. The created ASCII file contains information on the land use category in a local coordinate system, it requires, however, data to be additionally sorted and attributes to be assigned. The attributes are defined land use

categories and the occurrence of a given attribute in a discrete receptor is defined in the binary system. These operations can be performed in a spreadsheet, resulting in an input file of the MAKEGEO preprocessor.

As a result of the conducted activities, four computational grids with different input data source and computational grid resolution were prepared for the area in question. The description of the prepared land use data has been presented in Table 3. Graphical representation of the results of the conducted activities has been depicted in Figure 6. The generated computational grids with information on the land use category are characterized by distinct differences in terms of data sources and in terms of resolution. In a grid with a resolution of 1 km generated from the GLCC data, only three categories of land use can be distinguished, with the following codes: 10, 20 and 40. On the other hand, in a grid with the same resolution created from the CLC data, six categories of land use can be distinguished, with the following codes 10, 20, 30, 40, 51 and 70. In the computational grids CTG_02 and CTG_03, with 500 and 100 m resolutions respectively, there are also six categories of land use, however, they reflect the occurrence of natural elements such as rivers or forests much more clearly.

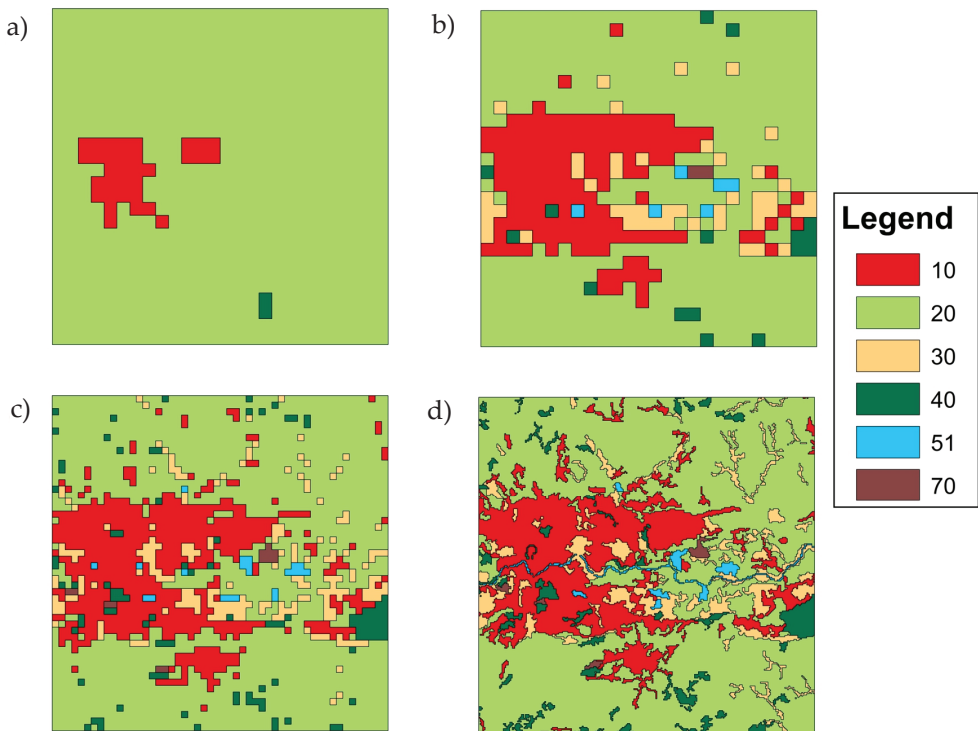


Fig. 6. Graphical representation of the adopted computational grids of land use categories: a) CTG_00; b) CTG_01; c) CTG_02; d) CTG_03

Table 2. Assignment of Corine Land Cover land use classes for the CALMET model

Classification of Corine Land Cover database		Land use classification for CALMET	
Level 3	GRIDCODE	Level 1	GRIDCODE
Continuous urban fabric	1	Urban of built-up land	10
Discontinuous urban fabric	2		
Industrial or commercial units	3		
Road and rail networks and associated land	4		
Port areas	5		
Airports	6		
Construction sites	9		
Non-irrigated arable land	12	Agricultural land – unirrigated	20
Permanently irrigated land	13		
Vineyards	15		
Fruit trees and berry plantations	16		
Olive groves	17		
Annual crops associated with permanent crops	19		
Complex cultivation patterns	20		
Land principally occupied by agriculture, with significant areas of natural vegetation	21	Agricultural land – irrigated	-20
Agro-forestry areas	22		
Rice fields	14	Rangeland	30
Sport and leisure facilities	11		
Pastures	18		
Natural grasslands	26		
Moors and heathland	27		
Sclerophyllous vegetation	28		
Transitional woodland-shrub	29		
Green urban areas	10	Forest land	40
Broad-leaved forest	23		
Coniferous forest	24		
Mixed forest	25		
Water courses	40	Small water body	51
Water bodies	41		
Coastal lagoons	42	Bays and estuaries	54
Estuaries	43		
Sea and ocean	44	Ocean and sea	55
Inland marshes	35	Forested wetland	61

Table 2. cont.

Classification of Corine Land Cover database		Land use classification for CALMET	
Level 3	GRIDCODE	Level 1	GRIDCODE
Peat bogs	36	Nonforested wetland	62
Salt marshes	37		
Salines	38		
Intertidal flats	39		
Mineral extraction sites	7	Barren land	70
Dump sites	8		
Beaches, dunes, sands	30		
Bare rocks	31		
Sparsely vegetated areas	32		
Burnt areas	33		
Glaciers and perpetual snow	34	Perennial snow or ice	90

Table 3. Characteristics of land use computational grids

Name of grid land use	Resolution grid [m]	Database
CTG_00	1000 × 1000	GLCC
CTG_01	1000 × 1000	CLC2006
CTG_02	500 × 500	CLC2006
CTG_03	100 × 100	CLC2006

3.3. Creation of Input Data for CALMET Model

The final stage of the preparation of spatial data for the CALMET processor is combining the previously prepared data sets of land use and the digital elevation model into one input file (GEO.DAT) of the CALMET model, as well as assigning the factors dependent on the land use category. In order to generate a geophysical data file (GEO.DAT), the MAKEGEO preprocessor is used.

This processor combines the two sets of data mentioned above and assigns the following factors with respect to the land use categories [2, 3]:

- aerodynamic surface roughness,
- albedo (the ratio of the reflected to incident radiation),
- Bowen ratio (the ratio of the heat collected by the water surface to the amount of heat used for the evaporation),
- soil heat flux,
- anthropogenic heat flux,
- leaf area index.

The values of the above parameters with respect to land use categories have been summarized in Table 4. It contains factors which are necessary for the proper designation of meteorological parameters by the CALMET processor. As it was in the case of the TERREL and CTGPROC preprocessors, in order to run the program, it is necessary to specify the mapping and the reference system of the input data, the data resolution and the coordinates of the reference point, which is the lower left corner of the assumed computational grid. As a result of the carried out works, four sets of geophysical data for the area of the city of Krakow and its surroundings were obtained, different in terms of the origin of the source data and the horizontal dimensions of the discrete receptors of the computational grid.

Table 4. Summary of geophysical factors depending on the land use category

GRIDCODE	Surface roughness length	Albedo	Bowen ratio	Soil heat flux constant	Vegetative leaf area index
10	1.0	0.18	1.5	0.25	0.2
20	0.25	0.15	1.0	0.15	3.0
30	0.05	0.25	1.0	0.15	0.5
40	1.0	0.1	1.0	0.15	7.0
51	0.001	0.1	0.0	1.0	0.0
70	0.05	0.3	1.0	0.15	0.5

4. Summary

The development of remote sensing and geoinformatics tools contributed to generating new data resources, which reflect the terrain and the manner of land use more accurately. These data are applied in calculating the dispersion of pollutants in the ambient air using the CALMET / CALPUFF model. The data of the GTOPO30 digital elevation model and the GLCC land use have a global reach, and despite lower accuracy, they allow for modeling at a macro scale using a rare computational grid. The advantage of the above data is their compatibility with geophysical data preprocessors, which greatly simplifies the data preparation process. These data can be successfully exploited in calculating the transport of pollutants over long distances using commercially reasonable computational costs.

The application of more accurate SRTM3 and CLC 2006 data enables more accurate reflection of the relief and land use by the CALMET / CALPUFF model, which may contribute to an increased accuracy in calculating the results. The use of these data, assuming a computational grid with horizontal dimensions of one receptor of 100 × 100 m, allows for the pollution dispersion modeling at a local scale at the most, as computational costs increase significantly with the growing density of a grid. Moreover, the CLC 2006 data are available only to 38 countries

in Europe, and the SRTM3 data does not cover the most remote north and south areas with its range.

When comparing the GLCC and CLC 2006 data it is evident that even with a 1 km resolution, the prepared computational grid based on the CLC 2006 data is characterized by a more accurate mapping of the actual coverage, compared with the grids of the GLCC dataset. Both for Poland and other European countries, it is recommended to use the SRTM3 and CLC 2006 data, as they cover almost the whole of Europe with their range and they are more detailed.

References

- [1] U.S. Environmental Protection Agency: *Revision to the Guideline on Air Quality Models: Adoption of a Preferred General Purpose (Flat and Complex Terrain) Dispersion Model and Other Revisions; Final Rule*. U.S. EPA, 40 CFR Part 51. Federal Register, vol. 70, no. 216, 2005, pp. 68217–6826.
- [2] Scire J.S., Strimaitis D.G., Yamartino R.J.: *A User's Guide for the CALPUFF Dispersion Model*. Concord, 2005.
- [3] Earth Tech, Inc.: *Development of the Next Generation Air Quality Models for Outer Continental Shelf (OCS) Applications (CALPUFF and Postprocessors)*. Concord, 2006.
- [4] Atmospheric Studies Group, [on-line:] www.src.com [access: December 6, 2013].
- [5] Liu M.K., Yocke M.A.: *Siting of wind turbine generators in complex terrain*. Journal of Energy, vol. 4, no. 1, 1980, pp. 10–16.
- [6] Allwine K., Whiteman C.: *Green river air quality model development. MESLAR – a mesoscale air quality model for complex terrain. Overview, technical description and user's guide*. U.S. Environmental Protection Agency, Washington 1985.
- [7] Mahrt L.: *Momentum balance of gravity flows*. Journal of Atmospheric Science, vol. 39, 1982, pp. 2701–2711.
- [8] Scire J.S., Robe F.R., Fernau M.E., Yamartino R.J.: *A User's Guide for the CALMET Meteorological Model (Version 5)*. Concord, 2000.
- [9] Rzeszutek M.: *Ocena oddziaływania Zakładu Termicznego Przekształcania Odpadów Komunalnych w Krakowie na jakość powietrza z wykorzystaniem matematycznego modelu dyspersji CALPUFF*. AGH, Wydział Geodezji Górniczej i Inżynierii Środowiska, Katedra Kształtowania i Ochrony Środowiska, Kraków 2013 [M.Sc. thesis, unpublished].
- [10] Levy J., Spengler J., Hlinka D., Sulluvan D., Moon D.: *Using CALPUFF to evaluate the impacts of power plant emissions in Illinois: model sensitivity and implications*. Atmospheric Environment, vol. 36(6), 2002, pp. 1063–1075.
- [11] Szczygłowski P., Mazur M.: *Zastosowanie modelu Calmet/Calpuff do obliczeń poziomu stężeń zanieczyszczeń pochodzących z wysokich emitorów punktowych*. Inżynieria Środowiska, t. 10, z. 2, 2005, pp. 195–205.

- [12] Szczygłowski P., Mazur M.: *Modelling dispersion of air pollutants over the area of diversified relief based on the Calmet/Calpuff model*. Environment Protection Engineering, vol. 32, no. 4, 2006, pp. 73–77.
- [13] Tartakovsky D., Broday D.M., Stern E.: *Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain*. Environmental Pollution, vol. 179, 2013, pp. 138–145.
- [14] Ranzatoa L., Baraussea A., Mantovani A., Pittarello A., Benzo M., Palmeri L.: *A comparison of methods for the assessment of odor impacts on air quality: Field inspection (VDI 3940) and the air dispersion model CALPUFF*. Atmospheric Environment, vol. 61, 2012, pp. 570–579.
- [15] Vieira de Melo A.M., Santos J.M., Mavroidis I., Reis Junior N.C.: *Modelling of odour dispersion around a pig farm building complex using AERMOD and CALPUFF. Comparison with wind tunnel results*. Building and Environment, vol. 56, 2012, pp. 8–20.
- [16] Biuro Studiów i Pomiarów Proekologicznych „EKOMETRIA” Sp. z o.o.: *Opracowanie prognozy zanieczyszczenia powietrza pyłem drobnym w Polsce na lata 2010, 2015 i 2020 wraz z analizą uwarunkowań i oceną kosztów osiągnięcia standardów dla pyłu określonych dyrektywą w sprawie jakości powietrza atmosferycznego i czystszej dla Europy*. Gdańsk 2009.
- [17] ATMOTERM SA: *Program ochrony powietrza dla województwa małopolskiego*. Kraków 2013.
- [18] Global 30 Arc-Second Elevation (GTOPO30), [on-line:] <https://lta.cr.usgs.gov/GTOPO30> [access: December 6, 2013].
- [19] Shuttle Radar Topography Mission (SRTM), [on-line:] <https://lta.cr.usgs.gov/SRTM2> [access: December 6, 2013].
- [20] Rodrigues E., Morris C.S., Belz J.E.: *A Global Assessment of the SRTM Performance*. Photogrammetric Engineering & Remote Sensing, vol. 72, no. 3, 2006, pp. 249–260.
- [21] Karwel K.A., Ewiak I.: *Ocena dokładności modelu SRTM na obszarze Polski*. Archiwum Fotogrametrii, Kartografii i Teledetekcji, vol. 16, 2006, pp. 289–296.
- [22] Global Land Cover Characterization (GLCC), <https://lta.cr.usgs.gov/GLCC> [access: December 6, 2013].
- [23] Büttner G., Kosztra B., Maucha G., Pataki R.: *Final draft – Implementation and achievements of CLC2006*. Institute of Geodesy, Cartography and Remote Sensing (FÖMI), Barcelona 2012.
- [24] European Environment Agency: *CLC2006 technical guidelines*. EEA Technical report, no. 17, Copenhagen 2007.
- [25] CORINE Land Cover (CLC), [on-line:] <http://clc.gios.gov.pl> [access: December 6, 2013].