ISSN 1068-3739, Russian Meteorology and Hydrology, 2020, Vol. 45, No. 7, pp. 515–521. Allerton Press, Inc., 2020.
Russian Text The Author(s), 2020, published in Meteorologiya i Gidrologiya, 2020, No. 7, pp. 112–119.

COMMUNICATIONS

Computation of City-descriptive Parameters for High-resolution Numerical Weather Prediction in Moscow Megacity in the Framework of the COSMO Model

T. E. Samsonov[a, b, c] **and M. I. Varentsov**[a, b, c*]

[a]*Hydrometeorological Research Center of the Russian Federation, Bolshoi Predtechenskii per. 11–13, Moscow, 123242 Russia*

[b]*Lomonosov Moscow State University, GSP-1, Leninskie Gory, Moscow, 119991 Russia*

[c]*Moscow Center for Fundamental and Applied Mathematics, GSP-1, Leninskie Gory, Moscow, 119991 Russia*

**e-mail: mvar91@gmail.com*

Received April 30, 2020
Revised June 4, 2020
Accepted June 4, 2020

Abstract—A new methodology is proposed for computing the city-descriptive parameters for the TERRA_URB scheme in the COSMO model. It is based on the GIS analysis of OpenStreetMap data, Sentinel-2 satellite images, and global land cover data. A set of parameters is obtained for Moscow megacity on a grid with the spacing of 500 m. The dataset is promising for high-resolution numerical weather prediction within the COSMO-Ru system and for urban climate research.

DOI: 10.3103/S1068373920070079

Keywords: Atmospheric model, numerical weather prediction, urban areas, city-descriptive parameters, urban canopy parameters, GIS technologies, COSMO, urban climate

INTRODUCTION

The elaboration of weather and climate services for megacities requires developing atmospheric modeling methods for urban conditions [6]. The atmospheric models with the grid spacing of 1 km and less, which take into account the mosaic pattern of urban landscapes, are currently used for numerical weather prediction (NWP) and scientific research. So called urban surface parameterizations are used to describe the interaction between the atmosphere and buildings on such scale [8]: it is assumed that a model grid cell or its part is occupied by the homogenous urban environment characterized by a set of parameters. The success of the modeling for cities is largely determined by the accuracy of specifying these parameters. Many studies deal with the problem of their quantification [10], but the choice of the optimum method is still debatable. The building parameters are most precisely identified from the vector-formatted cartographic data, and the distribution of the types of land cover such as vegetation is efficiently determined from satellite images [15]. At the same time, the ambiguity of correlating the simplified description of the urban environment in the model with reality, a limited access to official cartographic information, and the absence of uniform standards for its structure impede a development of the universal solution.

The recent studies on modeling the meteorological regime of Moscow revealed that the COSMO model with the TERRA_URB parameterization [16] and a grid spacing of 1 km is promising for urban climate research [2, 3, 14] and NWP within the COSMO-Ru forecasting system (Hydrometcenter of Russia) [4, 5]. The ongoing work on using a smaller grid spacing (for example, 500 m) grid spacing in the framework of the COSMO-Ru system development requires computing the city-descriptive parameters with a proper resolution. The present paper deals with solving this problem using new sources of information which allow proposing a more universal and potentially more precise method for determination of necessary parameters.

Thus, the objective of the present research is to develop a method for the calculation of city-descriptive parameters needed for modeling based on the relevant open sources of global data and its application for the Moscow region using the COSMO model grid with a spacing of 500 m.

THE TERRA_URB SCHEME AND REQUIRED CITY-DESCRIPTIVE PARAMETERS

The consideration of urban land cover features in the TERRA_URB scheme is carried out by the correction of surface properties in the parameterization of the land active layer [16]. A set of the city-descriptive parameters in TERRA_URB is largely comparable to other similar parameterizations [8]. The fraction of urbanized territory ${}_{\text{urb}}$ determines for which part of a grid cell TERRA_URB is applied. This part of the cell is considered “sealed,” i.e., it is impervious for moisture and devoid of vegetation. The surface thermodynamic parameters are corrected in order to take into account the effects of buildings whose geometry is determined by the following characteristics: the fraction of buildings ${}_{\text{bld}}$; their mean height H; and the ratio H/W, where W is the average width of urban canyons (streets with buildings stretched along them [3, 12]). The height H determines the roughness length of the urban surface, and the others parameters determine the values of its effective albedo, emissivity, heat capacity, and thermal conductivity. To consider the anthropogenic heat flux, it is necessary to specify its average annual value Q_{h}, based on which the flux for a specific time moment is calculated with account of predefined seasonal and diurnal variations.

According to the earlier estimates of the TERRA_URB sensitivity, the most important parameters are ${}_{\text{urb}}$ and Q_{h}. These parameters determine a possibility of simulating the urban heat island in the model where a temperature difference between the city and suburbs reaches up to 10 C, and other urban mesoclimatic effects [2, 3, 14]. The role of parameter ${}_{\text{urb}}$ is greater in summer, the contribution of Q_{h} is more significant in winter. The parameters ${}_{\text{bld}}$, H, and H/W determine the finer features of the temperature distribution in the city; the model sensitivity to their change does not exceed tenths of a degree [16]. In addition, a significant sensitivity of simulated wind speed to the change in the average height of buildings may be expected.

INITIAL DATA

The innovation of the proposed method consists in the joint use of three independent sources of data: OpenStreetMap (OSM), Copernicus Global Land Cover (CGLC) data with a resolution of 100 m [7], and Sentinel-2 satellite images with a resolution of 10 m.

The OSM data contain the vector polygons corresponding to individual buildings, the territories with different land use types, and the road and street network objects in the linear form. The accuracy of these data varies in space but, for thoroughly mapped urban areas, is comparable to official cartographic data [9, 11]. Previously, when using TERRA_URB for Moscow, the OSM data were used as the only source of information on the urban environment [2–4, 14]. There was a problem of estimating ${}_{\text{urb}}$, which cannot be explicitly determined from the OSM data, so some empirical assumptions were required [2, 3]. To avoid this, the CGLC data were used to estimate ${}_{\text{urb}}$. The unique feature of the CGLC data is the coding of the area fractions if the different land cover types, including the built areas. However, in the CGLC data such territory includes vegetation inside the built areas (lawns, yard landscaping, etc.). Since the urban surface in TERRA_URB is considered completely “sealed,” the CGLC data should be corrected by removing vegetation from the area of built territory. Data from Sentinel-2 satellite were utilized to identify vegetation in the urban environment (the image for August 28, 2019). Based on the threshold value of NDVI (Normalized Difference Vegetation Index) > 0.6, the vegetation pixels were identified [15]; then, based on the threshold value of green-channel brightness, they were discriminated between trees (500) and grass or bushes (>500).

THE METHOD FOR COMPUTING CITY-DESCRIPTIVE PARAMETERS AND ITS APPLICATION TO MOSCOW MEGACITY

The proposed algorithm for the determination of the required city-descriptive parameters includes three stages:

—the preparation of OSM data using the GIS tools: the conversion of linear road and street network objects to polygonal ones with account of the class and number of lanes, the generalization of land surface types, the elimination of the mutual intersections of the polygons [3, 12];

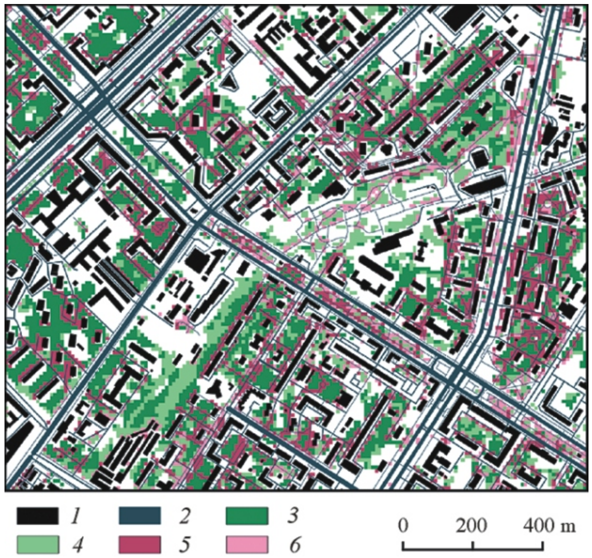

Fig. 1. The example of identification of vegetation in the urban environment and masking of pixels of vegetation that intersects with buildings and roads. (*1*) Buildings; (*2*) roads; (*3*) trees; (*4*) grass and bushes; (*5*) trees (masked); (*6*) grass and bushes (masked).

Table. The parameters for the model grid cells determined from the GIS analysis

Designation	Initial data	Parameter description
$_{\text{CGLC URB}}$	CGLC	Fraction of urbanized (built) territory
$_{\text{OSM BLD}}$	OSM	Fraction of buildings
$_{\text{OSM ROADS}}$	OSM	Fraction of roads
$_{\text{OSM GREEN}}$	OSM	Fraction of green spaces according to OSM data
L_{OSM}	OSM	Average number of floors for the buildings for which it is known
$_{\text{OSM LV}}$	OSM	Fraction of buildings with a known number of floors relative to their total area
$N_{\text{OSM BLD}}$	OSM	Total number of buildings in a grid cell
$N_{\text{OSM LV}}$	OSM	Number of buildings with a known number of floors in a grid cell
$_{\text{SENT GRASS}}$	Sentinel-2 and OSM	Fraction of grass vegetation
$_{\text{SENT TREES}}$	Sentinel-2 and OSM	Fraction of wood vegetation

—the masking (identification) of vegetation pixels that intersect with roads and buildings due to their overlapping by tree canopy and due to data discreteness (Fig. 1); the issue of interpretation of such intersections remains debatable: at the current research stage, it is proposed to neglect such pixels when calculating the ground fraction covered by vegetation; however, alternative ways of their account will be considered in further studies;

—the determination of the set of intermediate parameters (see the Table) for the grid cells (the 200 200-km COSMO grid with a spacing of 500 m including Moscow and most of the Moscow region was used) and the calculation of the final required parameters based on them.

The values of $_{\text{urb}}$ were determined from the CGLC data corrected with account of information about vegetation, buildings and roads. It was taken into account that $_{\text{urb}}$ cannot exceed the fraction of the vegetation-free cell area and cannot be smaller than the fraction of roads and buildings:

$$ {}_{\text{urb}} \quad \max[\ {}_{\text{OSM BLD}} \quad {}_{\text{OSM ROADS}};\ \min({}_{\text{CGLC URB}};\ 1 \quad \max({}_{\text{OSM GREEN}};\ {}_{\text{SENT GRASS}} \quad {}_{\text{SENT TREES}}))]. \qquad (1) $$

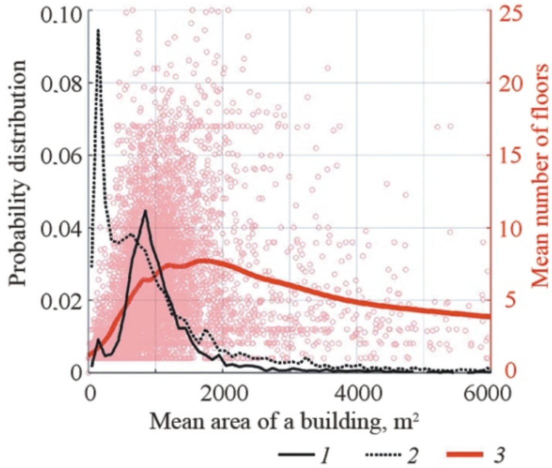

Fig. 2. The probability distribution function for the average area of a single building with a (*1*) known and (*2*) unknown number of floors in the grid cells (weighed by the area occupied by the respective buildings and normalized by the total area of all buildings, i.e., the area under the solid and dotted curves in the graph is equal to 1 in total) and (*3*) the dependence of the number of floors on the area of the single building *f*(*S*).

The designations and description of the terms are given in the Table. If the fraction of buildings is less than 2%, the value of $_{\text{urb}}$ was assumed equal to zero. This excluded territories with the sparsest buildings which do not exert essential influence on the atmosphere, and territories for which information on buildings is not available in the OSM data.

The height of buildings was determined based on the information about the number of floors L_{OSM} that is available as one of the attributes in OSM. However, it is available only for 20% buildings in the study region, which occupy 34% of their total area. The number of floors is usually unknown for private houses, garages, and technical structures with a small height and area (the ratio of the area of buildings with the known and unknown number of floors for buildings of different sizes is shown in Fig. 2). To retrieve information on the number of floors, we firstly determined the mean area of the building with the known number of floors for each cell:

$$S_{\text{LV}} \quad S_0 \quad _{\text{OSM BLD}} \quad _{\text{OSM LV}} / N_{\text{OSM LV}} \tag{2}$$

where S_0 is the grid cell area. The mean area of the building with the unknown number of floors S_{noLV} was determined in a similar way. Based on data on L_{OSM} and S_{LV}, the model $L = f(S)$ was obtained, which is a local regression dependence of the number of floors in a building on its area (Fig. 2). Based on it and on S_{noLV}, the number of floors was retrieved for the buildings for which it was unknown. Despite the great scatter of dots in Fig. 2, it is possible to discriminate between low buildings with an area of 100–300 m^2 and higher buildings. The mean height of buildings H was determined assuming that the height of one floor H_{L} = 3 m and the height of the foundation and attic H_0 = 2 m:

$$H \quad H_{\text{L}}[L_{\text{OSM}} \quad _{\text{OSM LV}} \quad f(S_{\text{noLV}})(1 \quad _{\text{OSM LV}})] \quad H_0. \tag{3}$$

To evaluate the ratio of the aspect of urban canyons (H/W), it is also necessary to estimate their average width W. The original method for the identification of canyons in the real buildings [12] was applied to estimate their width in the previous studies by the authors. The alternative methods often used in foreign practice [10] are based on simpler analytic considerations. For example, assuming that buildings in a grid cell are situated regularly within its urbanized part and have the same size and are squared, the canyon width is determined as follows:

$$W \quad \sqrt{S_0} \; \sqrt{\quad _{\text{urb}0}} \quad \sqrt{\quad _{\text{OSM BLD}}} \Big/ \sqrt{N_{\text{OSM BLD}}} \tag{4}$$

where $_{\text{urb0}} = \max[\ _{\text{CGLC URB}};\ _{\text{OSM BLD}} + \ _{\text{OSM ROADS}}]$. The parameter $_{\text{urb0}}$ is used instead of $_{\text{urb}}$, because the exclusion of vegetation from the built area leads to a decrease in the width of canyons, filled by vegetation, and to the distortion of geometric characteristics of the built environment. The comparison of W estimated using (4) with the width of canyons identified in the real built environment by the method from [12]

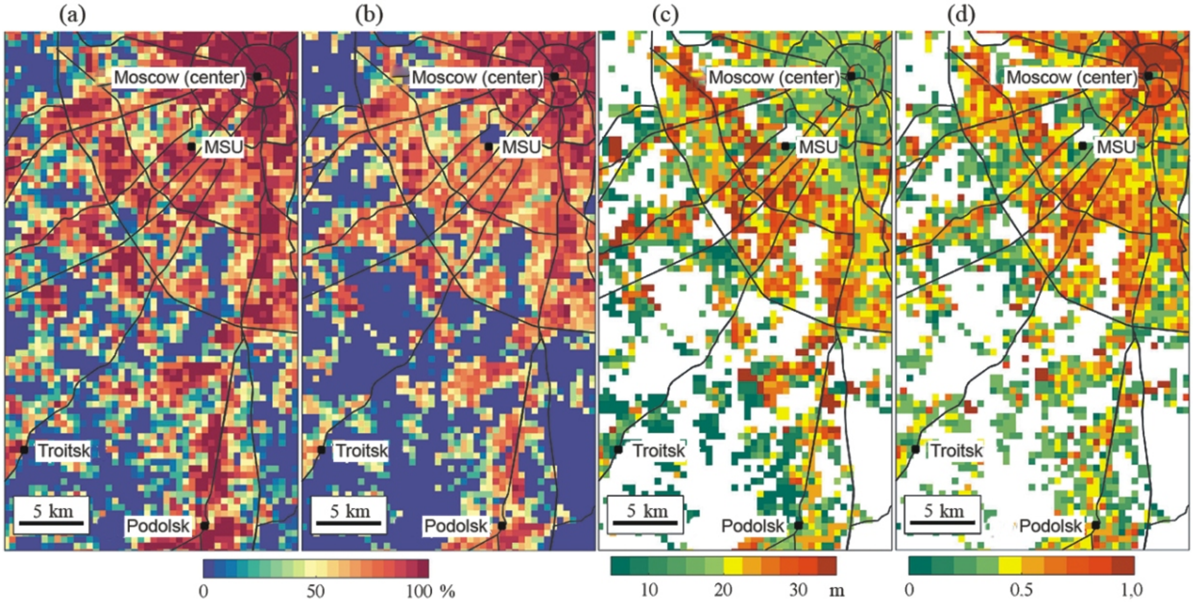

Fig. 3. The values of some city-descriptive parameters computed on the grid with a spacing of 500 m for the segment of the study area (southwestern Moscow): the fraction of urbanized territory based on (a) initial and (b) corrected CGLC data, (c) the mean height of buildings, and (d) the mean ratio H/W. The dark gray lines are the main highways.

revealed their good agreement, with the correlation coefficient of 0.81 (by the example of the most urbanized 1-km grid cells). For this reason, as well as due to the low sensitivity of the model to the parameter H/W [16], we further determined this parameter on the 500-m grid based on equation (4) which is much simpler for mass calculation. The fraction of urbanized territory $_{\mathrm{urb}}$ was determined as $_{\mathrm{OSM\,BLD}}/_{\mathrm{urb0}}$ (the use of $_{\mathrm{urb0}}$ is explained above).

The exact determination of Q_{h} with the resolution of hundreds of meters remains a difficult and unsolved problem. In this study we use a simple approach, which proved itself well in the previous studies [2–4, 14]. According to literature data [13], the average annual value of Q_{h} with the borders of Moscow (excluding the new areas joint to Moscow in 2011) is 52 W/m^2. To obtain a detailed spatial pattern, this value was "weighted" by the product $_{\mathrm{OSM\,BLD}}\,H$, that gave the variability of Q_{h} on the 500-m grid from 0 to 250 W/m^2, with the same prescribed average value.

The spatial distribution of the values of $_{\mathrm{urb}}$, H, and H/W on the 500-m grid is shown in Fig. 3 for the fragment of the territory in the southwest of Moscow. Certainly, the accuracy of the estimates is an important issue. It depends both on the accuracy of initial data and on the assumptions made. Initial OSM data were verified before by the comparison with reference information from official sources that showed their agreement in presenting built areas [11]. There are no reference values for the derived parameters, so only their qualitative assessment is possible, in particular, in the form of comparison with alternative data sources. The obtained values of $_{\mathrm{urb}}$ and Q_{h} demonstrate a more detailed and realistic pattern as compared to the default values from the global databases used in COSMO [16]. The values of $_{\mathrm{urb}}$ indicate better the "sealed" nature of grid cells as compared to the CGLC data (Figs. 3a and 3b). For example, $_{\mathrm{CGLC\,URB}}$ reaches 100% for residential areas in the southwest of the city and more than 50% for a number of suburban areas. The values of $_{\mathrm{urb}}$ for these territories make up 70–80 and 10–30%, respectively, that seems to be more realistic. Alternative estimates for the parameters H and H/W are absent, so only the consistency of the spatial pattern to expectations may be noted: the values of H are minimal in the historical center and suburbs, H/W is maximal in the center and in the areas with high buildings (Figs. 3c and 3d). The estimates of the spatial variability of Q_{h} qualitatively agree with literature data [1]. More reliable conclusions on the quality of obtained data can be made after the analysis of the results of simulations based on this data. Such research is already underway, and the results will be published later.

CONCLUSIONS

The method for computing the city-descriptive parameters was developed for the needs of meteorological modeling (by example of the TERRA_URB scheme in the COSMO model). It differs from the previous methods by the joint use of open global cartographic and satellite data. It was used to obtain the set of parameters for the Moscow region on the 500-m grid which has already been used in test model simulations and is promising for use in operational NWP task within the framework of COSMO-Ru system. The obtained data may also be useful for other various urban climate studies. The use of the initial data with global coverage allows an easy application of the method to other regions and grids. The priority directions of its further development were identified: a more accurate consideration of differences between the built and "sealed" surfaces, a refinement of approaches to the determination of the building height, the width of urban canyons, and anthropogenic heat flux. The further studies of uncertainty in the calculated city-descriptive parameters and its effect on the simulation results are also needed.

ACKNOWLEDGMENTS

The authors thank G.S. Rivin, I.A. Rozinkina, A.V. Kislov, and P.I. Konstantinov for valuable advice and organization of work at different research stages.

FUNDING

The research was supported by the grant for the implementation of activities on the development of the system for monitoring, forecasting, and warning of severe and adverse weather events in Moscow city (the Resolution of the Government of Moscow No. 257-PP, April 3, 2018) and was partly supported by the Russian Foundation for Basic Research and by the Government of Moscow (grants 19-35-70009 and 18-35-20052).

REFERENCES

1. G. G. Aleksandrov, I. N. Belova, and A. S. Ginzburg, "Anthropogenic Heat Fluxes in the Capital Agglomerations of Russia and China," Dokl. Akad. Nauk, No. 1, **457** (2014) [Dokl. Earth Sci., No. 1, **457** (2014)].
2. M. I. Varentsov, T. E. Samsonov, A. V. Kislov, and P. I. Konstantinov, "Simulation of Heat Island over the Moscow Agglomeration with the COSMO-CLM Regional Climate Model," Vestnik Mosk. Un-ta, Ser. 5: Geografiya, No. 6 (2017) [in Russian].
3. *Climate of Moscow under Global Warming*, Ed. by A. V. Kislov (MGU, Moscow, 2017) [in Russian].
4. G. S. Rivin, R. M. Vil'fand, D. B. Kiktev, I. A. Rozinkina, K. O. Tudrii, D. V. Blinov, M. I. Varentsov, T. E. Samsonov, A. Yu. Bundel', A. A. Kirsanov, and D. I. Zakharchenko, "The System for Numerical Prediction of Weather Events (Including Severe Ones) for Moscow Megacity: The Prototype Development," Meteorol. Gidrol., No. 11 (2019) [Russ. Meteorol. Hydrol., No. 11, **44** (2019)].
5. G. S. Rivin, I. A. Rozinkina, R. M. Vil'fand, D. B. Kiktev, K. O. Tudrii, D. V. Blinov, M. I. Varentsov, D. I. Zakharchenko, T. E. Samsonov, I. A. Repina, and A. Yu. Artamonov, "Development of the High-resolution Operational System for Numerical Prediction of Weather and Severe Weather Events for the Moscow Region," Meteorol. Gidrol., No. 7 (2020) [Russ. Meteorol. Hydrol., No. 7, **45** (2020)].
6. A. Baklanov, C. S. B. Grimmond, D. Carlson, D. Terblanche, X. Tang, V. Bouchet, B. Lee, G. Langendijk, R. K. Kolli, and A. Hovsepyan, "From Urban Meteorology, Climate and Environment Research to Integrated City Services," Urban Climate, **23** (2018).
7. M. Buchhorn, M. Lesiv, N. E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, "Copernicus Global Land Cover Layers Collection 2," Remote Sens., No. 6, **12** (2020).
8. G. F. Garuma, "Review of Urban Surface Parameterizations for Numerical Climate Models," Urban Climate, **24** (2018).
9. M. Haklay, "How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets," Environ. Plann. B: Plann. Des., No. 4, **37** (2010).
10. V. Masson, W. Heldens, E. Bocher, M. Bonhomme, B. Bucher, C. Burmeister, C. de Munck, T. Esch, J. Hidalgo, F. Kanani-Suhring, Y.-T. Kwok, A. Lemonsu, J.-P. Levy, B. Maronga, D. Pavlik, G. Petit, L. See, R. Schoetter, N. Tornay, A. Votsis, and J. Zeidler, "City-descriptive Input Data for Urban Climate Models: Model Requirements, Data Sources and Challenges," Urban Climate, **31** (2020).
11. T. Samsonov and P. Konstantinov, "OpenStreetMap Data Assessment for Extraction of Urban Land Cover and Geometry Parameters Required by Urban Climate Modeling," in *Extended Abstract of Proceedings of the GIScience 2014, September 23–26, Vienna, Austria. GeoInfo Series*, Vol. 40 (Hochschu-lerschaft, TU Vienna, Vienna, Austria, 2014).

12. T. E. Samsonov, P. I. Konstantinov, and M. I. Varentsov, "Object-oriented Approach to Urban Canyon Analysis and Its Applications in Meteorological Modeling," Urban Climate, **13** (2015).
13. I. D. Stewart and C. A. Kennedy, "Metabolic Heat Production by Human and Animal Populations in Cities," Int. J. Biometeorol., No. 7, **61** (2017).
14. M. Varentsov, H. Wouters, V. Platonov, and P. Konstantinov, "Megacity-induced Mesoclimatic Effects in the Lower Atmosphere: A Modeling Study for Multiple Summers over Moscow, Russia," Atmosphere, No. 2, **9** (2018).
15. M. M. F. Wong, J. C. H. Fung, and P. P. S. Yeung, "High-resolution Calculation of the Urban Vegetation Fraction in the Pearl River Delta from the Sentinel-2 NDVI for Urban Climate Model Parameterization," Geosci. Lett., No. 1, **6** (2019).
16. H. Wouters, M. Demuzere, U. Blahak, K. Fortuniak, B. Maiheu, J. Camps, D. Tielemans, and N. P. M. van Lipzig, "The Efficient Urban Canopy Dependency Parametrization (SURY) v1.0 for Atmospheric Modelling: Description and Application with the COSMO-CLM Model for a Belgian Summer," Geosci. Model Dev., No. 9, **9** (2016).