Where does a settlement end? Defining urban areas for more than just runoff analyses

LUDĚK STROUHAL, PETR KAVKA

ABSTRACT

The article presents a methodology for the spatial delineation of significant urban areas in the Czech Republic, primarily for the purposes of hydrological analyses and flash flood risk assessment. First, the definition of urban areas was refined in relation to existing terminology, followed by the development of a comprehensive procedure for creating layers of urban and natural features using the planimetric layers of the Fundamental Base of Geographic Data of the Czech Republic (ZABAGED®) supported by additional open data sources. The methodology includes hierarchical classification and concurrent geometric processing of selected topographic layers, as well as subsequent cleaning and filtering of the final polygons of both urban areas and natural features. The result is a dataset of significant urban areas that supports better evaluation of rainfall-runoff processes in their vicinity and analysis of their internal structure, providing a starting point for designing and assessing adaptation measures. The outputs will soon be publicly available for further hydrological applications and research.

INTRODUCTION

One of the classic, yet still highly relevant tasks of applied hydrology is the analysis of flow paths and areas of surface runoff concentration, with the aim of predicting and preventing the impacts of pluvial flooding on settlements and infrastructure [1]. Although the basic principles of runoff analysis are straightforward and the development of GIS and data sources over the past 20 years has provided powerful tools for their objectification and automation, two major challenges await those undertaking large-scale territorial studies.

The first challenge relates to the extent, detail, and limitations of elevation data (terrain models) and the computational methods used to analyse them. Put simply, the goal is to ensure that water in the model 'flows' along paths that reflect reality and take into account both the terrain and built structures. The second challenge concerns the quality, interpretation, and processing methods of planimetric data. Determining a catchment area or the point where a runoff path enters a water reservoir is relatively straightforward, as there are data layers of moderate quality available for water bodies. However, when attempting to identify high-risk points where surface water may enter major settlements or, more generally, urban areas, researchers face the issue of how to define such areas and how to delineate the boundary between open landscape and settlement.

In neither general nor specialised planimetric databases of the Czech Republic does a dataset exist that delineates the boundaries of settlements or urban areas. The concept of the built-up area, or areas designated as built-up and developable, is used in spatial planning, but their spatial definition exists only in a decentralised form and is created as part of individual land-use plans. Land consolidation projects (LC), which are primarily focused on agricultural land, use the term *Internal boundary of the LC area*. This boundary delineates the extent of the land consolidation, including its interface with settlement areas, and is also prepared for the purposes of a specific project. However, for hydrological analyses at larger scales, these boundaries are neither readily accessible nor suitable for use. In a number of research projects addressing similar issues ("*Risk Maps Resulting from Flood Hazards in the Czech Republic*" – MoE, no. SMZP2007SP SP/1C2/121/07, "*Impact of Erosion on Water Bodies*" – MoI, no. BV VG20122015092, "*Enhancing the Preparedness of Urban Areas in the Czech Republic by Linking the Critical Point Method with a Flash Flood Indicator*" – TA CR, no. SS06010059), the delineation was therefore always derived by the researchers ad-hoc according to their own methodology (e.g. [2]) and the intended purpose. The level of detail in the documentation of the data sources and derivation methods used was highly variable and not always satisfactory.

This article focuses specifically on the second challenge highlighted by runoff studies, i.e. clarifying the boundary between settlements and the surrounding landscape from the perspective of the hydrological regime and potential adaptation to climate change. Approaches to managing rainfall, as well as the risks and options for adaptation measures associated with extreme precipitation, differ between open countryside and settlements. The boundary between them is not sharply defined and depends, among other factors, on the size and fragmentation of urban areas. This paper describes the delineation of only *Hydrologically significant urban areas*; however, for the sake of readability, the shortened term *Urban areas*, or its abbreviation UA, will be used in the following text. The delineation of their boundaries utilised detailed and openly accessible national data from the Basic Geographic Data Database of the Czech Republic (ZABAGED®) [3], supported by additional specialised data sources (Digital Technical Map, Register of Territorial Identification, Addresses and Real Estate, etc.). This article presents, in limited detail, the procedures used to define the boundaries of UA, which were created as auxiliary outputs within the framework of the TA CR project no. SS06010386 "*Adaptation of Urban Areas to Flash Floods and Drought*". Parts of the text are based on the interim report on the project's progress for 2024 [4].

METHODOLOGY

Definition of the built-up area of a settlement and urban areas

Before spatially delineating the boundaries of an urban area, it is first necessary to define it as precisely as possible. Despite the common use (*in Czech, translator's note*) of the term 'intravilán' (of settlements), it is not a terminologically

well-defined concept. It also does not translate well into English – it is simply 'built-up area' or 'inner urban'. There are differences between the property-law perspective and the spatial planning or urbanistic perspective. From a legal standpoint (Act No. 128/2000 Coll., on municipalities), intravilán is defined as the built-up area of a municipality that was recorded as intravilán in the cadastral map as of 1 September 1966. This is a historical legal term, primarily relevant in connection with soil fund protection. Both older and current building legislation (Act No. 183/2006 Coll. and Act No. 283/2021 Coll.) no longer use this term directly; instead, they operate with the concepts of built-up area (zastavěné území; ZÚ) and developable area (zastavitelné plochy; ZP), which are delineated in the spatial plan. According to the wording of § 116 of Act No. 283/2021 Coll., the built-up area (ZÚ) includes:

— built-up building plots,
— building gaps,
— other fenced gaps between built-up building plots,
— public spaces,
— roads or parts thereof, including entrances to other plots within the built-up area and railway tracks where they pass through the intravilán and other plots within the built-up area,
— other plots that are surrounded by other plots of the built-up area, excluding vineyards and hop gardens.

From an urban planning and spatial development perspective, the built-up area is understood as the compact part of a municipality or town where built-up functions predominate (residential, commercial, industrial). Considering possible spatial arrangements, settlements can be distinguished as compact (the main contiguous built-up area of the municipality), dispersed settlements, and isolated homesteads. The spatial delineation of urban areas itself is a relatively complex and expert process within this field. For example, the *Methodology for the Identification and Classification of Areas with Urban Values* [5], which establishes an objective framework for identifying and classifying areas with urban values as one of the phenomena monitored for Territorial Analytical Documents, recommends the use of a combination of diverse cartographic materials and results from field surveys. The precise delineation is left to expert assessment and analysis of the character of the specific area.

From both of the above-mentioned approaches (property-law and urban), it is evident that delineating the built-up area is not an easy task, where individual classes of topographical features could simply be grouped without regard to other spatial relationships with their surroundings. The definition of (*hydrologically significant*) *Urban areas* applied in the aforementioned project no. SS06010386 is also based on a combination of these approaches. It refers to spatially extensive compact areas consisting of contiguous zones with residential or commercial-industrial functions, along with smaller internal islands or plots of a predominantly natural character. Areas of UA defined in this way have, compared to the surrounding *Natural features* (NF), despite including a certain proportion of anthropogenic areas, a different management of stormwater, risks, and impacts of fluvial and pluvial floods, and employ different methods of mitigation. Threshold values and criteria for classifying objects or their classes were established during the development of the methodology on a test area exceeding 2,300 km², which included parts of the capital city Prague, medium-sized and smaller towns, as well as dispersed village settlements. The GIS analyses described below were further processed using the ArcGIS Pro environment.

Processing topographic objects and iteration of the natural features layer

The methodology for creating the data layer of urban areas was developed using positional data from the ZABAGED® database from spring 2024 concerning the test area and was applied to the entire Czech Republic using the updated content of the same database from December 2024. After an initial analysis of the spatial relationships between individual object classes (layers) and their attribute sets, a gradual classification of the positional data layers or their individual objects into natural and urban areas was carried out, starting with the most numerous and clearly natural features and progressing to less numerous and ambiguously classifiable layers, including possible adjustments to the geometry of certain objects. The gradually expanded layer of natural features (NF) was used to classify objects of ambiguous classes, with one of the main criteria being the ratio of each object's shared boundary with this layer. At the end of the classification process, the selected urban areas were aggregated into an initial approximate layer of urban areas (UA). In the final phase, topological cleaning and filtering were performed based on the size and significance of the resulting UAs. In various parts of the process, besides ZABAGED, selected classes of objects from the Register of Territorial Identification, Addresses, and Real Estate (RÚIAN) [7] and the geometric delineations of land parcels recorded in the Land Parcel Identification System (LPIS) [6] were also used.

Classification of overlapping ZABAGED layers

In the Czech context, the indispensable topographic database ZABAGED contains, in terms of spatial relationships, two types of polygon object classes: overlay and base (background). All base classes represent continuous areas with relatively homogeneous ground cover or type of use (e.g., permanent grassland, functional development) and together create a topologically clean, continuous representation of the entire territory of the Czech Republic. In contrast, overlay classes most often represent man-made objects of various kinds (buildings, castles), but also natural features (swamps, peat bogs), which overlap with one or more base layers. In terms of their mutual positions, they are disjoint, but sharing a common boundary is not excluded. Relevant overlay objects provide valuable information when deciding on the classification of certain base layer objects into urban areas and defining their boundaries. Following an exploratory analysis, relevant classes were selected from the overlay layers according to *Tab. 1*. All base layers were used to ensure the integrity of the derived UA/NF polygons.

Tab. 1. Selection of relevant ZABAGED overlay layers for sorting of the base layers

Relevant	Irrelevant
Individual building or block of buildings	Swamp, marsh
Castle	Cooling tower
Shed, greenhouse, polytunnel, shelter	Heliport
Covered structure	Above-ground storage tank
Chateau	Ground storage tank
Airport runway perimeter	Peat bog
	Ruin, remains
Irrelevant — regions	Landslide, scree
Small-scale protected area (ZCHÚ)	Rock formations
Large-scale protected area (ZCHÚ)	Grandstand
Special protection area	Tower-like structure
Special area of conservation	Railway turntable, transfer table

As is evident from their names in *Tab. 1,* even the selected overlay object classes do not necessarily have the same relevance when deciding whether to classify a base layer object as part of a significant urban area or not. The most problematic class is the largest layer, *Individual building or Building block* (hereafter *Buildings*), which includes an enormous number of objects of varying sizes and importance. The attribute data of this layer contributes only partially to their differentiation, as less than 2 % of nearly 3.9 million objects have a defined specific building type from about 40 categories; the remainder of the objects are classified as *Unspecified buildings*. Therefore, a supplementary attribute called *ZB_podklad*, named after the base ZABAGED layer, was first assigned to these objects.

Because many of the background layers (e.g. *Functional development area*) include more detailed information on land use, which can play a role in classification of *Buildings*, this specific attribute was transferred to an additional field called *ZB_detail*. Subsequently, *Buildings* were classified according to these three attributes, and where applicable, by the auxiliary criterion of their area, into three levels of significance (0–2) based on a key, which is not detailed here due to its excessive length. In general terms, buildings of a rather non-urban type (e.g. pumping stations, livestock farms, etc.) are initially marked as insignificant (significance = 0), while most of the remaining specific types are assigned significance level 2. Furthermore, buildings that are unspecified but located on a clearly non-urban background (such as forest land, solar power plant) are also marked as insignificant, with varying area limits of 16/500/1,000 m² applied in cases of ambiguous data. Specific combinations of characteristics were used for smaller buildings situated on the *Functional development area* layer. Remaining unspecified buildings were assigned a significance level of 1. In the following phase, significance of *Buildings* was taken into account when classifying objects in the background layers; in some cases only significant *Buildings* were considered, while in specific cases all *Buildings* were used.

Sorting and adjustment of base ZABAGED layers

All polygon object classes in ZABAGED – both overlapping and background – were analysed during the exploratory phase in terms of the existence, reliability, and relevance of their internal attributes. Careful classification of more than 60 categories of *development types* required the use of objects from the *Functional development area* layer, which includes similar but not identical categories to the *building type* from the overlapping *Buildings* layer.

Objects from the *Arable land* layer and *Other unspecified areas* – specifically one of its two categories, *Other unspecified areas* – as well as the layers *Other area in settlements* and *Ornamental garden, park*, were found to be highly problematic and required advanced approaches when deriving urban areas (UA).

From the linear object classes in ZABAGED, datasets related to transport infrastructure were used, specifically: *Motorway, Unregistered road, Road under construction, Street, Railway line,* and *Railway siding*. In cases where a *Tunnel* object was found to run concurrently along the route of any of these objects, the covered sections were removed from the dataset.

The process of classifying and modifying individual ZABAGED background layers for subsequent compilation into the final UA/NF layer was tested and then automated using the graphical programming environment ModelBuilder, combined with simple Python scripts within ArcGIS Pro. A set of approximately 30 interconnected tools was created to support a more transparent workflow and allow for intermediate product checks. Rather than directly aggregating objects that make up urban areas, the proposed method focuses on the inverse task – identifying which objects represent typical natural features and, with a high degree of certainty, do not belong to urban areas. Objects with ambiguous classification are then gradually analysed and sorted into either natural features (NF) or candidates for inclusion in UA, based on the proportion of their shared boundary with a NF. Geometric adjustments are made throughout to complex objects, and finally, all potential UA objects are aggregated. Both working layers – NF and UA – are maintained in two variants throughout the process: the MERGE variant preserves internal object boundaries and retains the relevant portion of their original attribute set, supplemented with a source attribute indicating the name of the parent input layer. The DISSOLVE variant consists only of merged polygons without internal boundaries, containing overall or average characteristics of the area (e.g. area size, total surface area and building footprint, shape index, etc.). A brief description of the processing of background layers follows below; a detailed explanation exceeds the scope of this article.

Hexagonal grid for decomposition of complex objects

For analysis and adjustment of complex topographic classes, an auxiliary polygon grid of regular hexagons – hexes – was created at two scales. In the base resolution, hexes were defined with an area of 400 m², corresponding to a height of approximately 21 m. In the lower-resolution version, the hex height was set to 60 m. In both cases, the hex grid covered the entire territory of the Czech Republic with a slight overlap. To reduce computational demands, hexes located entirely within polygons of arable land and all forest land classes were removed, each reduced by an 80 m buffer. To support spatial analysis, hexes containing significant *Buildings*, *Castle*, or *Covered structures* were then marked with attribute tags, as were their neighbouring hexes. Optionally, hexes containing all *Buildings*, and those adjacent to them, were also marked.

Tab. 2. Buffer widths for linear transport infrastructure according to type for purpose of their separation from Other unspecified areas

Road type	Code	Buffer [m]
1st- or 2nd-class motorways	D1, D2	18
Roads for motor vehicles	M	15
Branches and spurs of 1st- and 2nd-class motorways	D1v, D2v, D1p, D2p	9
Branches and spurs of roads for motor vehicles	Mv, Mp	7.5
1st-class roads	S1	7
2nd-class roads	S2	5
3rd-class roads and unregistered roads	S3, –	4
Branches and spurs of 1st-class roads	S1v, S1p	3.5
Branches and spurs of 2nd-class roads	S2v, S2p	2.5
Railway line and siding	–	2.5
Branches and spurs of 3rd-class roads	S3v, S3p	2

Other unspecified areas

In the first step of processing the base layers, adjustments were made to the *Other unspecified areas (OUA)* within the polygon layer *Arable land and other unspecified areas*. As shown in *Fig. 1, OUA* include both linear land objects such as roads in non-urban areas, along with adjacent greenery, and polygonal elements that may potentially contain buildings or other anthropogenic structures. These polygonal elements were automatically separated for potential later inclusion in the UA layer. First, uniform buffers around linear road objects were extracted from the *OUA* objects, as well as any overlaps with the LPIS layer. The buffer widths listed in *Tab. 2* were determined for individual road classes based on a review of technical standards, experience from previous projects, and random verification using aerial imagery. The areas remaining after the separation of road buffers were subsequently divided using the basic hexagonal grid. For each fragment, the relative position with respect to the edge of the original polygon, to the road, and an indication of the presence or proximity of *Buildings* was determined. Fragments

with a suitably chosen combination of the described attributes were re-aggregated into a contiguous unit (see the red-hatched object in *Fig. 1*), and the total area of included *Buildings*, shape index, etc., were calculated. Finally, the resulting polygons were filtered and classified – based on experimentally determined parameter combinations – into potential UA elements and elements to be incorporated into the NF layer.
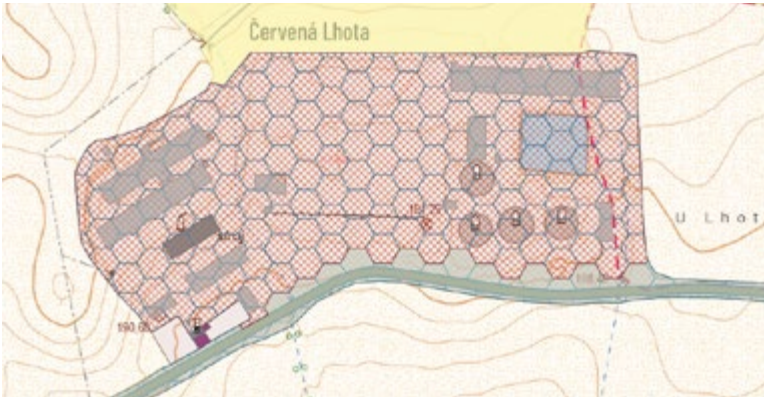

Fig. 1. *Other unspecified areas* including both the areal feature of anthropogenized area (red hatch) and the pseudo-linear road parcel (grey) subject to automated separation

Permanent grassland

Before creating the first iteration of the NF layer, a hypothesis was tested that, in the case of compact development, buildings are always situated on a typically urban-type base layer, such as *Other settlement area*, *Ornamental garden/park*, *Orchard/garden*, etc. This assumption proved invalid in the case of the *Permanent grassland (TTP)* layer. Although this layer includes more or less maintained grassy areas in the non-urban area, it also fills small gaps within settlements or even serves as the base layer for relatively compactly arranged *Buildings* in settlements located in specific landscape types, such as mountain villages. An example of such a settlement is shown in *Fig. 2*, where it is evident that a blanket classification of *TTP* as natural feature in these locations would lead to unacceptable fragmentation and the subsequent exclusion of urban areas that in reality form a fairly compact settlement centre. Therefore, significant segments of *TTP* within a 20 m buffer around *Buildings* were excluded, with the selection criteria being an area larger than 0.25 ha, the presence of at least three buildings, and contact with or proximity to objects from the layers *Other settlement area*, areal features with buildings, and polygons of roads extracted from *OUA*. The remaining *TTP* areas were then incorporated into the first iteration of the NF layer.

Fig. 2. Significant clipped areas of permanent grassland (deep yellow) for inclusion in urban areas; light yellow is open grassland and shaded insignificant clipped areas with scattered settlements

Tab. 3. Selection of base layers or their subsets for the first iteration of the Natural Feature layer

Base layer	Category / subset
Orchard, garden	Category Orchard and other permanent crops
Arable land and other unspecified areas	Category Arable land
OUA roads	Road polygons extracted from OUA
OUA without buildings outside roads	Areas extracted from OUA
OUA LPIS	Overlap of OUA and LPIS excluding buildings
Permanent grassland other	Permanent grassland without significant buffers around buildings
Forest land with trees	
Forest land with shrubs	
Forest land with dwarf pine	
Hop field	
Water body	
Surface mining, quarry	
Storage site	

Ornamental parks and gardens

In the third step, the first iteration of the polygon layer for NF was created, incorporating either entire classes of objects or their subsets according to *Tab. 3*. This was followed by a rather complex processing of the *Ornamental parks and gardens* feature (hereinafter *Parks*), which unfortunately lacks any relevant internal attributes and includes mainly urban residential greenery, but also larger maintained suburban green areas such as Průhonice Park. These polygons often contain more or less significant buildings, both isolated and in groups, as well as numerous internal islands (especially around water bodies), which complicates spatial analyses and decisions regarding the classification of the object as part of either UA or NF, as shown in *Fig. 3*. First, polygons in *Parks* that are potentially natural (*P_Park*) are separated if their shared boundary with NF accounts for at least two-thirds of their outer perimeter (excluding internal islands). This forms the second iteration of the NF layer. From the potentially settlement-related *Parks* (*U_Park*), 75 m buffers around significant buildings and *Castles* are extracted. Due to frequent overlaps with neighbouring polygons, only those parts in direct contact with buildings or near *Chateau* are isolated, while buffer areas forming isolated islands within the original *U_Park* are disregarded. Small fragments of *U_Park* or fragments without contact with NF are reassigned back to the subsets around *Buildings*. For large fragments and those adjacent to PP, the length of their outer perimeter (again excluding internal islands) and the proportion of this perimeter shared with NF are calculated. Golf courses (a category of the *Functional development area* layer) are temporarily included in the NF set. The shared perimeter is determined using a 10 m buffer around NF to include cases where parks are separated from the non-urban area by only a narrow road (most often represented by the *Other settlement area* base layer). Large fragments with less than 20 % shared perimeter with NF form the final component that, together with the building buffers, constitute the resulting *U_Park*. The remaining objects, along with the primarily designated *P_Parks*, are assigned to the NF layer, creating its third iteration.

Fig. 3. Průhonice park as the biggest challenge for processing and sorting *Ornamental parks and gardens (green objects with circles)*. A typical natural feature, only sharing a high percentage of perimeter with residential elements and containing a number of isolated buildings and island objects; light green elements were included into the first version of *Urban areas* layer

Gardens

After processing the *Parks*, the objects from the *Orchard, garden* layer, specifically its *Garden* category, were classified (the remaining categories, *Orchard* and *Other permanent crops*, had already been included in NF in previous steps). From the perspective of inclusion in UA, this category is also ambiguous, as it encompasses both family house plots within contiguous settlement areas and scattered complexes of suburban satellites, villages, or former allotment gardens whose function has gradually shifted to residential use (*Fig. 4*). Compared to *Parks*, however, these objects are much simpler and more compact, which was reflected in the classification approach applied to them. First, gardens not containing centroids of significant *Buildings* (*Gardens without buildings*) were isolated, and their objects with a boundary shared with NF over two-thirds were

assigned to the fourth iteration of NF. Again, internal islands within the polygons were disregarded. From the remaining *Gardens without buildings*, objects with no contact with any *Building* polygons and with a boundary shared with NF over 40 % were further removed and assigned to NF, resulting in the fifth iteration of this layer. The remaining *Gardens without buildings*, as well as all gardens containing centroids of significant *Buildings*, proceeded to further processing as potential UA features.
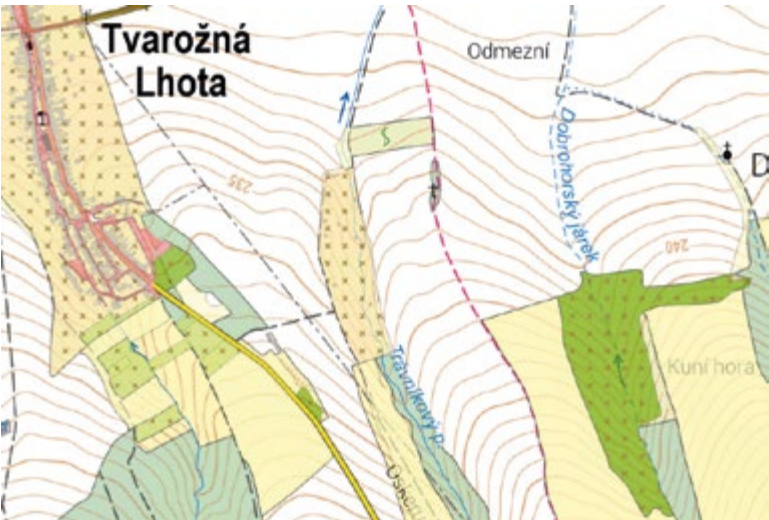

Fig. 4. The layer *Orchards and gardens* (objects with crosses) consists of orchards (light green) classified as *Natural feature and gardens* differentiated according to the built-up area to be classified as NF (dark green) or UA (ochre)

Classification of simple features

In the next step, a relatively straightforward classification of a number of features was carried out purely based on their shared boundary with NF. Features with a proportion of natural boundary exceeding 66 % (or 50 % in specific cases) were excluded from the potential UA features and incorporated into the next

iteration of the NF layer. The following features were processed in this manner: *Vineyards, Cemeteries, Substations, Power stations, Landfills, Car parks/Rest areas, and Pipeline pumping stations*. In the case of *Power stations*, it was necessary to merge neighbouring features by type in advance, due to the occasional presence of multiple adjacent objects. This was followed by a more complex processing of features from the *Airport* layer, in order to distinguish small rural airfields with few features and often natural runways from larger airports with complex infrastructure and a significant degree of surface anthropogenization, such as Vodochody Airport shown in *Fig. 5*. First, surface type information was assigned to the *Airport runway boundary* features based on the attribute in the *Runway axis* feature. Next, a 40m buffer was clipped from the *Airport* polygons around all *Buildings*, including *Shed, Greenhouse, Polytunnel*, and *Shelter* features. A convex hull was then defined around these subsets and the paved runways, and used to extract a portion of the original *Airport* polygons. Small fragments were merged with adjacent polygons. Larger fragments (often nearly entire airports) were considered natural features and incorporated into what is now the thirteenth iteration of the NF layer. In the next, fourteenth iteration, anthropogenized subsets of the airports were finally classified into UA and NF category, this time using stricter criteria: at least 80 % natural boundary share and a maximum area of 10 ha for inclusion in the NF category.

Tab. 4. Specific treatment for groups of development types in the object Functional development area

|	
Unconditionally urban areas *included in UA*	Castle grounds; bus station; other health and social facilities; depot; wood processing and paper industry; fire station; playground; metallurgical industry; chemical industry; barracks and military facilities; monastery; swimming pool; other cultural building; summer stage; museum; hospital; other unspecified industry; swimming complex; police complex; printing industry; food industry; glass, ceramics, and building materials industry; port; open-air museum; warehouse; hangar; greenhouse crop cultivation; group garages; stadium; engineering industry; school; educational facility; technical services; textile, clothing, and leather industry; prison; underground water tank; exhibition grounds; amusement park.
Conditionally urban areas *clipping and based on surroundings*	Auto-moto-cycling area, botanical garden, campsite, golf course, holiday cottage area, meteorological station, recreational development, water treatment plant, zoo, safari.
Other types *classified based on surroundings*	Castle grounds (ruins); archaeological site; fuel filling station; wastewater treatment plant; racecourse, show jumping arena; deep mining; observatory; livestock breeding; church; dog training ground; areas for UL aircraft; sports complex; shooting range; transmitter; other agricultural area.
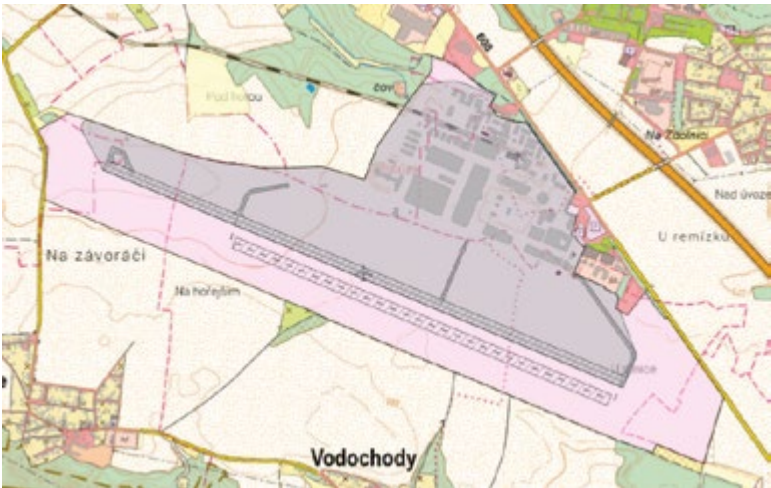
Fig. 5. Separation of anthropogenized (grey) and rather natural areas (pink) of Vodochody Airport

Functional development area

Processing the *Functional development area* layer presented numerous challenges. This layer contains a large number of areas with various types of use, specified in the attribute field *typzast*. A total of 61 development types were divided into three groups according to *Tab. 4*, with each group subsequently handled in a different manner. For so-called conditionally UA features, 100 m buffers were clipped around significant buildings. These subsets were then merged with the unambiguous (i.e. unconditionally UA) categories. The remaining types, along with spatially significant fragments from the conditional categories, were ultimately classified as either UA or NF categories based on the proportion of natural external boundary. To determine the shared boundary, the NF layer was expanded by 10 m in order to suppress artificial separation caused by roads or narrow polygons of *Other areas in settlements* (highlighted in red in *Fig. 6*).

Fig. 6. Separation of the anthropogenized part of the botanical garden (orange) from the areas of natural character (hatched) within the layer of the *Functional development area*; the blue background is an iteration of the *Natural Feature* layer

Other areas in settlements

The most problematic class of ZABAGED objects is undoubtedly the layer *Other areas in settlements* (hereinafter *OAS*), whose features cannot be unequivocally classified as part of settlements and, moreover, complicate the classification of surrounding layers. As shown in *Fig. 7*, despite its name, this layer also includes a large number of non-urban areas, but primarily problematic polygons along all types of roads. It is desirable to remove these and reclassify them into the NF layer, but due to their vast quantity, this cannot be done manually. For automating the process, an approach similar to that used for *OUA* was chosen. In this case, buffer polygons along roads were not generated, because a large portion of *OAS* run alongside the linear *Street* object, and clipping these would create an enormous number of problematic cases within settlements. The *OAS* layer was therefore divided using a regular hexagonal grid, and hexagons containing two or more segments of the external boundary of *OAS* were identified. These "source" hexes were then iteratively expanded by adding their neighbours until the resulting subset encountered proximity to a *Building*, a potential UA object, or hexes located deeper within the polygonal parts of *OAS*. The resulting subsets were further classified based on geometric attributes (number of hexes, shape coefficient, etc.). The entire process had to be performed twice with different scales of the hexagonal grid because, besides narrow polygons along streets and lower-class roads, the *OAS* layer also includes much wider strips around motorway and major access roads and feeder roads, which somewhat inconsistently are not classified in the *OUA* layer. Even hexes of 60 m size were insufficient to identify all cases, such as Prague Ring Road before Komořany Tunnel (*Fig. 7*). However, these polygons were filtered out from the UA layer at a later stage thanks to the successful separation of the narrow *OAS* strips using the described procedure.

Fig. 7. The complex object *Other areas in settlements* (red) includes linear and areal features both within and outside of settlements; unwanted linear features along various road types, removed by automated process, highlighted in black

Railway stations and rail yards

As the final base layers of topography, features of the classes *Railway station area, Stops* (hereinafter *Stop area*), and *Rail yard* were processed. Due to their frequent but not always simultaneous occurrence, as well as the specific nature of both types of features and their location relative to settlements, these layers were processed using slightly different methods. First, subsets of the *Stop area* within a 50 m buffer around features from the *Buildings, Covered construction*, and *Shed, Greenhouse, Polytunnel, Shelter* layers were performed. Remaining larger fragments in contact with NF were reassigned to this layer, and buffers around features were included in the UA layer only if they had contact with previously identified UA features, regardless of *Rail yards*. In this way, *Stop areas* isolated further from settlements were excluded. Finally, *Rail yard* features were processed using a similar approach employing hexagons as in the case of *OAS*.

Refinement of urban areas and gap infilling

After adjustment and classification of all base ZABAGED topographic layers, features identified primarily as settlement-related were aggregated into the first rough version of the *Urban area* layer, both in the MERGE and the DISSOLVE variant. Alongside this, the latest iteration of the NF layer also existed in both MERGE and DISSOLVE variant. Together, these layers covered the entire study area completely and without overlap. As shown in *Fig. 8*, the initial UA version (in orange) exhibits several shortcomings and complications for defining meaningful urban area boundaries. Notably, there is undesirable fragmentation of UA by watercourses where these are represented not only as linear features but also as corresponding polygons in the water body layer. Furthermore, the first iteration of the UA layer contains a large number of gaps – internal islands of varying sizes. This fact is evidenced by more than 36,000 polygons in the NF layer, of which over 33,000 do not exceed an area of 1 ha.

Fig. 8. First incomplete version of the UA layer with sub-areas separated by watercourses and with a number of gaps

Before performing any filtering of UA polygons, it was first necessary to remove fragmentation caused by watercourses. Therefore, a double buffering of UA feature hulls was performed – 40 m outward and 50 m inward – after which the area of the original UA features was removed. Within the resulting polygons larger than 0.25 ha, "linear" polygons of flowing water bodies, roads from the *OUA* layer, strips of *OAS*, and fragments of *Rail yards*, which had previously been assigned to the NF layer, were extracted and transferred to the second iteration of the UA layer.

Given the previously stated definitions of intravilán and built-up area, the process then proceeded with the controlled infilling of gaps in the following steps:
— Potential settlement features from the NF layer within gaps up to 5 ha were transferred to the UA layer.
— Features of all NF categories within gaps up to 1 ha were transferred to the UA layer.
— Features of all categories within gaps up to 5 ha located more than 40 m from other NF were transferred to the UA layer.

After the described infilling of gaps, especially due to the removal of "linear" interruptions caused by watercourses, the number of UA polygons slightly decreased to just over 97,000. However, the infilling of smaller gaps led to a dramatic reduction in the number of NF polygons, from the original 36,000 to only 1,120 features.
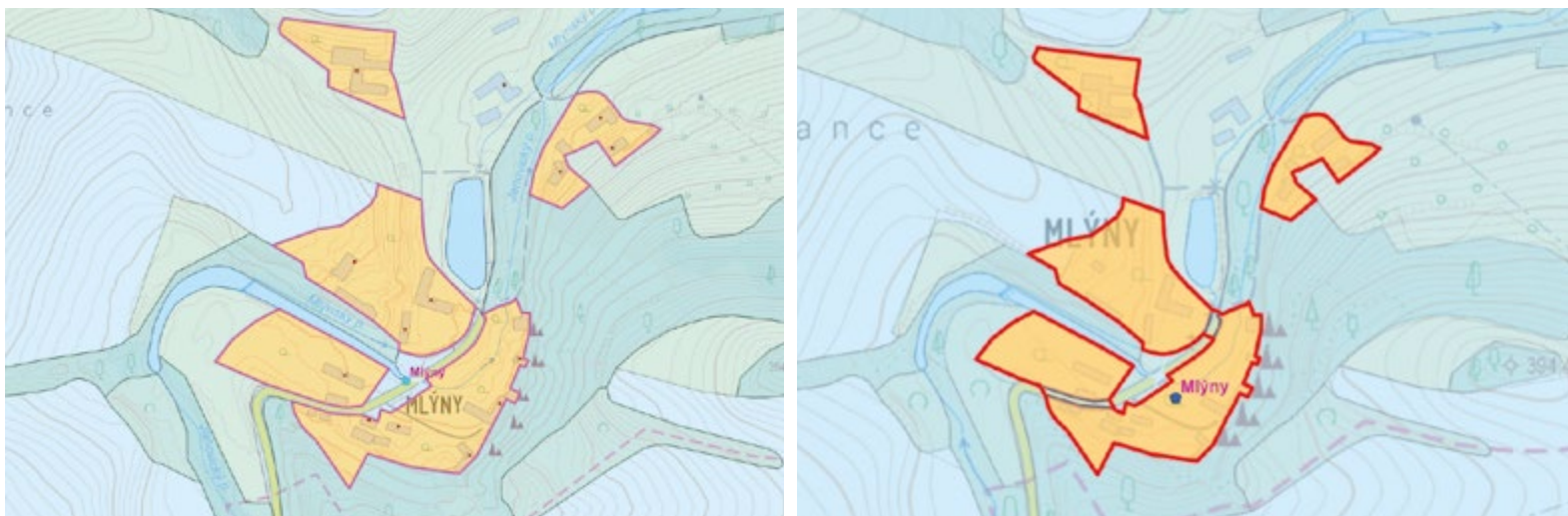
Fig. 9. On the left, separated polygons of the urban area of the Mlýny village with the definition point outside the derived settlement boundaries; on the right, polygons connected by the Aggregate polygons tool and after definition point correction

Filtering of significant UAs

As shown in *Fig. 8*, the derived UA layer contains a number of small, isolated polygons outside the main settlements. Compared to large contiguous urban areas, these patches have a negligible impact on the hydrology of the region; therefore, they were filtered out. In the first step, filtering was performed on UA features significant from the perspective of administrative units.

To determine significant UA features, besides size, the presence of a settlement's defining point was also considered. Within ZABAGED, two point features are available: *Defining point of administrative unit* and *Defining point of part of municipality*. The former proved too conservative, as too many compact UAs of significant municipalities would be excluded, so the latter, more detailed feature was used. This object also showed minor flaws, particularly positional inaccuracies – several points were clearly located outside the defined UA boundaries. Since it is a borrowed point layer, the original, more complete data source – the *Part of municipality* layer from the RÚIAN database – was used instead of ZABAGED. Spatial inaccuracies were also found here, with some defining points lying completely outside the delineated UA, with errors of up to hundreds of metres. For points with larger errors (distances from UA layer features), manual position corrections were therefore necessary.

A second complication for filtering insignificant UA features is shown in *Fig. 9*. Some otherwise compact settlements, significant enough to have their own defining point, consisted of several disconnected but nearby polygons in the UA layer, because the commonly used *Other areas in settlements* object was not delineated in the ZABAGED source data. Therefore, before filtering, the Aggregate polygons tool with a threshold of 40 m was applied to merge polygons closer than this threshold. Additionally, all newly formed internal NF gaps up to 1 ha in size, which arose as a result of merging nearby UA polygons, had to be infilled.

For the final filtering, two threshold values were chosen based on the analysis of size and spatial relationships of all UA features defined so far: 1 ha (approximately the area of two football pitches) and 5 ha. All isolated UA features with an area up to 1 ha were classified as insignificant and transferred to the NF layer. All UA features larger than 5 ha are considered significant, even without the presence of a *Defining point of part of municipality*, thus preserving large logistics parks and warehouse complexes. In the category between 1 and 5 ha, at the current stage of the project (methodology validation is ongoing), UA polygons are regarded as significant if they meet at least one of the following conditions:

They contain a *Defining point of part of municipality*.
— A dominant building (church, chateau, etc.) is located within their area.
— They intersect with the sewer layer from the Digital Technical Map.
— A discharge record point (WWTP) is located nearby (within 100 m).
— They are crossed by a first-class road or a motorway.

CONCLUSIONS

This article presented a methodology for delineating the boundaries of significant urban areas (built-up areas) primarily for hydrological analysis purposes, using Czech open data from ZABAGED. The resulting datasets of natural features and urban areas will be publicly available at rain.fsv.cvut.cz in two versions. The first is a clean version in the form of the final product of significant urban areas as presented above; the second is a broader version that also includes all originally defined urban area patches between 1 and 5 ha in size. These layers are intended primarily for hydrological analyses, especially in relation to assessing risks from surface runoff and pluvial (flash) flooding.

The criteria for filtering small urban areas between 1 and 5 ha may evolve slightly in the future, primarily in connection with the gradual completion of the Digital Technical Map, which is not yet complete for the whole country but is undergoing intensive development.

The vulnerability of individual urban areas to pluvial flooding and soil erosion, as well as the potential impacts of drainage from these areas on the condition of watercourses, are related to the use of surrounding natural areas, their morphology, and technical features (ditches, road networks, etc.) that influence the path of concentrated surface runoff. Equally important are the characteristics of the urban areas themselves, such as the proportion and connectivity of permeable and impermeable surfaces, and the condition of settlement infrastructure. Correct spatial delineation of significant urban areas is a crucial initial prerequisite for assessing these aspects, which are the subject of ongoing research activities.

Acknowledgements

References

[1] BAUER, M., DOSTÁL, T., KRÁSA, J., JÁCHYMOVÁ, B., DAVID, V., DEVÁTÝ, J., STROUHAL, L., ROSENDORF, P. Risk to Residents, Infrastructure, and Water Bodies from Flash Floods and Sediment Transport. *Environmental Monitoring and Assessment*. 2019, 191(2), 85. Available at: https://doi.org/10.1007/s10661-019-7216-7

[2] ROSENDORF, P., DOSTÁL, T., KRÁSA, J., HANEL, M. et al. *Metodika stanovení území potenciálně ohrožených dopady přívalových srážek spojených s erozí půdy při zohlednění očekávané změny klimatu. Certifikovaná metodika*. Prague: T. G. Masaryk Water Research Institute, p. r. i., 2016. 58 pp. Certified by the Ministry of Agriculture of the Czech Republic under ref. no.30205/2016-MZE-10052. ISBN 978-80-87402-35-6.

[3] State Administration of Land Surveying and Cadastre (ČÚZK). *Katalog objektů ZABAGED®*. [on-line]. Prague: ČÚZK, continuously updated [accessed 2025-05-12]. Available at: https://geoportal.cuzk.cz/Dokumenty/ZABAGED_katalog/CS/

[4] KOŽÍN, R., VIZINA, A., KAVKA, P., STRÁNSKÝ, D., STROUHAL, L., HANEL, M., MÁCA, P., STRNAD, F., ŠTĚPÁNEK, P., BERNSTEINOVÁ, J., FISCHER, M., METELKA, T. *Odborná zpráva o řešení projektu SS06010386 – Adaptace urbanizovaných území na přívalové povodně a sucho. Průběžná zpráva za rok 2024*. Prague, T. G. Masaryk Water Reseach Institute, p. r. i., 2024. 79 s. Non-public internal document of project no. SS06010386.

[5] KUBEŠ, M. et al. *Metodika identifikace a klasifikace území s urbanistickými hodnotami* [on-line]. 1st ed. Prague: National Heritage Institute, 2015 [accessed 2025-05-29]. Available at: https://www.npu.cz/cs/e-shop/18517-metodika-identifikace-a-klasifikace-uzemi-s-urbanistickymi-hodnotami

[6] Ministerstvo zemědělství ČR. *Veřejný registr půdních bloků LPIS* [on-line]. Prague: Ministry of Agriculture, [accessed 2024-09-06]. Available at: https://eagri.cz/public/app/lpisext/lpis/verejny2/

[7] Český úřad zeměměřický a katastrální (ČÚZK). *RÚIAN – výstupní data ve výměnném formátu (VFR) [geografická data]*. Prague: ČÚZK, 2024. Available at: https://vdp.cuzk.cz/vdp/ruian/vymennyformat

Authors

Ing. Luděk Strouhal, Ph.D.[1,2]
✉ ludek.strouhal@fsv.cvut.cz
ORCID: 0000-0002-3979-4894

Petr Kavka[1]
✉ petr.kavka@fsv.cvut.cz
ORCID: 0000-0002-6511-9518

[1]Czech Technical University in Prague, Faculty of Civil Engineering, Prague (Czech Republic)
[2]T. G. Masaryk Water Research Institute, Prague (Czech Republic)

The Czech version of this article was peer-reviewed, the English version was translated from the Czech original by Environmental Translation Ltd.