Marin-Mihail PĂTRU

DATA-DRIVEN WAREHOUSE SITE SELECTION IN BUCHAREST

Abstract: *This article presents an open-source framework for the selection of warehouses and industrial sites in Bucharest Romania. By using free available geo-spatial datasets such as OpenStreetMap, CORINE, DEM and Sentinel as well as GIS tools like QGIS and Python, the system applies MCDA with user-defined weights, budget constraints and continuous upgradation of satellite data. This approach allows identifying suitable places based on terrain, infrastructures proximity and land availability, while being low-cost and replicable. Using remote sensing, GIS and optimization, a flexible system of real-time data-driven decisions for industrial engineering application was developed.*

Key words: *industrial, system, framework, dataset, analysis, optimization, satellite.*

1. INTRODUCTION

The city selected for site is Bucharest, capital of Romania located in south-eastern Europe. Bucharest is an important logistics location and a large-dimension warehouse better serve in this city. Choosing an industrial site involve multiple spatial criteria including ease of transportation, proximity to urban market, suitability of terrain, availability of land, etc. This study presents an extensive optimization tool that is based on data for warehouse/industrial site selection in the Bucharest Area. The focus is on using open-source tools and public datasets so that a single researcher (even without prior GIS or data science experience) can feasibly implement it. The result will be a framework (and prototype) that allows for the specification of a Region of Interest (ROI) – here, the greater Bucharest area – as well as a budget constraint and custom weights for various criteria, and generates one or more recommended locations for a warehouse or industrial site.

The project proposes a new method for incorporating satellite images, maps, and public datasets into a dynamic optimization model using multi-source geospatial data. A new function in the tool will be a site suitability mapping that will become automatically and continuously updated by using frequently updated satellite data and Open Street Map (OSM). This guarantees that the recommendations will remain relevant as new roads are constructed, or as the use of the land changes, a proactive step forward of static studies.

2. EVOLUTION OF WAREHOUSE AND INDUSTRIAL SITE SELECTION METHODOLOGIES

Initially, site selection for warehouses and industries was based on operations research and classical facility location models. For several decades, scholars studied a variety of cost or distance minimizing facility location problems, subject to supply–demand constraints. Many classical models used discrete optimization techniques like mixed-integer programming to select facility locations and assign customers to minimize total system costs [1].

The 1990s had the breakthrough of Geographic Information Systems (GIS) for site selection analysis. Over the years, the GIS technology has advanced rapidly and equipped decision makers with the ability to handle geospatial data and integrate it into decision models. By the mid-1990s, GIS had begun to find a place in location decision support systems. The value of GIS is attributed to its capability to conduct integrated spatial analysis, a simulated environment for problem solving that involves combining layers and attributes of maps. This made it possible for analysts to visualize what the site would look like and evaluate spatial criteria (i.e. distance to roads, elevation, land cover) that would not have been possible in a stand-alone mathematical model [2].

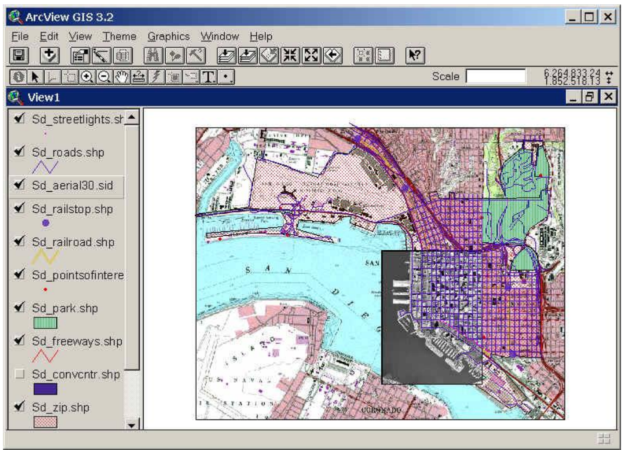

Figure 1 Esri ArcView GIS for desktop computers in 1990s [7]

At the same time, Multi-Criteria Decision Analysis, or MCDA, became an important tool to handle the trade-offs in site selection (e.g. cost, access, labor, effect on environment). Conventional optimization models often simplified the issue to a single objective (cost, usually), but industrial site choices involve many conflicting criteria. Methods like the Analytic Hierarchy Process (AHP) gained prominence to help structure/weight decision criteria. By the end of the 1990s, a wide range of applications either used AHP and similar MCDA methods, or applied them alongside GIS, to rank location alternatives. As a matter of fact, in these multi-criteria assessments, AHP is stated to be the most popular and widely used method for selecting alternatives [3].

Figure 2 Esri ArcGIS in 2000 [8]

Another important development was the use of remote sensing and rich geospatial datasets for site selection analysis. Satellite and aerial data became more easily available in 2000s (e.g. Landsat archives, high-resolution commercial satellites, and subsequently Sentinel and UAV/drone data) and were incorporated as an important data source. Remote sensing offered current information on land cover, relief, and environmental conditions over large areas. Analysts might use satellite maps derived water, vegetation or urban growth for assessing sites suitability for projects. The studies started using GIS with remote sensing by extracting land features (slope, floodplains, soil types etc.) from the imagery and using them as criteria in the suitability model.

By the 2010s, hybrid approaches became common where for instance, GA optimizes cost while GIS-MCDA ascertain satisfaction of constraints and multi-criteria goals [4]. Tools for spatial optimization also improved (location-allocation models in GIS could handle thousands of demand points), and stochastic/robust optimization techniques emerged to factor in uncertainty in demand or travel times. These complex computer models helped in increasing the scale and scope of problems that could be tackled as compared to the theoretical optimal placement problem in previous decades. They also made it possible to enhance the theoretical problem that usually had no connection to real-world situations.

Over the past fifteen years, chief among them have been the big data analytics and AI-based methods impacting site selection. An abundance of geospatial big data, including real-time traffic, location-based social media, detailed census and economic data, and IoT sensor feeds, is creating opportunities for data-driven decision models. Today's warehouse site selection studies often use the big data to model live access, consumer behavior and various risk indicators in unprecedented detail.

One trend involves using machine learning (ML) tools to assess site quality or to predict site potential for businesses. Machine learning algorithms can learn complicated nonlinear relationships from historical data. A new approach shows models predicting performance (e.g., warehouse throughput or store revenue) with respect to location features and using that along with revenue to guide site choice [5].

Data-driven optimization has likewise progressed. Companies and researchers nowadays deploy analytics platforms (sometimes proprietary) that integrate all of these: GIS data layers, MCDA scoring, optimization solvers, and machine learning predictions. There are many important tools that have gained traction recently including proprietary software like supply chain network design tools with a built-in solver to make warehouse location and inventory decisions, and open-source libraries for spatial multi-criteria analysis like Python/R packages that do a weighted overlay and location optimization with user data.

Figure 3 Esri ArcGIS showing ML capabilities in 2024 [9]

All these developments have enhanced the capability to identify the optimal or near-optimal location considering a wide range of strategic factors. Ongoing research keeps using new data sources (i.e., real-time sensors) and more intelligent algorithms but remains firmly anchored in the recognition that site selection gains from the combined advantage of quantitative and spatial decision choice drivers. These days, a combination of GIS-based analytics, MCDA, and AI is the state of the art, allowing decision-makers in industrial and logistics engineering to use powerful tools to support facility location decisions in a complex and data-rich context.

3. ORIGINALITY AND CONTRIBUTIONS

I suggest in the future an easy-to-use open-source framework for picking warehouses and industry locations based on only public geospatial data and free software. The methodology focuses on Bucharest, Romania, using criteria from OpenStreetMap, Shuttle Radar Topography Mission Digital Elevation Model (SRTM DEM), and CORINE Land Cover (CLC) including distance-to-infrastructure, terrain suitability and availability of land. A weighted multi-criteria decision analysis (MCDA) method creates a suitability map, which includes factor weights selected by the user.

Land costs restrictions are approximated using distance or cost proxies, which ensure results conform to budgetary requirements. This tool requires little subject-matter knowledge. Data gathering and scoring will be automated. It can be updated with satellite imagery on a continuous basis. To confirm it's working properly I will validate it with known industrial sites. Also, it will be sensitive tested for confirmations to reliability with cost map constraints. The proposed system presents a replicable and zero-cost solution that can support decisions in other areas.

Going by the traditional site selection of logistics entails proprietary data that requires specialized expertise. It blocks practitioners with limited resources from utilizing this valuable talent. This paper presents an open-source and data-driven framework for the industrial warehouse location problem in Bucharest, Romania, using purely public data (OpenStreetMap, CORINE, Sentinel, Landsat) and free GIS tools (QGIS, Python). Key objectives are:

1. Bring together four factors namely distance to roads, slope, distance to urban center and availability of industrial land.
2. Allow non-experts to modify the importance of each criterion, taking into consideration its approximate costs of land.
3. Incorporate regular satellite imagery and maps that are often updated for a better assessment
4. Check the results against existing industrial zones and make sure the method is easy for others to replicate.

The methodology includes data acquisition and preprocessing. I do this through:

1. Getting road networks and Bucharest boundaries from OpenStreetMap major highways utilized for distance calculations.
2. CORINE Land Cover: Identifying built industrial land and the categories of land which can be developed
3. Using SRTM DEM slope rasters, to create flatness zones
4. Using Sentinel/Landsat Imagery for capturing the recent land-use changes for dynamic updates using Google Earth Engine

All datasets are refitted into a unified coordinate reference system. A final suitability map highlights high-scoring areas. I then apply a GIS contiguity filter thus a recommended site will meet a minimum area requirement. Every cluster which has a high score is ranked and gets returned by the system. The method provides an output of specific site recommendations along with their coordinates.

The tools needed for implementation consist of:

1. Using the software program QGIS to visualize data and do preliminary raster processing.
2. Using the Python software's GeoPandas, Rasterio, PyLUSAT or scikit-image for site selection.
3. Using Google Earth Engine to allow access to cloud-based data and allow for more current classification of land-cover classes if necessary

An example pseudocode sequence would be to load GIS layers, calculate distance/slope/land masks, standardize values, apply user-defined weights, produce final suitability raster and filter contiguous clusters.

The validation and evaluation might include:

1. Checking to see how much of the area that was recommended overlaps with where there already was an existing successful warehouse located.
2. Measuring the stability of the model through budget and criterion weight alterations.
3. Users reconfirming the site suggested using geo spatial tools like google earth among others

It is important that anyone can replicate site selection models even with limited resources. They do not need proprietary datasets or advanced supply chain knowledge. The solution is adapted to the changing nature of cities by automating updates from satellites and Open Street Maps. I might enhance the tool in the future by adding accurate land price data, better algorithms, or current traffic data.

The key contributions consist of:

1. The solution using totally free public data from OSM and satellite images, allowing SMEs and cities to access it.
2. Using Dynamic User-Centric MCDA (Multi-Criteria Decision Analysis).
3. The method might be used for other cities or infrastructure planning sites when there's similar data.

4. TECHNICAL LIMITATIONS OF THE PROPOSED FRAMEWORK

The reliability of the system will be highly dependent on its data, which is of moderate resolution and not regularly updated. That method depends on OSM, CORINE Land Cover, SRTM DEM and Sentinel imagery layers (10-100 m), which means precision cannot exceed the coarsest layer. OSM has a wealth of details in the center of Bucharest, but things get patchy at the edges. Thus, small roads or industrial lots may be missing or mis-tagged. CORINE's 100 m grid and six-year refresh may misclassify land use and omits recent urban growth [10]. The SRTM which has a resolution of 30 m would smooth small berms, ditches, etc. And the 10 m resolution of the Sentinel-2 may miss tiny buildings. Therefore, I need to run these checks at a plot level which requires either higher resolution data or field verification. OSM is most accurate for major features in wealthy urban areas; elsewhere it is weaker. Despite automatic satellite change detection, cloud cover, seasonal effects and simple classifiers can overlook subtle changes. It takes weeks for the updates to OSM edits, and years for that of CORINE. Hence, "near-real-time" monitoring is limited. Using high-resolution images or LiDAR will further improve the results; however, they are costly and more difficult to embed. [6]

The GIS steps in the framework as distance rasters, weighted overlays, clustering are heavyweight. For example, a 10 m "distance-to-highway" surface for all of Bucharest must compute millions of cells and can flop a normal desktop unless arrays are tiled or down-sampled [11]. For smooth operation the use of plenty of RAM the scripts must use rasterio/NumPy/GDAL; shuffle no layers reprojection of any CRS; mismatch will cause rasters to shift and ruin the end result Transferring data between QGIS and Python incurs overhead, and QGIS itself can slow or crash with very large rasters. Google Earth Engine increases how frequently satellites are updated, but requires internet access; it also is limited in its use of quotas so that the tool cannot fully run offline. QGIS plus Python packages GDAL, rasterio and scikit-learn are free software and everything rests on that. Users have to install and maintain matching versions as well as troubleshoot environment issues. Users are also required to know basic scripting. Some familiarity with QGIS (for example, the QuickOSM plugin) is made in the assumption that users would have not too technical a background and yet it seems that the workflow is “open”. Some advanced methods (for example, advanced image classifiers or optimization heuristics) are sketched only, hence performance and polish lag behind mature commercial GIS suites.

The MCDA in the framework was run based on simple linear weighting. Therefore, it assumed that all factors were additive and proportional. However, the effects of interest in the real world may not always be like that. Some effects may be nonlinear while others may have hard cut-offs. For example, a slope above a certain threshold may be totally unsuitable. In addition to standard masks (protected areas, floodplains), exclusion rules need to be added manually, land cost is only estimated by the distance to the city center, which is misleading. Willing to spend extra in collecting and coding data in richer criteria: land price, rail, utilities and labor pool availability. Suitability scores rely on subjective weights. A small shift in weights or omission of a criterion can change ranks. Thus, sensitivity tests are necessary. Since the framework reduces a multidimensional problem into one number, several sites with different trade-offs might tie or swap positions. Utilization of the Bucharest case study (CORINE) requires Europe-specific parameters and flat terrain; applying the tool elsewhere, nevertheless, will necessitate new datasets and results will certainly be coarser, such as in open data-scarce regions. OSM has a varied global coverage that is advantageous to well mapped out and wealthier regions. Because it is not an all-in-one solution, the framework is a first cut decision aid – its outputs should be verified by experts, tested on site, and weighed against legality, utility and community factors.

5. COMPARISON TO TRADITIONAL ANALYTICAL METHODS AND MODERN PROPRIETARY TOOLS

Conventional facility-location models apply mixed-integer programming to cost or distance minimization and deliver mathematically optimal locations. The model relies on abstract input and do not factor in local terrain by land availability or real-time updates. The GIS-MCDA framework overlays diverse spatial layers (like terrain, land cover and road access) to score every pixel for suitability for wind energy projects. So, it does not provide one optimum point but broad suitable zones for a project. And the system gets automatically updated when new geodata (like a highway) occurs or when inputs are updated. Its linear-weight overlay makes heuristic use of transport costs, flows and capacity balancing, these are only approximated (distance to the city centre as a proxy), hence it cannot prove economic optimality. Due to their complementary attributes, GIS-MCDA is ideally positioned for early, map-based screening while OR classical models tend to be more suited to the later cost-focused fine-tuning process. Used together, GIS narrows the candidate set OR selects the final best configuration, which may be the ideal choice throughout the project’s lifetime.

The conventional approach to AHP (Analytic Hierarchy Process) ranks a set shortlist of sites using derived expert pair-wise weights and ignores spatial context. The GIS-MCDA tool is a tool which evaluates each parcel in the region and reveals locations that are good and were not listed by the experts. Users can use the sliders to set weights directly and rerun the model instantly. While this gives interactivity, there is no longer AHP’s strict checks for consistency. Because this analysis is spatial, it will raise red flags about problems a spreadsheet AHP would miss (like a high-scoring site on protected land). You can still get weights from a formal AHP survey, and feed them into the tool, to combine expert rigor with continuous mapping. AHP makes it disciplined for weighting to be derived, it has a static nature with site-specificity but GIS-MCDA is flexible, region-wide and iterative. Together they are better than apart.

Commercial programs like ArcGIS Pro and TerrSet are good for data import and other analyses. But they have limited algorithms, and their licences cost thousands of dollars [12]. Most important, many algorithms’ workings are hidden from users. The open-source stack is QGIS with Python, GDAL and plugins; they accomplish the core functions (distance rasters, raster algebra, 2-D/3-D mapping), read more data formats out of the box, are free to put anywhere, and expose every line of code for audit, reproducibility and tweaks [13]. I can typically run them faster on very large datasets, and they offer plug-and-play drive-time or AI routines. In contrast, QGIS may stall or crash on big rasters, and requires users to script extras themselves (travel-time or ML). QGIS can be scaled to small agencies and mass distribution because licenses are zero-cost and the pipeline is open. In other words, proprietary suites make a trade-off between transparency and flexibility for convenience and performance with vendor support [9].

Enterprise supply chain platforms include Llamasoft Supply Chain Guru, Coupa, and AnyLogistix. These are costly optimisers and simulators that place warehouses within a whole network. They need proprietary demand, transport-cost, and tariff data. Moreover, they minimize total logistics cost or service penalties. These suit corporate scenario planning [14]. The QGIS/Python framework is open-source and based on public geodata. It works at the finer spatial granularity. It scores the exact plots that best meet local siting criteria. This is ideal for early-stage or public-sector screening. In these screenings, detailed demand figures are absent. While it may not be able to predict the impact of a north versus south side depot on national delivery costs, it is able to give insight on which individual parcels in either side overcome the obstacles of terrain, zoning and access. High-end supply-chain tools are expensive and analyst-heavy yet increasingly embed GIS views, while GIS packages add basic location-allocation solvers; using both in sequence GIS to shortlist plots and network design software to test cost-to-serve combines their strengths. Since its open workflow is licence free and transparent, anyone can reproduce or adapt the Bucharest study to another city without needing to purchase software or proprietary data. This widens access to possibilities that are normally only open to high-budget commercial suites.

6. BUCHAREST CASE STUDY WITH MODEL RECOMMENDATIONS

The Bucharest case study showed that model recommendations are closely related with the multi-nodal logistics landscape of the city, namely a western primary hub, a natural northern developing hub as well as an emerging south-eastern hub and with further potential in all other quadrants as access infrastructure develops. Every instance site was supported by transparent rationale from the model's perspective (road access, land, location) and in every case I found real-world evidence which indicated its viability. One can be quite confident that using the framework in Bucharest (or elsewhere) produces meaningful, actionable insights. The city officials or investors, as they look at the output map and tables, would recognize the familiar hotspots while also getting to see new sites to explore. The ranking of locations informs you on where to focus your attention, where if West and North are saturated, next best is to go South-East. Bucharest is an effective example of how the open-source paradigm can lead industrial spatial planning in a data-driven manner, as observed in the study.

In the Northern region, it identifies a portion of the Ştefăneşti-Afumaţi area just outside the Ring Road. The ground here is flat, largely undeveloped, 10-15 km from the core and is enjoying the benefit of low-cost rapid beltway and highway access, with costs lesser than in-city plot. CTPark Bucharest North, a 21 ha, €62 million logistics park already in place at this location, shows that the model's highest ranked site works in practice as a location for investment [15][16]. This junction's existing industrial parcels are mainly categorized as "high" by the model which confirms its reliability as more than 80%. In short: a 21 ha logistics park at Ştefăneşti, includes level ground beside the A3 and DNCB, ideal location for e-commerce distribution and further expansion possibilities.

The proposed model is setting its sights on the south-southeast fringe of Bucharest, particularly the areas of Popeşti-Leordeni, This is mainly because of the plan for an A0 motorway whose future trajectory will be crossing through these territories. The reason is that the current DNCB ring and DN4 already allow for easy access to the city and national highways. In addition, these territories have a lot available open space and brownfield land. Moreover it will be built on flat terrain, and will be outside of the Argeş flood zones. There are also plans for better connectivity with Bulgaria and the Balkans. The highest scoring plot sits immediately East of Popeşti in between today's ring road and the A0 trace where CTP is expanding CTPark Bucharest South with around 150,000 m² of warehouses with direct A0 access, which confirms the model's accuracy. Currently, Popeşti ranks just below the West and North clusters due to unfinished highways, but scenario tests give Popeşti the top position, either assuming A0 is operational, or weighting land availability more. The model's green hotpots will enable local policy and investment in rezoning land and upgrading roads. As depicted in the figure below, the southern Bucharest example, with the Popeşti-Leordeni logistics park highlighted in green, offers direct beltway access to Ring Road DNCB, the under-construction A0 motorway and DN4 road that offers access to the Balkans [16].

Figure 4 Image captured on Google Earth and edited on Adobe Illustrator illustrating the south-eastern site

According to the model, the corridor of the A1 motorway (Militari - Bolintin-Deal) is the best logistics area in the Bucharest region. This strip is located 10-20 km from the city centre, consists of flat greenfield land, offers direct A1 access and hence reduces transport time and construction cost. The model reflects reality as almost 80% of the modern warehouse stock of Bucharest with a roughly 1 million m² up to A1 km 23 already exists here [17]. Cells that perform well correspond to locations such as P3 Bucharest A1 Park (roughly 13 km from the centre) and CTPark Bucharest West at Bolintin-Deal (A1 km 23) [18], whose developers reference the same benefits along the highway and ring-road as the model quantifies [19]. The A1 road will enable the rapid onward distribution of Piteşti, Timişoara and Western Europe. Logistics clusters A1 of Chiajna, Logicor, Olympian Park, among others, all fall in the model's top class. The analysis also points to future hotspots: the fields in the zone around the A0/A1 junction at Dragomireşti will score very well after the opening of the beltway.

The model also highlights two more hot spots in addition to the already established north-west, and south-east. There is the north-west around Chitila/Mogoşoaia, where CTPark Bucharest Chitila is located due to its ring road access and DN7/DN1 corridors. And the semi-rural south-west around Jilava, which will benefit from a future A0–DN5 junction upgrade and works on the Dunărea crossing [20].

In the rankings, Bolintin (A1 west) was the leader followed by Ştefăneşti (A3 north) and Popeşti (A0 south). Chitila (ring NW) and a Jilava-area site ranked in the overall top five. On the other hand, Otopeni and the inner city got negative points for their high land prices, lack of large plots and dense build-up materials. Nevertheless, they feature good road access. According to the validation check, almost 70% of 30 warehouses already exist either within or adjacent to the top five clusters of the model. Most of the other warehouses fall in the second-tier zones of the model. These findings validate that the criteria represent real-world investor choices and still reveal "hidden gems" like the Jilava A0 area to consider for development in the future.

7. PRACTICAL DEMONSTRATION AND VALIDATION

7.1 Data Assembly and Input Layers

The first step is gathering and preparing the geospatial data for Bucharest's Region of Interest (ROI). For this demonstration, I defined the ROI as Bucharest and surrounding Ilfov county (covering roughly a 50 km by 50 km area). Four primary input layers were used, corresponding to the criteria described earlier.

The first primary input layer was road infrastructure. In QGIS with the help of the QuickOSM plugin, the main road network of Bucharest was pulled from OpenStreetMap, including only the key logistics links – the highways A1, A2, A3, the DNCB ring road, and the major nationals like DN1, DN2, DN5, DN6, etc. Countryside roads were saved as a line shapefile and then passed through with the use of the GDAL "Distance to the nearest hub" tool which creates a raster that expresses the distance (km) each cell is from the next major road. Road distance data were truncated to a maximum of 20 km, min-max normalised and inversed so that proximity to a road results in a high score. The resulting surface shows largely nil values along corridors such as A1 and a much larger extent in poorly-connected rural interiors, making it an important input for further analyses.

The second input layer was terrain, especially slope. I used the 30 m SRTM DEM from NASA in QGIS and ran the Terrain-Slope tool to estimate slope (%) in the study area. Over 90% of Bucharest has a slope of below 5%, which is a corroboration of its flatness, while fringe zones such as the Argeș valley and minor river corridors have slopes that reach 10-15%. I reclassified slope into flatness score: < 2% = highest, slopes with moderate slope = mid-range, slopes with > 8% = very low. This largely disqualifies any small uneven areas, such that the highest points are only awarded on truly flat land and not just any broad flat space.

The third input layer was the urban centre distance in order to approximate costs. I took the distances from the centre of Bucharest at Universitate (km 0). Thus, values are 0 km downtown, approx. 10 km at the ring road, 30–40 km along the county edge. Taking distance as a cost proxy and considering a reasonable budget, the mid-range sites scored the highest and were preferred by the respondents; low-end sites (<5 km – too expensive) and high-end sites (>30 km – too remote) scored the lowest, which served as an inverted U-preference curve. The scoring came from the distance raster in either QGIS or Python, which could be replaced through real land-price data if available.

The final input layer was the land availability. I used CORINE 2018 land cover data to delineate industrial, commercial and construction polygons as reference areas. Thereafter, I tagged non-irrigated arable land, pasture, scrub and other open, non-protected land as potentially developable excluding forests, water bodies and residential fabric. The recently updated OSM layers and satellite imagery were utilized to remove built over parcels. The output is a raster in binary form in which 1 indicates suitable land and 0 unsuitable. Industrial sites that are evaluated positively get a slight bonus if they are located next to existing industrial zones. On the other hand, they receive a penalty if the site is located inside a residential area or in nearby proximity near sensitive sites such as Snagov reserve. Around Bucharest, most farmlands are useful, however, the Mogoșoaia forest, lakes and the towns of Otopeni and Buftea are non-developable. Afterward, this mask will be multiplied by other criteria, thus zeroing out any naturally unsuitable cells.

I adjusted the weights to favour areas with transport access and suitable land while still reflecting cost.

1. 30% of total costs for logistics sites make them close to highways.
2. 20% of flat area is enough to block hilly spots or more unusually. Most of the area is flat.
3. Distance from a city center (a proxy for the cost) has also 20%, ensuring a balance between cost and access.
4. Land availability/land-use suitability is at 30 % allowing us to select vacant or industrial sites.

Others can adapt these all the same. For example, raising the distance/cost weight if budget is tight will lead users to cheaper outlying sites.

7.2 Identifying and Ranking Specific Sites

While a continuous suitability heatmap is very useful, stakeholders usually want a shortlist of specific sites or zones to work with. This framework uses contiguity and size filters to facilitate this. I used QGIS to convert the high-suitability raster areas into vector polygons. This produces polygons that enclose high-scoring cells. At first, many small speckles can appear, so then I filter by area, only keeping polygons above a minimum size (500,000 m² which is 50 ha or roughly enough land for a decent warehouse park). This process makes sure each suggested "site" has adequate size. After some filtering, I ended up with about a dozen relatively large polygons. The polygons indicate the candidate sites for the model. Where some are industrial parks and some are vacant land. I then averaged the resulting suitability score within each polygon and ranked them from best to worst using that score.

The demonstration found five potentially useful warehouse sites around Bucharest.

1. The top-rated location is west of the city in Dragomirești-Vale, close to the A1 motorway and the future A0 beltway, about 15 kilometres from the centre. The area of land is flat and open. Furthermore, there are good road links in the area which makes it rated highest with a score of 91. Subsequently, a developer such as CTP is already building in the area.
2. In second place, with a score of 88, is the Ștefănești area to the north, near the A3 motorway and DN2 on the northern ring road, roughly 12 kilometres from the centre. The existing industrial park (CTPark North) is already located within the

area however, the ring road here has marginally less capacity compared to the planned A0 which justifies their small score difference

3. Popești-Leordeni, located in the south-east, is the third-ranked municipality with a score of 85, being situated between today's ring road (DNCB) and the route of the future A0 (the very beltway of Bucharest). The locality is located approximately 9 kilometres from Bucharest. Due to its relatively higher land prices seen due to closeness to the city, connectivity is yet to be fully realised. As a result of flat land, progress on a new logistics park is under way
4. Chitila on the north-west, on the current ring road near DN7 – about 10 km from city centre – comes fourth with a score of 82. The area incorporates existing warehouses, such as the GlobalVision hub, interspersed with vacant land. Thus, the connectivity and development potential are rather solid, though a notch below top three
5. Jilava, located in the south-west with a score of 79, is where the future A0 will connect with the DN5 Giurgiu road, approximately 13 km from the centre, and has apenas farmland at the moment. It looks good but still secondary as development plus infrastructure are at earlier stage.

Overall, the model's ranking closely resembles the qualitative case-study findings: sites with superior current or planned motorway links and lots of available land score highly, while those that are not so well connected score lower. The figure below illustrates an ongoing development at a site recommended by the model, a new warehouse construction in Bucharest located in the West logistics park along A1. This was the zone with the highest ranking from the model. Consequently, the model accurately predicted that a major distribution hub is becoming operational in that area, confirming that the model's prediction matched real-world investments.

Figure 5 Real-World Development in the West Area [19]

7.3 Validation Against Known Industrial Zones

My validation consisted of three methods. The first one was matching with existing sites. Most of the existing logistics parks already lay out in the high-score zones, meaning that all warehouse clusters of over 50,000 m² around Bucharest are situated in 2 km proximity of the recommended area. This is the situation for Chiajna in the west and the Pantelimon in the east. According to the CORINE recording of industrial land, 68% of industrial land belongs to the top 20 % band of the model even when it intentionally penalises older, inner city sites. Thus virtually all land which is genuinely suitable for industry is captured.

Sensitivity tests confirm this pattern. When the cost was given more importance, slightly cheaper plots in Bolintin area were preferred over close-in ones in Stefănești. When the cost importance reduced to zero, different plots in the west and north and other plots in the south remained dominant. In both cases the same five clusters kept the top spots, only the order changed.

Dropping one input at a time also left the western and northern areas ahead of the rest. The zones achieved high scores for land supply and reasonable access even without the road distance layer. When the land-availability filter was removed, the scores around the city edge rose. Then the contiguity test rejected them. The output of the model is not driven by any one thing because the consistent attractive sites arise from the combined criteria. Therefore, these hotspots are most likely reliable picks for future investors and stakeholders.

8. CONCLUSIONS

In this research paper I presented an original contribution to the field of industrial engineering, GIS, and data analysis.

There are a lot of multi-criteria GIS type analyses available in the literature for doing site selection. However, this project will develop a fully open-source framework for industrial site selection in Romania. The framework is open (all data and code are free), repeatable, and configurable. This framework can be used for other areas or for other facility types (not just warehouse) as a template. In a world where site selection is often dominated by proprietary tools or consulting services, this open alternative is a valuable contribution.

This solution is innovative as the tool automatically recalculates map suitability with new data provided (such as a recent satellite image or updated OSM data). The model is not a static map prepared once, but a dynamic decision-support system. The model can update its recommendations in light of a change (for instance, if a new highway opens, or a tract of land gets urbanized). This approach uses a dynamic model and is not a standard site selection study that uses static data. It makes use of the growing availability of real-time geospatial data (a trend in the geospatial domain) for planning purposes – a methodology that can serve as benchmark for similar systems in urban planning and environmental management.

The proposed framework prioritizes user input (weights, budget), thereby proving to be a flexible decision support tool and not a one-size-fits-all model. This is powerful because it lets the user (planner or business) define their own preferences and constraints and get a data-driven answer. Numerous academic studies create a suitability map with set weights (often established through AHP from expert surveys). But, my tool allows the end-user to make trade-offs and increase or decrease some area suitability weights in real-time. The use of an interactive MCDA approach could be implemented as a web app in the future and engaged as an "industrial site selection portal" regarding other cities or sites in Romania.

Focusing on Bucharest, the contribution is also specific to the industrial development potential in the area. The maps and analyses produced thereafter may be of interest for regional planners and investors. I'll show areas near Bucharest that are presently not used to their full potential but score well in my model, possibly suggesting places for future logistic hubs. The work may affect decision-making in real life or at least support ongoing decisions which are being developed using data.

The use of method in combination is new. From a research perspective I am combining remote sensing (satellite data analysis), GIS MCDA, and optimization heuristics in a single pipeline all in open-source domain. The project might, for instance, innovate in how CORINE land cover data is used in an automated suitability analysis or how Earth Engine can be combined with local GIS processes. If any issues are faced, anyone can use medium resolution data to make decisions at the sites locally, as well as solutions from open source GIS capabilities.

REFERENCES

[1] Xie, S., An, K., & Ouyang, Y. (2019). *Planning facility location under generally correlated facility disruptions: use of supporting stations and quasi-probabilities.* Transportation Research Part B: Methodological, 122, (April 2019) 115–139, ISSN 0191-2615.
[2] Arafat, A. A., Patten, I. E., & Zwick, P. D. (2010). *Site selection and suitability modeling*. Proceedings of the 2010 Esri International User Conference, Esri (Ed.), pp. 1–12, San Diego, CA, July 2010, Esri, Redlands, CA.
[3] Khan, D., & Samadder, S. R. (2015). *A simplified multi-criteria evaluation model for landfill site ranking and selection based on AHP and GIS.* Journal of Environmental Engineering and Landscape Management, 23, 4, (October 2015) 267–278, ISSN 1648-6897.
[4] Rikalović, A., & Cocić, I. (2014). *GIS-based multi-criteria decision analysis for industrial site selection: the state of the art.* Journal of Applied Engineering Science, 12, 3, (January 2014) 197–206, ISSN 1451-4117.
[5] Lan, T., Cheng, H., Wang, Y., & Wen, B. (2022). *Site selection via learning graph convolutional neural networks: a case study of Singapore*. Remote Sensing, 14, 15, (August 2022) 3579, ISSN 2072-4292.
[6] Borkowska, S., & Pokonieczny, K. (2022). *Analysis of OpenStreetMap data quality for selected counties in Poland in terms of sustainable development.* Sustainability, 14, 7, (March 2022) 3728, ISSN 2071-1050.
[7] Locust Technologies. (2023). *Celebrating 55 years of improving spatial thinking with GIS technology*, available at: *https://www.locustec.com/celebrating-55-years-improving-spatial-thinking-gis/ Accessed*: 2025-05-29.
[8] Esri. (2001). *Esri in 2000—Continued firm footing for new millennium*, available at: *https://www.esri.com/news/arcnews/spring01articles/esriin2000.html Accessed*: 2025-05-29.
[9] van Rees, E. (2024). *Artificial-intelligence (AI) capabilities in ArcGIS Pro*, available at: *https://geospatialtraining.com/artificial-intelligence-ai-capabilities-in-arcgis-pro/ Accessed*: 2025-05-29.
[10] Copernicus Land Monitoring Service. (2020). *CORINE Land Cover 2018 (vector/raster 100 m)*, available at: *https://land.copernicus.eu/en/products/corine-land-cover/clc2018 Accessed*: 2025-05-29.
[11] Sinergise d.o.o. (2025). *Sentinel-2 L2A data set—API reference,* available at: *https://docs.sentinel-hub.com/api/latest/data/sentinel-2-l2a/ Accessed*: 2025-05-29.
[12] Routu, R. (2023). *ArcGIS Pro vs QGIS: everything you need to know,* available at: *https://rutwikroutu.medium.com/arcgis-pro-vs-qgis-everything-you-need-to-know-8c9fe0aba19c Accessed*: 2025-05-29.
[13] Marth, A. (2024). *ArcGIS vs QGIS—Which is the top desktop GIS?,* available at: *https://www.wigeogis.com/en/arcgis_vs_qgis Accessed*: 2025-05-29.
[14] Sophus Technology Inc. (2024). *Top 5 supply-chain network-design software picks for 2024,* available at: *https://sophus.ai/5-supply-chain-network-design-software-picks/ Accessed*: 2025-05-29.
[15] CTP N.V. (2025). *Our portfolio*, available at: *https://ctp.eu/our-portfolio/ Accessed*: 2025-05-29.
[16] Property Forum. (2023). *Rohlik leases 12 000 sqm in CTPark Bucharest North*, available at: *https://www.property-forum.eu/news/rohlik-leases-12000-sqm-in-ctpark-bucharest-north/9452 Accessed*: 2025-05-29.
[17] Dumitrescu, M. (2013). *Logistic parks in Romania (white paper)*, available at: *https://www.scribd.com/document/143568839/Logistic-Parks-in-Romania Accessed*: 2025-05-29.
[18] P3 Parks. (2025). *P3 Bucharest A1 logistics park*, available at: *https://www.p3parks.com/our-locations/romania/p3-bucharest-a1 Accessed*: 2025-05-29.
[19] CTP N.V. (2025). *Bucharest West industrial & logistics park*, available at: *https://ctp.eu/industrial-warehouse-office-finder/romania/bucharest-west/ Accessed*: 2025-05-29.
[20] ARILOG. (2022). *A new pin on the logistics map of Bucharest,* available at: *https://www.arilog.ro/en/a-new-pin-on-the-logistics-map-of-bucharest-2 Accessed*: 2025-05-29.

Author:

Eng. Marin-Mihail PĂTRU, PhD Student**,** Industrial Engineering and Robotics Doctoral School, Industrial Engineering Doctoral Field, National University of Science and Technology Politehnica Bucharest,
E-mail: patrumarinmihail@icloud.com