

Integrated a Suitability Model to Locate the Potential Storage of Stormwater in New York City and Sacramento

Guangyi Zhang

*School of Environmental Science and Engineering, Southwest Jiaotong University, Chengdu, China
13335649101@my.swjtu.edu.cn*

Abstract: How to store stormwater has become a popular topic this decade since severe precipitations occur more frequently, especially in metropolises. Traditionally, the stormwater storage was located by manual planning and discussion based on the GIS data. In this paper, we developed an automatic method to identify the locations using geo-intelligent tools in ArcGIS Pro. We first selected two identical locations (New York City and Sacramento) in the US as examples since these two cities have been affected by the huge precipitations in the past 25 years. Then, we applied the suitability model from ArcGIS Pro to locate the potential areas to store the stormwater based on multiple geospatial datasets. Finally, we picked five locations for each city and validated our results with the published reports. Based on cross-validation, our results have similar patterns to the published reports; however, our method is more reliable and reproducible than the traditional urban planning methods. Moreover, a geo-intelligent method is easy to adjust based on parameter tuning and can adapt to changes in the situation.

Keywords: suitability model, stormwater storage, green infrastructure, geo-intelligent

1. Introduction

Global warming has caused severe issues that have impacted human activities and settlements [1]. With rapid industrialization and urbanization, a city became the center of business activities and the home of most people. Recently, extreme precipitation has become more frequent in urban areas associated with global warming [2]. It also has caused construction damage and life loss [3]. A green infrastructure strategy was created to minimize or control the destruction caused by stormwater [4]. A GIS-based method was deployed to manage the stormwater in Melbourne, Australia [5]. A stage-based model was built to detect the best location to store the stormwater in Detroit, Michigan, and Addis Ababa, Ethiopia [6]. In this paper, a data analysis was applied to the precipitation data, and a suitability model was created to find the best locations for stormwater storage in New York City and Sacramento.

2. Methodology

2.1. Study area

There are two locations in this study. The first study area is Sacramento (38°34'N, 121°29'W), the capital of the state of California, U.S. Sacramento has a Mediterranean climate, which is characterized

by wet and mild winters as well as dry and hot summers. The annual precipitation is 470mm. The temperatures in July and December are 24 and 7.9 Celsius. The population (2023) of Sacramento is 0.526 million (US Census), and the GDP (2023) of Sacramento is \$189 billion (Fred).

The second study area is New York City (40°43'N, 74°00'W), the biggest city in the U.S. The climate type of New York City is a humid continental climate, which is characterized by a wide variation of temperature and evenly distributed precipitation. The annual precipitation is 820-1100mm. The temperatures in July and January are 23 and 0.3 Celsius. The population (2023) of New York City is 8.258 million (US Census), and the GDP (2023) of New York is \$1.9 trillion (Fred).

2.2. Data

This study uses six datasets: monthly precipitation from 2000 to 2024, DEM (Digital Elevation Model), land cover, stream, soil, and urban data.

- Precipitation was used to determine which counties have an extremely large change rate of precipitation and a potentially large impact on the stormwater. The data was downloaded from the Department of Agricultural and Resource Economics, University of California, Davis by Prof. Aron Smith (<https://www.aaronsmithagecon.com/download-us-weather-data>)
- DEM is an embranchment of DTM (Digital Terrain Model), it was used to simulate ground terrain with digital data. DEM data was used to calculate the slope of the land surface. The data was downloaded from USGS (United States Geological Survey) Earth Explorer and obtained from SRTM (The Shuttle Radar Topography Mission). This mission used interferometric radar to compare two radar signals, which were taken by two separate radar antennas, an antenna on board the space shuttle and the other located at the end of a 60-meter mast that extended from the shuttle. The difference between the two signals calculates the surface elevation.[7]. The Horizontal Datum is WGS84 and Vertical Datum is EGM96. The spatial resolution is 1 arc-second for global coverage (~30 meters).
- Land cover was used to evaluate the suitability of land types to store the water with a minimum cost. Data was downloaded from the USGS Gap Analysis Project[8]. The distribution of native vegetation types, modified and introduced vegetation, developed areas, and agricultural areas of the U.S. is included. The spatial resolution is 30 m [8].
- Stream data was used to identify the location to drain the overflow from the stream water. The stream data was downloaded from ArcGIS Hub, and USGS Current Water Data was used to build the data for the Nation project [9]. The data contains the major rivers and streams for the whole U.S.
- Soil data was used to find locations with soil types that absorb water. Soil data was downloaded from the USDA (United States Department of Agriculture) Soil Survey Geographic Database (SSURGO) [10]. It includes information on soil collected by the National Cooperative Soil Survey [11]. Over a century, it was collected at scales ranging from 1:12,000 to 1:63,360.
- Urban data was used to point out the major city facilities and buildings. The urban data of Sacramento was downloaded from the Sacramento County Open Data Hub (<https://data-sacramentocounty.opendata.arcgis.com/>). The datasets include animal shelters, apartments, bus and train stations, fire stations, highways, hospitals, hotels, libraries, railways, stadiums, and arenas. The urban data of New York City was downloaded from <https://data.gis.ny.gov/>.

2.3. Methods and model parameters

There are two phases in the whole process of this study. In the first phase, Python 3.12.4 was used to analyze the precipitation data and found the two study areas. In the second phase, a suitability model

in ArcGIS Pro 3.1.4 was used to locate the most suitable areas to store the stormwater. The workflows of ArcGIS Pro are shown in Figure 1.

2.3.1. Using python to analyze data

After downloading the precipitation data, the data from 2000 to 2024 of California and New York state were extracted, and the mean of precipitation in summer (June, July, and August) and winter (December, January, and February) for each year was calculated [12]. Then, we calculated the annual change rate of each county for the two seasons. Finally, we identified the anomalies of the summer and winter precipitation. It turned out that the stormwater of New York City and Sacramento is worth considering based on the annual change rate of precipitation.

2.3.2. Build a suitability model to identify the area

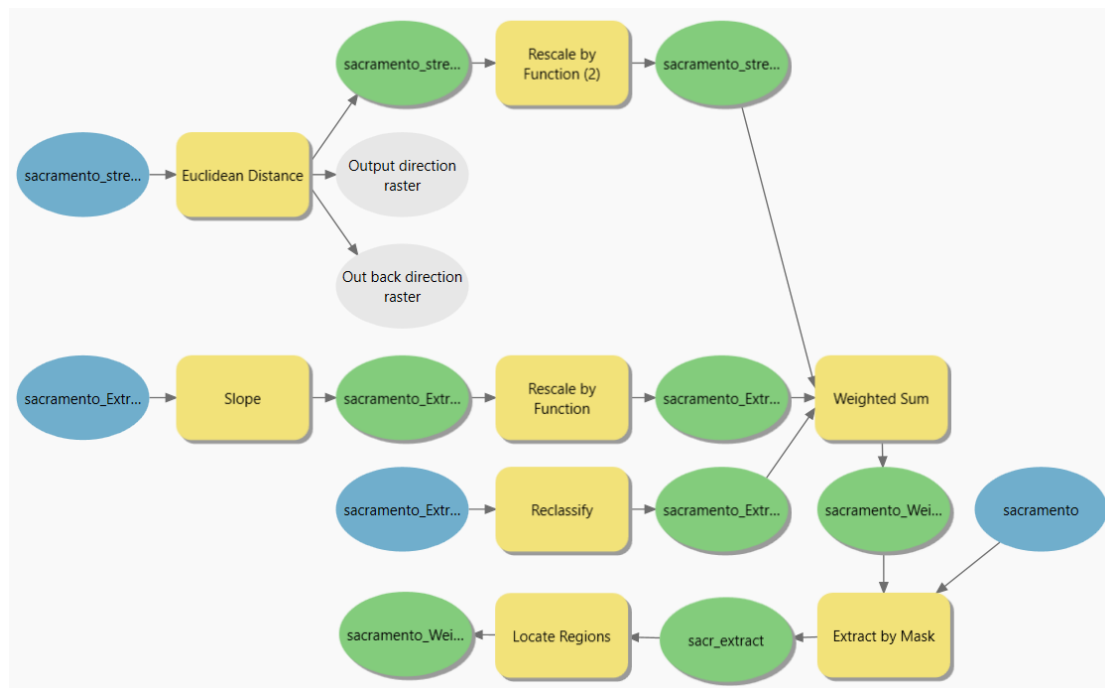


Figure 1: Workflow of suitability model

- The Euclidean Distance tool analyzes stream data preliminarily. Euclidean Distance is a commonly used definition of distance, the distance between two points in three-dimensional space. This tool calculates the Euclidean Distance from each cell to the closest source. Feature class and raster can be input as source data. The input of this tool is the clipped stream feature layer. The output cell size is 30 m, and the distance method is planar.
- The Slope tool analyzes DEM data preliminarily. This tool can calculate the slope (gradient or steepness) from each raster cell, and a low slope value means flat terrain, which is better. The Slope tool uses a 3 by 3 window of cells to compute the value, planar and geodesic calculations can be used to calculate slope. The parameters that need to be set in this tool are input raster, output raster, output measurement, method, Z factor, and target device for analysis. The input raster is DEM data, and the output measurement is the degree; the method we choose is planar, and the number of Z-factor is 1. The target device for analysis is “GPU then CPU”.
- The Rescale by Function tool analyzes the Euclidean distance tool's output and the slope tool's output. This function adjusts the input raster values by a chosen transformation method and then

maps the transformed values onto a designated continuous scale. It applies the function to the input raster values at first, and then the function value is linearly transformed to the specified evaluation scale, and the minimum and maximum function values are set from the scale and to the scale. For the Rescale by Function tool that is used to analyze the output of the Euclidean Distance tool, the option of transformation function is Logarithm, and the value of the output elevation scale is from 10 to 1. The parameter we set is according to past experiences; the places for stormwater storage tend to be close to streams [13, 14]. For the Rescale by Function tool that is used to analyze the output of the Slope tool, the transformation function option is “Logarithm,” and the value of the output elevation scale is from 1 to 10.

- The Reclassify tool analyzes landcover data preliminarily. The value in the raster can be reclassified or changed in this tool. There are two options in the reclassification parameter, the classify will open a dialog, and the unique option will show a remap table. The parameters in this tool include input raster, reclass field, reclassification, and output raster. The input raster is land-cover data, which will produce an output raster. The option of the Reclass field is “NLCD_Land.” The reverse new values of reclassification in two cities are shown in Table 1

Table 1: Value and reverse new value of Sacramento and New York City in the reclassify tool

Land type	Rescaled value of Sacramento	Rescaled value of New York City
Open Water	1	1
Perennial Snow/Ice	1	1
Developed, Open Space	10	10
Developed, Low Intensity	1	10
Developed, Medium Intensity	1	10
Developed, High Intensity	1	10
Barren Land	5	
Deciduous Forest	1	
Evergreen Forest	1	10
Mixed Forest	1	
Shrub/Scrub	10	
Herbaceous	10	10
Hay/Pasture	10	10
Cultivated Crops	5	10
Woody Wetland	3	10
Emergent Herbaceous Wetland	1	15

- The Weighted Sum tool analyzes the output data of two Rescale by function tools and the reclassify tool. The Weighted sum tool overlays the raster and summarizes them by multiplying them by a given weight. In each raster cell, the cell value is multiplied by the weight vector and then added. Input raster can be integer or floating point. The parameters in this tool are input raster, which includes raster, field, and weight. And it will produce an output raster. In the Weighted Sum tool of both areas, the weight of output data of two Rescale by function tools and reclassify tool are all 1, and the option of Field is all “VALUE.”
- Extract by Mask tool retrieves raster cell values within the boundaries specified by a mask layer. The mask can be either a raster or a feature dataset. If the mask is a feature, it is internally converted to a raster. The tool retains the cell values of the input raster that align with the mask area, and

additional attributes from the input raster are carried over to the output. The parameters in this tool include input raster, input raster or feature mask data, output raster, and extraction area. The input raster is the output of the weighted sum tool. The input raster or feature mask data is the study area. The option of extraction area is “inside.”

- The Locate Regions tool analyzes the output of the Extract by Mask tool. This tool is used to find out the regions that satisfy the specified evaluation criterion. A parameterized region-growing (PRG) algorithm is implemented in this tool. Neighboring cells are systematically integrated into the expanding regions, which originate from predefined seed cells. This addition is done in a way that optimally maintains the given shape and maximizes the region's utility. In this tool, the input raster is the output of Extract by Mask. The input content of region minimum area, region maximum area, minimum distance between regions, and maximum distance between regions are null, and the option of islands not allowed in regions needs to be checked. The rest of the input contents of parameters are shown in Table 2.

Table 2: Settings of the locate regions tool in Sacramento and New York City

Parameters	Input feature of Sacramento	Input feature of New York City
Total area	2	1
Area units	Square kilometers	Square kilometers
Number of regions	5	5
Region shape	Square	Hexagon
Region orientation	0	0
Shape/Utility tradeoff (%)	10	10
Evaluation method	Highest average value	Highest average value
Distance units	Meters	Meters
Number of neighbors to use in growth	Eight	Eight
Number of seeds to grow from	Based on input	Based on input
Resolution of the growth	Based on input	Based on input
Region selection method	Based on the number of regions	Based on the number of regions

2.3.3. Build the filters to refine the results

- Select tool is used to preliminarily analyze the soil data. It is a tool to extract some needed features from the input feature class or the input feature layer, and a select or Structured Query Language (SQL) expression is used. The features that are selected are stored in an output feature class. We wrote a query in the query builder, which is used to search for and select a subset of features and table records. The parameters we should set are input features, output features, and Expression. The input feature is soil data. The Expression in this study is “Where drclasswet is equal to Very poorly drained Or drclasswet is equal to Poorly drained Or drclasswet is equal to Somewhat poorly drained Or drclasswet is equal to Moderately well drained.” “drclasswet” refers to the natural drainage capacity of soil and is closely related to the frequency and duration of its wet periods [15]. This column displays the wettest drainage class assigned to an individual component of the map unit. A poorly drained soil is good at storing water; otherwise, the water will flow away. This tool is not used in New York analysis, because bare soil is rarely found in the New York area. The second Select tool is used to analyze the output of the first select tool. The parameters we should set are input features, output features, and Expression. The input feature is the output of the first select tool. The expression is “Where hydgrpdcd is equal to A Or hydgrpdcd is equal to B Or hydgrpdcd is equal to C” “hydgrpdcd” refers to Soils are categorized into Hydrologic Groups according to their analogous potential for runoff formation when exposed to similar rainfall and surface conditions [15]. Choosing soil with a high runoff potential can help store stormwater. Both Select tools are used in Sacramento’s model but not in New York.

- The Buffer tool produces a buffer area of input features. In this tool, the input feature is landcover data. For Sacramento, the input data is the location of apartments, bus and train stations, highways, and railways. For New York City, the input data is national register buildings. The distance around the input features that will be buffered is 100 meters, the side type is full, and the Method is planar. The locations within the buffer area will be considered non-ideal choices because building infrastructures will cost a lot.

3. Results

3.1. Result of study area selection

3.1.1. Precipitation change ratio result

The definition of abnormal precipitation change is the change ratio is greater than 100% or smaller than 100%. The number of abnormal precipitation changes was counted from 2000 to 2024. Table 3 shows the rank of abnormal precipitation changes. SA was ranked 2, and NYC had no abnormal year. The counties that are not on the table are because the number of times their precipitation is abnormal is 0.

Table 3: Number of abnormal years in some counties

County of CA	Number of abnormal years in CA	County of NY	Number of abnormal years in NY
Napa	10	Kings	35
Yolo	9	Orange	30
Contra Costa	9	Dutchess	1
Sacramento	9	Putnam	1
Merced	8	Richmond	1
Sutter	8	Suffolk	1

3.1.2. Anomaly result (z-score)

A Z-score, also known as a standard score, measures the number of standard deviations a data point is away from the mean of a distribution. The county with the biggest anomaly in CA in summer is Sacramento, it has a Z-score of -0.570278. The Z-score of precipitation in the summer of Kings is -0.209988, in Queens is -0.149991, and in New York is -0.166477.

3.1.3. Area choosing result

In CA, Sacramento was chosen as the study area because it has a big precipitation change ratio and a large anomaly. These two reasons make stormwater storage in Sacramento worth considering. In NY, though the precipitation change ratio is small and the anomaly is teeny, New York City was chosen as our study area because it has a large population, which made stormwater storage in New York City have more economic meaning.

3.2. Result of the stormwater storages

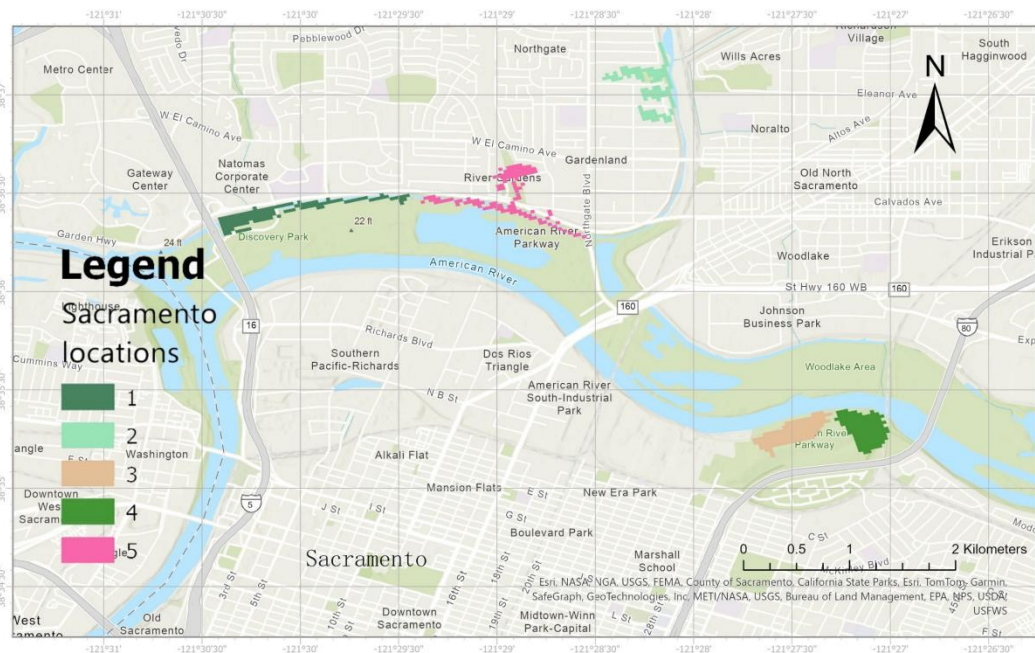


Figure 2: Locations of stormwater storage in Sacramento

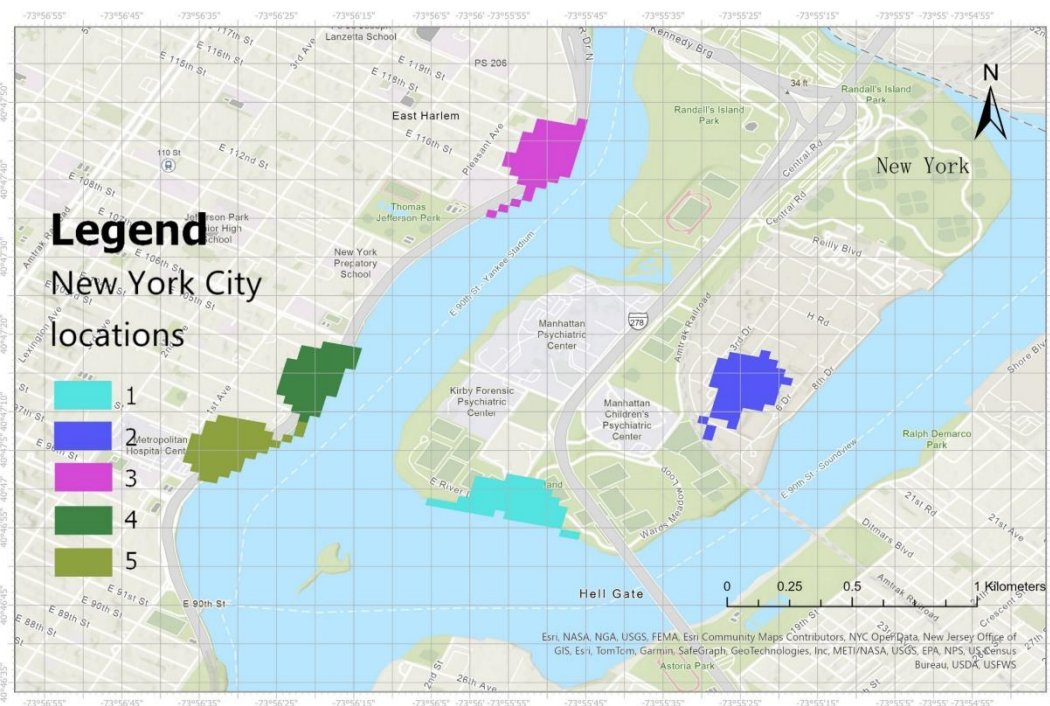


Figure 3: Locations of stormwater storage in New York City

- Based on our results for Sacramento, which is shown in Figure 2, the best stormwater storages are located near the American River. The top 1 storage is in the grassland of Discovery Park, Top 3 and top 4 storage are both in American River Parkway. The top 5 storage is in River Garden. These locations reflect 3 criteria as we mentioned in Section 2: these locations are all out of the buffer

area, so they will not influence the existing facilities; all these locations are near the water bodies, and it is easier to drain the extra water. Finally, these locations are all in parks, which have less population, and reduce the “Not in my backyard” (NIMBY) attitude [16].

- Based on our results in New York City, which is shown in Figure 3, the best stormwater storages are located near the Harlem River. The top 1 and 2 storage are in the grassland at Randalls Island. These locations reflect three criteria we mentioned in Section 2: the land cover in these locations has a relatively low cost to rebuild a stormwater storage facility, the surface in these locations is relatively flat to store the extra water, and the distance between these locations and nearby water bodies is relatively short to drain the extra water. The result also indicates that the water storage locations are clustered at Inwood and Randalls Island. That also points out our consideration of building the stormwater storage at a low cost since Inwood and Randalls Island have less population compared to other areas in New York City.

4. Discussions

Compared to the reports [13,14], the majority of locations created by our model fit. However, some locations did not match. The plausible reasons could be: 1. The input dataset of our model is different from the reports. 2. The storage of the report is based on the existing infrastructure [17]. 3. The results were made by strategic planning, not quantitative planning [18].

5. Conclusions

In this paper, New York City and Sacramento were selected as the demos to show the application of a suitability model on the potential stormwater storage. With the more severe global warming, the risk and chance of extreme precipitation dramatically increased. The advantages of using an automation process to plan or project are as follows: 1. It can be updated in a relatively short time period; 2. It can be tuned parameters; 3. It is a time-saving and cost-saving tool.

Declaration of interests

The author has nothing to disclose.

Authorship

Guangyi Zhang: Conceptualization, Investigation, Data collection, Data analysis, Data analysis, Writing, Review & editing, Visualization, Project administration, Funding acquisition.

References

- [1] Grimmond, S. U. E. (2007). *Urbanization and global environmental change: Local effects of urban warming*. *Geographical Journal*, 173(1), 83–88. https://doi.org/10.1111/j.1475-4959.2007.232_3.x
- [2] Ishmam, Z. S., Rohli, R. V., & Mostafiz, R. B. (2025). *Exploring the spatiotemporal trends of extreme precipitation indices for the continental United States*. *Physical Geography*, 46(2), 156–174. <https://doi.org/10.1080/02723646.2025.2465011>
- [3] Alcoforado, M.J., Andrade, H. (2008). *Global Warming and the Urban Heat Island*. In: Marzluff, J.M., et al. *Urban Ecology*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73412-5_14
- [4] Wang, J., Ellen, Banzhaf. (2018). *Towards a better understanding of Green Infrastructure: A critical review*. *Ecological Indicators*, 95, 1–15. <https://doi.org/10.1016/j.ecolind.2018.07.049>
- [5] Jayasooriya, V. M., Ng, A. W. M., Muthukumaran, S., & Perera, B. J. C. (2014). *Tools for modeling of stormwater management and economics of green infrastructure practices: A review*. *Water, Air, & Soil Pollution*, 225(5), 1933. <https://doi.org/10.1007/s11270-014-1933-x>

- [6] McFarland, A. R., Larsen, L., Yeshtela, K., Engida, A. N., & Love, N. G. (2019). Guide for using green infrastructure in urban environments for stormwater management. *Environmental Science: Water Research & Technology*, 5(4), 643–659. <https://doi.org/10.1039/c8ew00498f>
- [7] Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., et al. (2007). The shuttle radar topography mission. *Reviews of Geophysics*, 45(2), 361. <https://doi.org/10.1029/2005RG000183>
- [8] U.S. Geological Survey Gap Analysis Program. (2016). GAP/LANDFIRE National Terrestrial Ecosystems 2011. U.S. Geological Survey.
- [9] U.S. Geological Survey. (2025). National Hydrography Dataset. Retrieved March 1, 2025, from <https://www.usgs.gov/national-hydrography/national-hydrography-dataset>
- [10] U.S. Department of Agriculture, Natural Resources Conservation Service. (2025). Soil Survey Geographic Database (SSURGO). Retrieved March 1, 2025, from <https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo>
- [11] Schoeneberger, P. J., Wysocki, D. A., Benham, E. C., & Broderson, W. D. (1998). Field book for describing and sampling soils, Version 2.0. USDA Natural Resources Conservation Service.
- [12] Wang, J., Yang, B., Wang, Z. et al. (2025). ENSO and volcanic forcing of winter and summer interannual temperature variability in East Asia over the past six centuries. *Climatic Change* 178, <https://doi.org/10.1007/s10584-025-03864-0>
- [13] Wilk, N. (2025). Green stormwater infrastructure in Sacramento, California. *Urban Water Atlas*. Retrieved February 12, 2025, from <https://www.urbanwateratlas.com/author/nico-wilk/>
- [14] New York City Department of Environmental Protection. (2024). New York City Stormwater Manual. <https://www.nyc.gov/site/dep/stormwater/manual.page>
- [15] Natural Resources Conservation Service. (2014). SSURGO 2.3.2 Table Column Descriptions. U.S. Department of Agriculture.
- [16] Shen, C., & Wang, Y. (2023). Order from chaos: Deconstructing the interactions of multiple online stakeholders in NIMBY conflicts. *Environmental Science & Policy*, 145, 29–39. <https://doi.org/10.1016/j.envsci.2023.03.020>
- [17] Kwak, Y., Deal, B. (2024). Multi-scaled green infrastructure optimization: spatial projections and assessment for dynamic planning and design. *Landscape and Urban Planning*, 249. <https://doi.org/10.1016/j.landurbplan.2024.105128>
- [18] Vandersmissen, L., & George, B. (2023). Strategic planning in public organizations: reviewing 35 years of research. *International Public Management Journal*. <https://doi.org/10.1080/10967494.2023.2271901>