VISUALISING SYDNEY'S URBAN GREEN

A Web Interface for Monitoring Vegetation Coverage in the Greater Sydney Area between 1992 and 2022 using Google Earth Engine

AARO LAHTINEN[1], NICOLE GARDNER[2], CRISTINA RAMOS-JAIME[4] and KUAI YU[5]
[1,2,3,4]*University of New South Wales.*
[1]*lahtinenaaro77@gmail.com, 0009-0008-3281-1445*
[2]*n.gardner@unsw.edu.au, 0000-0001-6126-6716*
[3]*c.ramos@unsw.edu.au, 0000-0002-6868-8855*
[4]*daniel.yu@unsw.edu.au, 0000-0002-7788-548X*

Abstract. With continued population growth and urban expansion, the severity of environmental concerns within cities is likely to increase without proper urban ecosystem monitoring and management. Despite this, limited efforts have been made to effectively communicate the ecological value of urban vegetation to Architecture, Engineering and Construction (AEC) professionals concerned with mitigating these effects and improving urban liveability. In response, this research project proposes a novel framework for identifying and conveying historical changes to vegetation coverage within the Greater Sydney area between 1992 and 2022. The cloud-based geo-spatial analysis platform, Google Earth Engine (GEE), was used to construct an accurate land cover classification of Landsat imagery, allowing the magnitude, spatial configuration, and period of vegetation loss to be promptly identified. The outcomes of this analysis are represented through an intuitive web platform that facilitates a thorough understanding of the complex relationships between anthropogenic activities and vegetation coverage. A key finding indicated that recent developments in the Blacktown area had directly contributed to heightened land surface temperature, suggesting a reformed approach to urban planning is required to address climatic concerns appropriately. The developed web interface provides a unique method for AEC professionals to assess the effectiveness of past planning strategies, encouraging a multi-disciplinary approach to urban ecosystem management.

Keywords. Google Earth Engine, land cover classification, Landsat imagery, urban vegetation, web interface

ACCELERATED DESIGN, Proceedings of the 29th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2024, Volume 2, 515-524.

1. Introduction

Between 1990 and 2022, the population of Greater Sydney increased from approximately 3.6 to 5.2 million residents (Macrotrends, 2023). This significant population growth has resulted in rapid urban expansion, characterised by outward low-density development or urban sprawl (Chen et al., 2017). This widespread conversion of forests and grasslands to hard, impervious surfaces, coupled with past planning strategies focused on affordable housing and employment targets has generated numerous adverse environmental effects (Wu and Murray, 2003). These effects include biodiversity loss, habitat fragmentation, noise pollution, increased building energy consumption and the urban heat island (UHI) effect (Li et al., 2022). The latter, combined with climate change-induced increases in the magnitude and frequency of temperature fluctuations, is already threatening the livelihood of urban inhabitants (Li et al., 2022). With a further population growth of 3.2 million expected by 2056 (Macrotrends, 2023), the severity and regularity of these issues is likely to increase without planning reform that recognises the social, ecological and environmental value of urban vegetation (Zhang et al., 2022). The sufficient presence of healthy vegetation within an urban setting provides a multitude of benefits to city dwellers, including sequestration of carbon, improved air quality, increased evapotranspiration, maintenance of biodiversity, reduction of building energy consumption (Yang et al., 2022) and the mitigation of the UHI effect (Davies et al., 2017). Despite this, past planning strategies by the NSW government have failed to properly leverage these benefits due to a lack of ecological knowledge (Davies et al., 2017).

Few researchers have explored the possibilities of sharing ecological information in an intuitive, timely and easily accessible manner. Therefore, a framework for delivering accurate knowledge of the state of urban vegetation to AEC professionals could encourage a multidisciplinary approach to urban planning and influence future regulation on vegetation management. The research detailed in this paper explores the possibilities of using the cloud-based geo-spatial analysis platform, Google Earth Engine (GEE), to develop a web interface for facilitating the transfer of this knowledge (Gorelick, 2017). The initial attainment of information involved an extensive review of previous historical vegetation analysis studies. Through this, a structured framework was devised using Landsat and a supervised machine learning model to create an accurate land cover classification. The results of the analysis were communicated through a variety of visualisation techniques, including land cover maps; depicting geo-physical characteristics under five distinct land cover classes, land surface temperature maps; showing areas most affected by the UHI effect, 'true colour' maps; displaying a natural colour composite of satellite imagery and line charts that compare the trends of population and land cover types. To facilitate the transfer of this information to AEC professionals concerned with mitigating the UHI effect and improving urban liveability, a user-friendly and functional web interface was developed. The completed web interface will allow built environment professionals to evaluate past planning strategies and make informed decisions about future urban ecosystem management.

2. Literature Review

Satellite remote sensing is a common method for understanding the interactions between human activities and natural environments (Zhang et al., 2022). This process involves monitoring the geo-physical characteristics of a region of interest (ROI) by measuring its radiation emission and reflectance from a distance (U.S. Geological Survey, 2023). It is especially valuable for the detection of land use and land cover (LULC) changes, as direct field studies are costly, time-consuming, and prone to human error (Coppin et al., 2004; Lu et al., 2004). Recent studies in the field of change detection have placed a significant emphasis on the identification and mapping of urban vegetation, due to the rapid expansion of cities and growing climatic concerns (Chen et al., 2017; Li et al., 2022; Zhang et al., 2022). The 'Landsat' collection (30 m pixel resolution), since its launch in 1972, has been the predominant choice among scholars for urban vegetation mapping (Coppin et al., 2004; Lu et al., 2004; Zhang et al., 2022). Like other satellite imagery, Landsat captures multispectral data by measuring radiation reflectance values across multiple bands, with each band corresponding to a particular wavelength range in the electromagnetic spectrum (Coppin et al., 2004).

Recent studies have employed machine learning (ML) algorithms as a method of classifying surface characteristics due to their ability to produce high accuracies in complex, heterogeneous environments (Gibril et al., 2018). Supervised learning models are the most documented algorithm for classifying LULC, as class labels are generally pre-defined under four major land cover types: water bodies, built areas, bare land and vegetation (Mehmood et al., 2023). A review of popular supervised models revealed that random forest (RF) algorithms consistently outperformed other models in a complex urban setting (Chaturvedi & de Vries 2021). Applications of this method by Duncan and Boruff (2023) and Shrestha (2023) both succeeded in accurately characterising changes over 20 years (1999 - 2019) to the abundance and configuration of vegetation within the Greater Perth and Alabama metropolitan areas, respectively and concluded that RF was an appropriate method for evaluating the effectiveness of vegetation management initiatives and policy changes. They also suggested that increases to dwelling and population density resulted in changes to urban vegetation, reflecting the importance of including external datasets in the analysis process to ensure that drivers of change are properly understood. A comparison of these results with the earlier image differencing models revealed a distinct discrepancy in the level of detail produced, with ML models demonstrating a significant advantage in their ability to classify sporadic clustering of vegetation precisely. Another significant advantage is the ability to validate the results of a specific classification model with an accuracy assessment that determines whether each pixel has been correctly fitted to the appropriate class. This form of immediate feedback allows researchers to make guided adjustments to their analysis techniques, ensuring that accurate results are produced. Conclusions of similar studies by Gibril (2018) and Hoang and Tran (2021) emphasised the significance of developing a streamlined framework for clearly communicating the results of multi-year vegetation analysis to ensure appropriate changes are made to future urban planning and management strategies.

For explicitly mapping changes across a time series of land cover maps, image differencing is used to identify alterations between two classes (Afify, 2011). A typical example used by researchers was a transition from grassland or forest to built land, with changed pixels visualised as red and unchanged pixels as white (Basheer et al., 2013). This process is beneficial for identifying the extent of urban expansion and consequent vegetation loss within cities, however, it does not necessarily illuminate key periods of temporal change. For this, a calculation of class area across a time series of images can be used to create line graphs highlighting the temporal relationships between urbanisation and vegetation (Zhang et al., 2022). By demonstrating the results of historical vegetation analysis through a variety of visualisation techniques, conclusions can be quickly made about the magnitude, spatial configuration and period of vegetation loss.

Although there are web platforms that allow built environment professionals to visualise geo-spatial data, few of these provide accurate mapping of historical changes to vegetation coverage. Conclusions by Mehmood (2023) and Zhang (2022) suggested that clear and timely communication of the state of urban vegetation is imperative for urban planners as they devise new strategies to mitigate the UHI effect and improve urban liveability. As such, a tool allowing design professionals to engage with expert knowledge and data relating to the complex nature of spatio-temporal changes to vegetation could be useful to inform future regulatory frameworks and urban planning strategies. Many of the investigated papers used Google Earth Engine and other GIS software such as ArcGIS to accurately analyse urban vegetation coverage (Basheer et al., 2022), however, they did not convert this information into an accessible and shared format. The possibilities of using Google Earth Engine to both analyse satellite imagery and design a functional user interface will be explored in this paper.

3. Methodology

This case study details how Google Earth Engine can be used to develop a web interface for visualising changes to historical vegetation coverage within the Greater Sydney area between 1992 and 2022. Within this framework, a supervised classification model was applied to a sequence of Landsat imagery to accurately identify vegetation coverage within a complex, heterogeneous urban environment. The results of this analysis were communicated through raster maps and trend charts within an interactive web platform, allowing built environment professionals to understand the magnitude, spatial configuration and period of historical vegetation loss in an intuitive and simple manner.

3.1. DATA

Satellite imagery from the Landsat program was used for the analysis due to several determining factors, including an extensive and consistent catalogue of imagery, a 30 m pixel resolution and its multi-spectral capabilities. Out of the six available Landsat collections, four were used to ensure a consistent sequence of imagery was achieved across the 30-year time span. These are listed along with their operational years:

- Collection 5 (1984 to 2012), collection 7 (1999 to 2021), collection 8 (2013 to present) and collection 9 (2021 to present).

3.2. FILTER

Each satellite collection contains global imagery from the entire span of its operation, meaning a large portion is not required for a region and time-specific analysis. To resolve this, a function was created and applied to each collection that iteratively filtered images based on a set of given parameters. These parameters were determined based on the region, time and cloudiness of scenes within the collection. Any images not overlapping with the Greater Sydney region were removed to eliminate unnecessary data processing. A time span between 1992 and 2022 was established to ensure a comprehensive analysis was conducted. Within this span, an interval of three years was determined to limit short-term fluctuations such as seasonal variability yet still accurately capture significant changes to vegetation extent. This operation culled any images that were captured outside of the following years: 1992, 1995, 1998, 2001, 2004, 2007, 2010, 2013, 2016, 2019 and 2022. Finally, any images that contained over 10% cloud coverage were removed to ensure a high level of accuracy and reliability was maintained in the analysis.

3.3. IMAGE PROCESSING

A function was created to detect cloud coverage and over-saturation within an image and then mask over any pixels within these areas. The band values of masked pixels are set to 'null,' meaning they will not be used as data points in the analysis. Another fundamental part of the pre-processing stage is normalisation, which involves the remapping of band values. The default range for any given band is 1 to 65455, meaning large fluctuations between recorded reflectance values of both bands and images is likely. These significant discrepancies can make distinguishing between different geo-physical features difficult and in turn substantially decrease analysis accuracy. To mitigate these issues, the original range was scaled to achieve consistency across all bands and images. The new range was defined as -1 to 1, aligning with commonly used normalised index values. With band standardisation completed, the images are now appropriately processed to ensure a reliable and precise analysis can be performed.

3.4. SEQUENCE

A unified sequence of imagery was achieved by first merging the collections and then sorting the combined collection by date to ensure chronological order. The combined collection was partitioned into 11 segments that corresponded with each year in the sequence. The total image count for each segment was calculated:

1992	1995	1998	2001	2004	2007	2010	2013	2016	2019	2022
15	20	17	19	58	49	33	49	42	63	35

A 'for' loop was created to apply a median operation to each imagery segment. This function calculates the median of all values at each pixel across the entire stack of 12 bands (Gorelick, 2017). The outcome is a series of 11 images that are unaffected by seasonal variability and cover the entire Greater Sydney region. An example is provided in Figure 1.

Figure 1: True Colour Composite 1992

3.5. CLASSIFY

A thorough investigation of machine learning models identified the random forest algorithm as the most accurate and reliable due to its ability to limit the under and over-fitting of classes. The land cover classification process is detailed under four main headings: training data, classifier build, accuracy assessment and area calculation. This structured method was applied to each image separately, to ensure variability across images did not affect the accuracy and reliability of the classification.

3.5.1. Training Data

Five land cover classes were chosen for the supervised classification model: water, built-forms, bare land, grassland, and tree cover. For each of these classes, a set of training polygons was manually created by identifying and labelling instances of the specific land cover type using a 'true colour' visualisation of the image. The number of polygons defined for each class ranged from 90 to 110. This small range ensured that each class was represented equally within the classification model. The spatial configuration of the polygons was altered across images to ensure that land cover types were not incorrectly labelled due to landscape changes.

3.5.2. Classifier Build

After a set of polygons had been created for each land cover class, they were combined to form a unified dataset. This collection was then split into two subsets, with 80% designated for training the classifier and 20% assigned for the later accuracy assessment. GEE facilitates the construction of classification models through several built-in machine-learning functions, which take the training polygons and image bands as inputs. A random forest algorithm consisting of 10 independent decision trees was selected for the classification. Each decision tree is constructed using a random section of data from the training subset, to ensure that no single class dominates the pattern detection process. Beginning with a root node, each tree recursively splits data based on specific thresholds, which are determined by the band values associated with each polygon sample. With each split created, the heterogeneity of classes is progressively minimised until an accurate class prediction can be made. The estimates of each decision tree are collected and a majority vote is conducted to determine the most probable class. This process is applied to each pixel within the image, resulting in a complete land cover classification that contains water, built-land, bare land, grassland and tree cover classes.

3.5.3. Accuracy Assessment

An accuracy assessment was conducted to validate the classifier's effectiveness by overlaying the validation subset (20%) onto the classified image. Each polygon's labelled land cover class was compared with the underlying classification. Depending on the amount of correctly classified pixels within the polygon, a score between 0 and 1 was assigned. The average of these scores was calculated, resulting in an overall accuracy percentage. The recorded accuracy of each classifier within this method was above 90%. It is noted by most papers discussed in the literature review, that a classification accuracy higher than 85% is deemed appropriate for a reliable identification of urban vegetation (Chaturvedi and de Vries, 2021). Therefore, each land cover classification accurately depicts the extent of urban vegetation within the Greater Sydney area.

3.6. WEB INTERFACE

The finished interface (Figure 2) is a functional and easy-to-use web platform that allows the industry partner and AEC professionals to view the outcomes of the research project in a timely, efficient, and interactive manner. The intuitive nature of the application ensures that insights about key spatio-temporal changes to vegetation can be quickly determined. The information obtained from the interface enables AEC professionals to validate the effectiveness of past planning strategies and devise smarter decision-making for future sustainable management of urban ecosystems and inhabitant satisfaction.

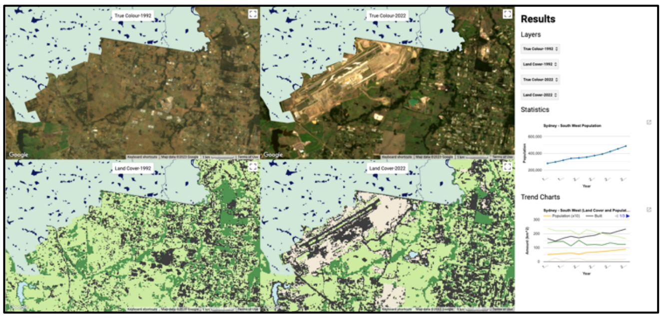

Figure 2: Web Interface

4. Findings

A thorough study of two SA4 areas was conducted to demonstrate the effectiveness of the web platform as a method of identifying critical changes to vegetation coverage.

4.1.1. Outer Southwest

There was a steady population increase between 1992 and 2022 in the Outer Southwest region, from 192,420 people to 303,902 people. To accommodate this growth, consistent urban expansion occurred within a critical area, located in the northern part of the region. The total area of built land increased from 138 to 206 km², leading to a significant decrease in grassland cover, from 437 to 372 km², indicating most developments occurred on greenfield sites. The proportion of grassland coverage declined by approximately 5%, with built land increasing by 6%.

4.1.2. Blacktown

Like the Southwest, Blacktown's population substantially increased between 2013 and 2022, with a rise from 327,424 people to 414,725 people. A key area of development was identified in the northern part of the region (Figure 3). Changes to land surface temperature within this area were investigated, indicating an increase in heat that corresponded with the spatial configuration of the new development. The proportion of grassland decreased by a staggering 9%, with built land increasing by roughly 6% between 2013 and 2022.

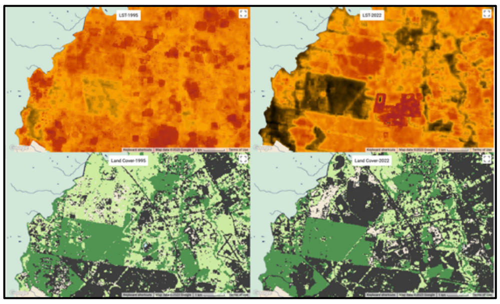

Figure 3: LST and Land Cover Changes 1995 to 2022 in

5. Discussion

A rigorous testing of the web interface was conducted to evaluate the platform's functionality, useability and applicability to the AEC industry. The timely identification of critical vegetation loss across a sequence of images was recognised as a primary benefit of the platform. This method was used to recognise a direct relationship between rapid urban expansion and heightened land surface temperature in the Blacktown area. Industry professionals can use these insights to determine how similar urban development may affect the micro-climate of a designated region, encouraging an informed approach to mitigating the UHI effect and improving urban liveability. Another valuable feature is the ability to determine trends of different land cover types. These observations can be used to accurately track the rate of vegetation loss and determine which land types are being cleared to accommodate new developments. Therefore, the web interface demonstrates notable applicability to AEC professionals concerned with improving Sydney's urban ecology.

Due to the limited ten-week timespan assigned for the research project, ground truth data still needed to be implemented to further validate the classifier's reliability. Although each land cover type was carefully chosen and equally represented in the classification model, manual labelling of classes could be considered a subjective method. The use of ground truth data, in conjunction with the accuracy assessment completed, would further validate the reliability and accuracy of the process. Also, due to the project's limited scope, a few features were omitted from the developed framework, including a change map; showing areas where a transition between two land cover types had occurred and a future prediction model, which would estimate prospective land cover changes. Further developments of this research project could investigate the potential of using a regression model to forecast future impacts of vegetation loss on urban climate and liveability. This could further encourage AEC professionals to adopt a planning approach that recognises urban vegetation's social, ecological and environmental value.

The developed web interface is a unique example of how knowledge, relating to the complex nature of spatio-temporal changes to vegetation can be shared with AEC professionals in an intuitive and accessible manner, to ensure future planning approaches recognise the ecological and environmental value of urban vegetation.

6. Conclusion

The research detailed in this paper explored the possibilities of using GEE to develop an accessible web platform for monitoring vegetation extent in the Greater Sydney area between 1992 and 2022. A supervised classification model was applied to a sequence of Landsat imagery to create an accurate and reliable estimate of urban vegetation. The results of this analysis were visualised in a clear and concise manner through a variety of raster maps and trend charts, allowing AEC professionals to easily interpret complex spatio-temporal changes to vegetation. A functional and accessible web interface was developed to facilitate the transfer of this knowledge to urban planners, encouraging an informed approach to mitigating the UHI effect and improving urban liveability.

With continued population growth and urban expansion, the severity of climatic concerns within cities is likely to increase without proper urban ecosystem monitoring and management. The research detailed in this paper provides a unique and intuitive method for AEC professionals to assess the effectiveness of past planning approaches and encourages a multidisciplinary strategy for urban ecosystem management. How could an informed approach to urban ecosystem management affect the growing climatic concerns within cities?

Acknowledgements

I would like to thank my supervisors Dr Nicole Gardner, Cristina Ramos Jaime and Daniel Yu for their continuous support and exceptional guidance throughout the research project.

References

Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T., & Liu, S. (2022). Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques. *Remote Sensing*, *14*(19), Article 19. https://doi.org/10.3390/rs14194978

Chaturvedi, V., & de Vries, W. T. (2021). Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Science*, *5*(3), Article 3. https://doi.org/10.3390/urbansci5030068

Chen, B., Nie, Z., Chen, Z., & Xu, B. (2017). Quantitative estimation of 21st-century urban greenspace changes in Chinese populous cities. *Science of The Total Environment*, *609*, 956–965. https://doi.org/10.1016/j.scitotenv.2017.07.238

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital Change Detection Methods in Ecosystem Monitoring: A Review. *International Journal of Remote Sensing - INT J REMOTE SENS*, *25*, 1565–1596. https://doi.org/10.1080/0143116031000101675

Davies, P., Corkery, L., Nipperess, D. A., van den Berg, F., Joei, C., Bishop, M., Staas, L., Hose, G., Pelleri, N., Osmond, P., Keane, A., Hochuli, D., Barnett, G., Lin, B., Threlfall, C., & Wilkinson, S. (2017). *Urban ecology: Theory, policy and practice in New South Wales, Australia*. National Green Infrastructure Network.

https://www.mq.edu.au/research/research-centres-groups-and-facilities/secure-planet/centres/centre-for-green-cities/publications
Duncan, J. M. A., & Boruff, B. (2023). Monitoring spatial patterns of urban vegetation: A comparison of contemporary high-resolution datasets. *Landscape and Urban Planning*, *233*, 104671. https://doi.org/10.1016/j.landurbplan.2022.104671
Gibril, M., Idrees, M., Yao, K., & Shafri, H. (2018). Integrative image segmentation optimization and machine learning approach for high quality land-use and land-cover mapping using multisource remote sensing data. *Journal of Applied Remote Sensing*, *12*, 1. https://doi.org/10.1117/1.JRS.12.016036
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031
Hoang, N.-D., & Tran, X.-L. (2021). Remote Sensing–Based Urban Green Space Detection Using Marine Predators Algorithm Optimized Machine Learning Approach. *Mathematical Problems in Engineering*, *2021*, 1–22. https://doi.org/10.1155/2021/5586913
Li, R., Zheng, S., Duan, C., Wang, L., & Zhang, C. (2022). Land cover classification from remote sensing images based on multi-scale fully convolutional network. *Geo-Spatial Information Science*, *25*(2), 278–294. https://doi.org/10.1080/10095020.2021.2017237
Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004). Change Detection Techniques. *International Journal of Remote Sensing*, *25*.
Mehmood, M. S., Rehman, A., Sajjad, M., Song, J., Zafar, Z., Shiyan, Z., & Yaochen, Q. (2023). Evaluating land use/cover change associations with urban surface temperature via machine learning and spatial modeling: Past trends and future simulations in Dera Ghazi Khan, Pakistan. *Frontiers in Ecology and Evolution*, *11*. https://www.frontiersin.org/articles/10.3389/fevo.2023.1115074
Macrotrends. (2023). *Sydney, Australia Metro Area Population 1950-2023*. Retrieved June 12, 2023, from https://www.macrotrends.net/cities/206167/sydney/population
Shrestha, M., Mitra, C., Rahman, M., & Marzen, L. (2023). Mapping and Predicting Land Cover Changes of Small and Medium Size Cities in Alabama Using Machine Learning Techniques. Remote Sensing, 15(1), Article 1. https://doi.org/10.3390/rs15010106
U.S. Geological Survey. (2023). *What is remote sensing and what is it used for? | U.S. Geological Survey*. https://www.usgs.gov/faqs/what-remote-sensing-and-what-it-used
Wu, C., & Murray, A. T. (2003). Estimating impervious surface distribution by spectral mixture analysis. *Remote Sensing of Environment*, *84*(4), 493–505. https://doi.org/10.1016/S0034-4257(02)00136-0
Yang, Y., Song, F., Ma, J., Wei, Z., Song, L., & Cao, W. (2022). Spatial and temporal variation of heat islands in the main urban area of Zhengzhou under the two-way influence of urbanization and urban forestry. *PLoS ONE*, *17*(8), e0272626. https://doi.org/10.1371/journal.pone.0272626
Zhang, L., Yang, L., Zohner, C. M., Crowther, T. W., Li, M., Shen, F., Guo, M., Qin, J., Yao, L., & Zhou, C. (2022). Direct and indirect impacts of urbanization on vegetation growth across the world's cities. Science Advances, 8(27), eabo0095. https://doi.org/10.1126/sciadv.abo0095