Oscar A. Alvarado

A REVIEW OF QUANTITATIVE TOOLS FOR MEASURING MULTIPLE URBAN ECOSYSTEM SERVICES

Abstract
A growing number of tools have been developed to measure the benefits that ecosystems provide to human well-being. However, few of these tools have been specifically designed for cities, which could be a reason for their limited adoption by urban decision-makers and spatial planners. This has resulted in widespread under-estimation of the ecosystem services that nature-based solutions can provide within cities. This study applied a comprehensive and systematic methodology for screening, comparing, scoring and ranking ecosystem services assessment tools according to scientific criteria and practical considerations. A scoring matrix was developed to evaluate the suitability of open-access, quantitative tools in capturing multiple ecosystem services across different urban landscape domains and societal challenges. This matrix was composed of evaluative criteria derived from the scientific literature, the needs and capabilities of practitioners, and expert interviews.

Keywords: urban ecosystem services, quantitative assessment tools, comparative tools evaluation, nature-based solutions

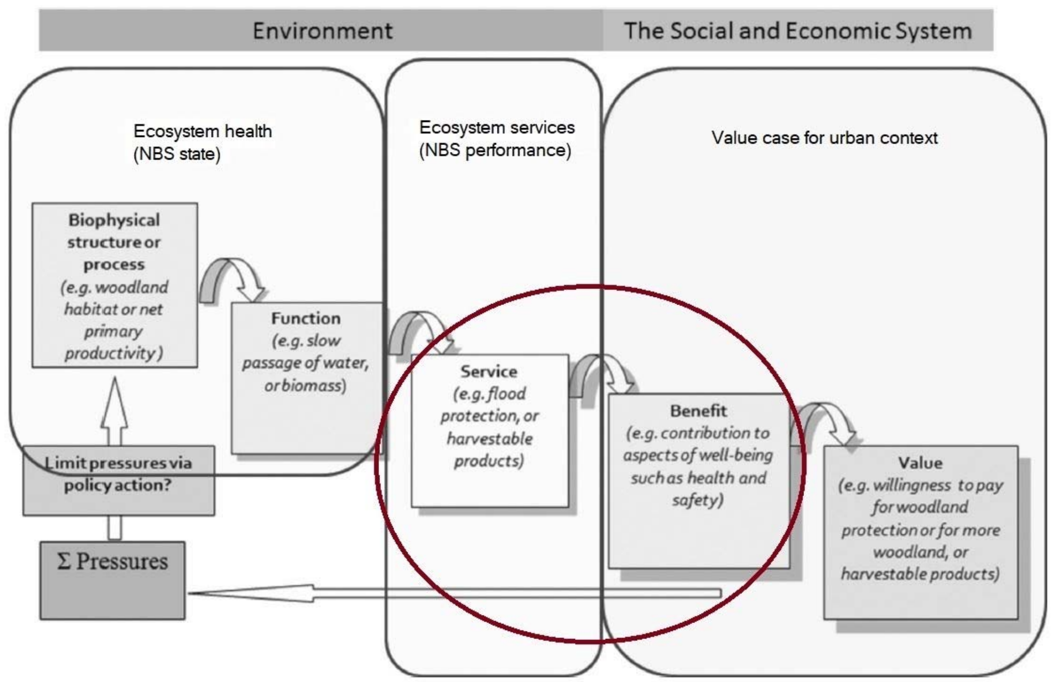

Fig. 1. Modified version of the cascade model for ecosystem services [19]. The red circle represents the elements of the model that ecosystem service assessment tools are often designed to capture.

Introduction
As cities and their populations continue to expand at an unprecedented rate, natural landscapes around the world continue to be transformed into sprawling urban settlements where 'grey' urban infrastructure such as roads, pavements, buildings and constructed assets displace previously existing natural habitats and ecosystems [1]. These patterns of development place intense pressure on local, regional and global natural ecosystems, resulting in extensive habitat fragmentation, biodiversity loss, collapse of natural resources and degradation of important ecosystem functions [2], [3]. At the same time, growing urban populations are facing numerous environmental, socioeconomic and public health challenges that significantly impact the liveability of cities. Lately, there has been growing acknowledgment of the role that nature-based solutions (NBS) can play in addressing these societal challenges while simultaneously providing a range of long-term benefits to human well-being and biodiversity (Fig. 1). Yet the unceasing, large-scale loss of natural areas in cities implies that local authorities are failing to consider the benefits of NBS, and thus incorporate their value, into urban spatial planning decisions [4].

A number of authors have highlighted the urgent need to compare, test and validate the performance of a rapidly growing range of benefit assessment tools across a wide variety of settings [5]–[7]. This study aims to apply a quantitative and systematic approach for screening, evaluating, comparing and scoring a variety of 'off the shelf' assessment tools that capture multiple ecosystem services (ES) from NBS in an urban context. The guiding research questions for this study are:

1. What are the main strengths and weaknesses of a selection of existing assessment tools for measuring multiple ES from a wide range of urban NBS?
2. How can such assessment tools be evaluated and ranked in terms of scientific validity as well as feasibility for everyday use in measuring ES?

Methodology
For this study, ES that were deemed relevant for common urban societal challenges and quality of life indicators, as classified by [8] and [9], were prioritized. As the main focus of this study, ES assessment tools are designed to provide comprehensive, replicable and quantifiable estimates of ES provision as well as their impact value to human welfare [10]. As opposed to ad hoc measurements of single ecosystem services, 'pre-packaged' or 'off the shelf' assessment tools are flexible enough in scope (multiple ES) and scale (spatial and temporal) to capture the multi-functional characteristics of NBS and their contribution to human well-being [2] while facilitating consistency in comparative analyses [7].

The overall approach (Fig. 2) first sought to identify and understand the current selection of assessment tools developed for measuring a wide range ES in diverse landscape settings. Only those assessment tools that met the screening criteria were chosen for further evaluation. The criteria used for scoring the remaining assessment tools were developed by the author and adapted from the literature. Finally, a scoring matrix was developed for ranking the assessment tools according to the selected criteria.

The identification of existing ES assessment tools was carried out through a review of online platforms related to NBS and/or ES (Ecosystems Knowledge Network, OpenNESS, Oppla), participation in NBS workshops, and a search on Google Scholar using the keywords "urban multiple ecosystem services assessment tools". A total of 30 'off the shelf' ES assessment tools were identified during this initial process.

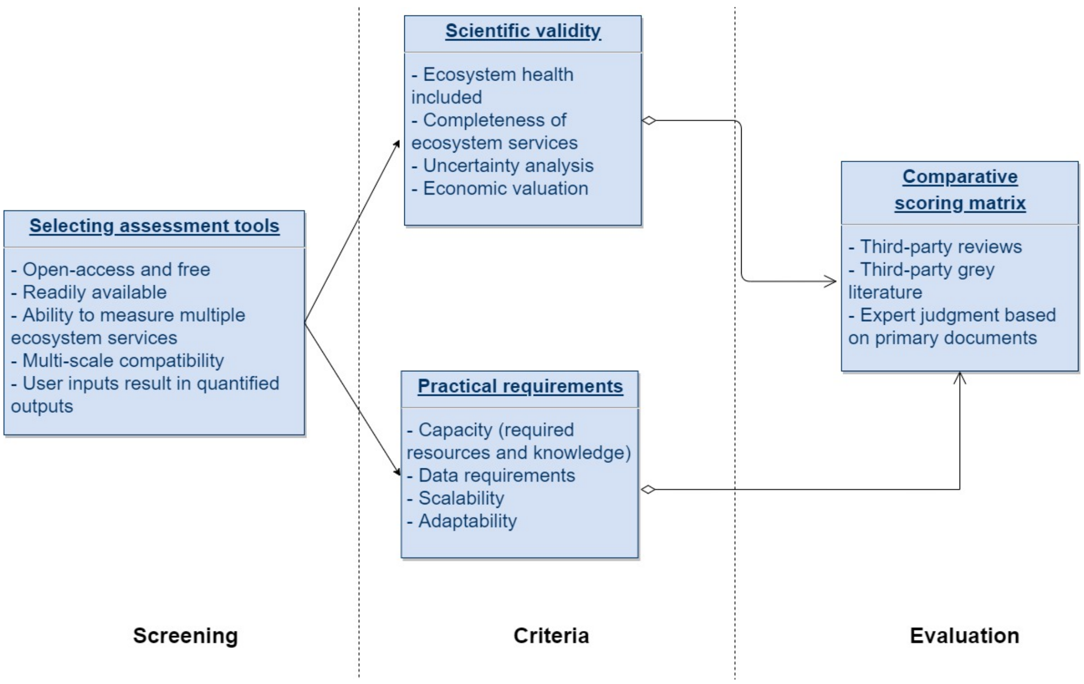

Fig. 2. Research framework.

Given the wide range of identified tools, screening criteria were developed (Tab. 1) to specifically select 'off the shelf' assessment tools that could be readily employed without proprietary restrictions across relevant urban scales to quantify multiple ES through the use of user generated data.

Scoring criteria

To systematically evaluate and compare the selected assessment tools, a set of criteria was developed for which scores would be applied. These criteria aim to address two equally important perspectives when undertaking ES assessments; scientific validity and practical requirements. The selection of criteria began with a desktop review of scientific and grey literature regarding previous urban ES assessments. Specifically, the work of [11] in synthesizing and organizing criteria for selecting individual indicators for ES was adapted for assessment tools. Additionally, informal discussions were conducted with a range of actors including colleagues, external researchers, and local proponents of NBS in cities. The purpose of these discussions was to understand the needs and priorities of different NBS stakeholders that may not be immediately evident when reviewing literature.

From a scientific validity perspective, the inclusion of *ecosystem health* and performance indicators within the each assessment tool was deemed to be of critical importance during the evaluation phase. An ecosystem's *structure* and *integrity* are key factors that enable resistance and recovery (i.e. resilience) to the kind of external perturbations that NBS often face in urban environments [12]. It is difficult to understate the importance that resiliency has on an ecosystem's ability to provide future flows of services, particularly those related to risk reduction and climate change adaptation. Thus in order to be relevant for urban areas, any comprehensive assessment of the performance of NBS should be able to link ecosystem service flows to the underlying ecosystem's structure and integrity.

To fully capture the multiple dimensions of human well-being that NBS can positively affect, assessment tools should encompass all three categories of *ES*; provisioning, regulating, and cultural [7]. Furthermore, the ability to include as many relevant ES as possible in an assessment can lead to a more in-depth understanding of the complex relationships between ES (i.e. trade-offs, synergies and bundles) [13]. Table 2 presents a checklist of relevant urban ES that was used to determine how each assessment tool scored in terms of *completeness*.

The inherent complexity of NBS, and the underlying ecosystem properties and functions, implies a great deal of *uncertainty* when attempting to quantify ES [14]. This uncertainty may differ across assessment tools, complicating the comparison and interpretation of outputs. Therefore, it is important to consider whether each assessment tool acknowledges this uncertainty and how it is incorporated into final outputs.

Some assessment tools include their own economic analysis of measured ES. Incorporating this feature within an assessment tool is beneficial but not critically necessary given the existence of independent valuation methods for ES, however discrepancies can arise when combining different economic approaches and so as much consistency as possible is preferred [2]. It should be noted that the inclusion of monetary valuation within an assessment tool does not prevent additional forms of ES valuation.

Criteria	Description	Rationale
Open access and free	Open access and free to use without the purchase of software licenses or contracting with third parties	Assessment tools that are in the public domain allow them to be independently applicable without restrictions for the current study and future research
Readily available	Available in English as of 20/04/2019	To ensure that every assessment framework included for evaluation is sufficiently well-developed and well-documented, which promotes greater transparency and credibility
Measure multiple ES	Explicit, central focus on measurement of multiple indicators for several ES	NBS are multifunctional by nature therefore capturing as many ES as possible is highly desirable
Multi-scale compatibility	Compatible with site, local and/or landscape scales that are most relevant to urban contexts (i.e. NBS, neighborhood, city scales)	Assessment frameworks that are applicable across multiple spatial scales are attractive because it is easier to learn one tool than many
User inputs results in quantified outputs	Quantitative and qualitative data can be input into the assessment tool to obtain outputs that reflect ES provision and distribution (supply)	Quantified results are essential for measuring ES and their trade-offs, though qualitative assessments are also appropriate for cultural ES and ranking preferences

Tab. 1. Screening criteria to select assessment tools.

Category	Urban ES	InVEST	ARIES	SolVES	B£ST	ESTIMAP	i-Tree
Provisioning	Food supply	√	√	√		√	
	Fresh water supply	√	√		√		
	Raw materials			√			
Regulating	Local cooling effects						
	Building energy use				√		√
	Air purification					√	√
	Carbon sequestration	√	√		√		√
	Water flow regulation	√	√		√	√	√
	Sediment regulation		√				
	Wastewater treatment	√	√		√		
	Coastal protection	√	√			√	
	Pollination	√	√			√	
	Noise attenuation				√		
Cultural	Recreation	√	√	√	√	√	
	Mobility				√		
	Aesthetic	√	√	√		√	
	Mental health			√			
	Physical health				√		√
	Child development						
	Spiritual			√			
	Social cohesion						
	Education			√	√		
	Jobs / productivity				√		
	Property prices						√
	Tourism			√	√	√	
	Heritage			√			
	Bequest			√			
Total ES included (max = 28)		9	10	10	12	8	6

Tab. 2. Checklist for measuring ES completeness of each tool.

From a practical perspective, a clear understanding of the *feasibility* requirements for using each assessment tool is essential for widening their adoption among technical and non-technical decision-makers [6]. Common limitations and barriers for the use of existing assessment tools include *data and resource needs*, especially *human or technical capacities* [15], as well as the *flexibility* to apply the same assessment tool over *multiple scales*. Therefore, the inclusion of user-centered feasibility criteria reflects the perceived and actual needs of end users undertaking ES assessments [11].
For each of the above criteria, every selected assessment tool was quantitatively evaluated through a scoring matrix (Fig. 3) with scores (ranging from 0-3 for simplicity) and colors assigned to each tool based on a review of relevant sources (Tab. 3). The creation of a scoring matrix allowed for a direct comparison of several assessment tools based on cumulative scores and resulted in one tool achieving the highest score based on the aforementioned criteria.

Results

The final six assessment tools that met the criteria of the screening process (Tab. 1) were selected for more in-depth evaluation. This final list features readily available assessment tools that are replicable and flexible enough for use in diverse but relevant urban spatial scales, using context-specific data to quantify multiple urban ES within the tool's own framework.
Fig. 3 shows the results of the scoring matrix for the selected assessment tools, including the nine evaluation criteria, total cumulative scores and color scheme used to highlight the type of literature source that was used for each assigned score. Overall, the six tools mostly focused on provisioning and regulating ES (Tab. 2) while cultural ES were often poorly taken into account within the tools' capabilities, thus ignoring the contributions that NBS can provide to social well-being, not just physical well-being. An obvious exception to this is the SolVES tool, which has been specifically designed to measure societal valuations of ES and thus is better equipped to capture cultural ES (Tab. 2).

Source	Tool					
	InVEST	ARIES	SolVES	B£ST	ESTIMAP	i-Tree
Peer-reviewed scientific papers that independently reviewed or applied a tool						
Abd-Elrahman et al., 2010						√
Vigerstol & Aukema, 2011	√	√				
Bagstad et al., 2013	√	√	√			
Nemec and Raudsepp-Hearne, 2013	√	√	√			
Peh et al., 2013	√	√				
Russo et al., 2014						√
Baró et al., 2015						√
Morales-Torres et al., 2016				√		
Ossa-Moreno et al., 2017				√		
Stange et al., 2017					√	
Zulian et al., 2018					√	
Third party grey literature that provide summaries and descriptions for a tool						
Ecosystems Knowledge Network, n.d.	√	√		√		√
Peer-reviewed scientific papers authored by a tool's own developers						
Sherrouse et al., 2011			√			
Zulian et al.,, 2014					√	
Ashley et al., 2017				√		
Primary technical documents per tool						
Zulian et al., 2013					√	
Villa et al., 2014		√				
Sherrouse & Semmens, 2015			√			
Sharp et al., 2018	√					
(Horton et al., 2019)				√		
USDA Forest Service, 2019						√

Tab. 3. Scientific papers, grey literature and primary tool documents used in scoring matrix.

In terms of spatial scalability, analysis was limited to whether tools could capture ES at scales relevant to urban settings, namely site (i.e. park), local (i.e. neighborhood) and landscape (city-wide). Most tools (except B£ST) were able to measure ES at a city-wide landscape scale, which would allow for an integrated assessment of ES across heterogenous urban settings, including surrounding peri-urban areas. In the end, *i-Tree Eco achieved the highest cumulative score* out of all assessment tools, narrowly beating out ARIES and B£ST.

Discussion

The wide range of available 'off the shelf' ES assessment tools necessitated narrowing down the list to a more manageable number for evaluation and scoring. Screening criteria (Tab. 1) reflected practical needs (open access, free, readily available) and the type of assessment desired (multiple ES, urban scales, quantified outputs). Some tools still in the prototype stage (LUCI, Naturvation Index) could eventually meet the current screening criteria once fully developed, while other tools restricted to narrow geographical areas (GI-Val, TESSA) are planning to expand their transferability in the future. In particular, TEEB Stad and Atlas atlas Natuurlijk Kapitaal may be promising options for future ES assessments based on their incorporation of extensive data sets and models, though they are limited by their availability in Dutch only.
By modifying the screening criteria according to specific requirements, it would be possible to further broaden or narrow the range of tools that can be evaluated. This would largely depend on the needs and capabilities of potential end users. For example, it may be an important prerequisite for tools to be open source so that the underlying methodology can be modified. Furthermore, a certain level of technical support may be necessary for studies that rely on community participation. While the exact composition of screening criteria is open to modification, those used in the current study represent a strong basis for selecting 'off the shelf' assessment tools that can be readily employed without proprietary restrictions across relevant urban scales to quantify multiple ES through the input of user generated data.
While there is a clear shift in the academic and research communities away from individual indicators that measure single ES towards 'off the shelf' tools that measure multiple ES at a time [16], there continues to be a lack of information on the existence, capabilities and requirements of these tools [6]. This study is the first attempt to comparatively score and rank a selection of ES assessment tools in a systematic fashion that takes into account two separate but equally relevant perspectives; scientific validity in incorporating ecosystem health alongside the measurement of multiple ES, and practical requirements that reflect the feasibility of applying each tool in situ.

Criteria		InVEST	ARIES	SolVES	B£ST	ESTIMAP	i-Tree
Ecosystem health	*Structure*	3	2	1	1	3	3
	Integrity	3	2	1	1	3	2
ES	*Completeness*	1	1	2	2	1	0
	Uncertainty	0	3	0	3	0	2
	Economic analysis	3	3	0	3	0	3
Feasibility	*Capacity*	1	1	2	3	1	3
	Data	1	2	2	1	2	1
	Scalability	2	2	2	2	2	3
	Adaptability	3	3	3	3	3	3
Total score (max = 27)		**17**	**19**	**13**	**19**	**15**	**20**

Score breakdown	Types of sources
3 = Desirable (excellent)	Peer-reviewed scientific papers that independently reviewed or applied a tool
2 = Acceptable (fair)	Third party grey literature that provide summaries and descriptions for a tool
1 = Undesirable (poor)	Peer-reviewed scientific papers authored by a tool's own developers
0 = Absent (not applicable)	Expert judgement based on a review of primary documents for a tool

Fig. 3. Scoring matrix. (a) top: final cumulative score for each assessment tool; (b) bottom: description of scoring methodology and color-coding.

Although there have been several useful reviews of standardized 'off the shelf' assessment tools in the literature, certain limitations of those previous reviews include: being limited in scope in the number [16] and type [14] of tools reviewed, being purely qualitative in nature [6], and focusing only on general trends in data sources [17], ES measured [2] or frequency of particular indicator types [18]. The screening and scoring methodology presented in this study is designed to overcome these previous limitations and in the process, facilitate the comparison and selection of relevant ES assessment tools for a given set of circumstances.

Beyond identifying i-Tree Eco as the most suitable tool within the parameters of the current study, the results of the scoring matrix (Fig. 3) also demonstrate the unequal representation of each tool across the scientific and grey literature. While there is some merit in weighing the individual scores according to source type so that independent, peer-reviewed scientific analyses are more highly valued, the disparity in sources would skew the results and prevent recently developed tools, for which third party reviews are rare, from outscoring more established tools. Instead, an additional criterion such as peer-reviewed scientific backing could be inserted to reflect a preference for objective analyses of the tools under evaluation.

The narrow edge that i-Tree Eco received over other tools implies that the results of the scoring matrix are not as definitive as one would prefer when deciding on a tool to use. The use of different criteria, or a different interpretation of the primary documents for each tool, could easily result in another tool achieving a higher score. Thus it is important that the source and reasoning (Tab. 3) behind each assigned score is presented as a reference for future comparative scoring exercises. Some of these explanations are based on the analysis and opinion of third party reviewers, whereas others are derived from interpretation of each tool's primary documents. Any attempt to score multiple tools across different types of literature sources will always involve some level of subjectivity. However, under the current circumstances and through the use of independent reviews where possible, the results of the scoring matrix are considered to be a reliable enough reflection of the performance of each tool according to the chosen evaluation criteria.

Future applied research in this field could include the use of additional methods for selecting appropriate evaluation criteria, whereby the scoring matrix is modified to accommodate different needs of end users. One possible option involves the use of surveys among end users to rank criteria. This could result in a straightforward process where the criteria with the highest votes are selected for insertion into the scoring matrix. Another approach could involve the use of multi-criteria analysis to rank criteria against each other, also through the use of surveys. Each of these approaches would better reflect end user priorities yet would require additional time and resources for collecting and analyzing survey results.

One of the limitations of the current evaluation approach is the lack of a "comparative concurrent application of multiple tools to a common location" as a way of measuring tool feasibility under practical conditions [6]. By simultaneously applying several tools to a common case study, feasibility criteria in the scoring matrix (i.e. capacity, data, adaptability) could be more accurately evaluated and compared, especially when contrasted with the current scoring method which is based on literature review. This type of practical assessment is an encouraging prospect for future research in evaluating these tools, however the significant amount of data and time that would be required to undertake such an assessment would necessitate a large team and coordination across each tool's application. An immediate next step to build on the current study could therefore involve a limited practical assessment carried out with two or three of the highest scoring tools (i-Tree, B£ST, ARIES) to further validate the feasibility scores that each of them received in the scoring matrix.

Conclusion

This study sets up a useful methodology to comparatively evaluate and rank a selection of open-access, readily available, 'off the shelf' ES assessment tools designed to quantify multiple ES across relevant urban scales. Screening criteria can help practitioners quickly narrow down suitable options from a wide range of existing ES assessment tools. The scoring matrix highlights the performance of each tool across several criteria, thus allowing for their direct, quantitative comparison in terms of scientific validity and feasibility. At the same time, every aspect of the methodology is flexible enough to accommodate specific project needs, which can be expressed through chosen criteria. The practical application of one or several of these tools on an urban NBS case study is a logical next step for subsequent research, which can validate (or update) the results of the scoring matrix while revealing additional strengths and weaknesses for each tool.

By demonstrating the value that NBS can have on human well-being in cities, through tools that measure and quantify urban ES, cities can increasingly incorporate ecosystems and biodiversity into spatial planning. Bringing nature back into our cities can restore our collective physical and mental health, improve our resiliency to meet future challenges, and actively strengthen the living ecosystems that all species depend on for survival. This study supports this transition to a more nature-inclusive urban landscape by facilitating the selection and use of existing ES assessment tools to capture the value of urban NBS.

REFERENCES

[1] Davies, C., & Lafortezza, R. (2019). Transitional path to the adoption of nature-based solutions. Land Use Policy, 80, 406–409.

[2] Haase, D., Larondelle, N., Andersson, E., Artmann, M., Borgström, S., Breuste, J., … Elmqvist, T. (2014). A Quantitative Review of Urban Ecosystem Service Assessments: Concepts, Models, and Implementation. AMBIO, 43(4), 413–433.

[3] Boumans, R., Roman, J., Altman, I., & Kaufman, L. (2015). The Multiscale Integrated Model of Ecosystem Services (MIMES): Simulating the interactions of coupled human and natural systems. Ecosystem Services, 12, 30–41.

[4] Bos, E., & Vogelzang, T. (2018, March). Groei versus groen drie casestudy's over de waarde van het stadsgroen in Amsterdam. Retrieved from: https://www.wur.nl/nl/nieuws/Stadsgroen-Amsterdam-mogelijk-meer-waard-dan-woningbouw.htm.

[5] Vigerstol, K. L., & Aukema, J. E. (2011). A comparison of tools for modeling freshwater ecosystem services. Journal of Environmental Management, 92(10), 2403–2409.

[6] Bagstad, K. J., Semmens, D. J., Waage, S., & Winthrop, R. (2013). A comparative assessment of decision-support tools for ecosystem services quantification and valuation. Ecosystem Services, 5, 27–39.

[7] Nemec, K. T., & Raudsepp-Hearne, C. (2013). The use of geographic information systems to map and assess ecosystem services. Biodiversity and Conservation, 22(1), 1–15.

[8] Gómez-Baggethun, E., & Barton, D. N. (2013). Classifying and valuing ecosystem services for urban planning. Ecological Economics, 86, 235–245.

[9] Veerkamp, C., Hanson, H., Almassy, D., Bockarjova, M., Botzen, W. J. W., Dammers, E., … Hedlund, K. (2018). Working Paper on NBS assessments: Review of current methods to assess the multiple benefits and values of urban Nature-Based Solutions (Working Paper No. Deliverable 3.3; p. 157). Naturvation.

[10] Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, Dr., … Shaw, Mr. (2009). Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. Frontiers in Ecology and the Environment, 7(1), 4–11.

[11] van Oudenhoven, A. P. E., Schröter, M., Drakou, E. G., Geijzendorffer, I. R., Jacobs, S., van Bodegom, P. M., … Albert, C. (2018). Key criteria for developing ecosystem service indicators to inform decision making. Ecological Indicators, 95, 417–426.

[12] Feld, C. K., Silva, P. M. da, Sousa, J. P., Bello, F. D., Bugter, R., Grandin, U., … Harrison, P. (2009). Indicators of biodiversity and ecosystem services: A synthesis across ecosystems and spatial scales. Oikos, 118(12), 1862–1871.

[13] Cord, A. F., Bartkowski, B., Beckmann, M., Dittrich, A., Hermans-Neumann, K., Kaim, A., … Volk, M. (2017). Towards systematic analyses of ecosystem service trade-offs and synergies: Main concepts, methods and the road ahead. Ecosystem Services, 28, 264–272.

[14] Crossman, N. D., Burkhard, B., Nedkov, S., Willemen, L., Petz, K., Palomo, I., … Maes, J. (2013). A blueprint for mapping and modelling ecosystem services. Ecosystem Services, 4, 4–14.

[15] Dammers, E., Veerkamp, C., Ruijs, A., Hedlund, K., Olsson, P., Alsterberg, C., … Bulkeley, H. (2019). Working paper: Set up, applicability and use of the Naturvation Index (Working Paper No. Deliverable 3.4; p. 136). Naturvation.

[16] Nelson, E. J., & Daily, G. C. (2010). Modelling ecosystem services in terrestrial systems. F1000 Biology Reports, 2.

[17] Martínez-Harms, M. J., & Balvanera, P. (2012). Methods for mapping ecosystem service supply: A review. International Journal of Biodiversity Science, Ecosystem Services & Management, 8(1–2), 17–25.

[18] Egoh, B., Drakou, E. G., Willemen, L., Maes, J., Dunbar, M. B., European Commission, … Institute for Environment and Sustainability. (2012). Indicators for mapping ecosystem services a review. Luxembourg: Publications Office.

[19] Potschin-Young, M., Haines-Young, R., Görg, C., Heink, U., Jax, K., & Schleyer, C. (2018). Understanding the role of conceptual frameworks: Reading the ecosystem service cascade. Ecosystem Services, 29, 428–440.