Independent Geographical Study

Evaluating the algorithms used in site suitability analysis: an open source implementation for photovoltaic solar farms in Arizona

Yik Ching Tsui (1931393)

2022

This dissertation is submitted as an Independent Geographical Study as a part of a BA in Geography at King's College London.

KING’S COLLEGE LONDON

UNIVERSITY OF LONDON

DEPARTMENT OF GEOGRAPHY

INDEPENDENT GEOGRAPHICAL STUDY

I, ……………………..Yik Ching Tsui ………………………… Hereby declare (a) that this dissertation is my own original work and that all source material used is acknowledged therein; (b) that it has been specially prepared for a degree of King’s College London; and (c) that it does not contain any material that has been or will be submitted to the Examiners of this or any other university, or any material that has been or will be submitted for any other examination.

This Dissertation is ……………8756………….. words.

Signed………Tsui Yik Ching …………………………………

Date…………7 April 2022……………………………………

Contents

1 Introduction

In the context of global anthropogenic climate change, renewable energy sources are as important as ever. With the Paris Agreement limiting the warming to below 1.5°C, there is a crucial and urgent need to replace carbon-heavy energy sources with clean energy. The drive for increased solar energy capacity is increasing the demand for improved spatial analysis for suitable solar farm sites. Site suitability analysis is important not just for climate change in the abstract, but also finds practical use for all stakeholders, including the energy industry, the government, and the people near solar farms.

The practice of site selection tends to be closed to internal firms, or discretely done by government employees or policymakers with insufficient public input. Public consultations are not enough; change can start in the very first steps of planning.

The first issue is a practical issue, and the solution would be a practical solution. Site suitability analysis needs to be held with a high standard of transparency and reproducibility. Achieving this and setting an example is the first aim of this IGS – to make a Geographical Information System that is transparent and reproducible so that analysts and readers are confident of the results and aware of strengths and limitations.

One of the key contributions is making the code open source, which furthers the goal of transparency because everyone can audit the code and reproduce a given analysis. It crowdsources bug fixes and feature improvements. Finally, it allows adjacent fields and industries to modify it and apply it elsewhere. Being open source will democratize geographic knowledge, enhance public cooperation and confidence, and lower the barriers to entry for planning and spatial analysis.It is flexible and applicable to other places and topics. It is not simply uploading a messy code that is very specific to the situation, but it handles things in a more abstract way so that it can be applied to other topics. Most importantly, it empowers more people to do their own analysis.

The second aim is to critically evaluate the inputs and parameters used in site suitability analysis. There is a need to synthesize knowledge in the field so future work can take advantage of new developments in buffers and continuous standardization functions.

To summarize, the aims of this IGS are to:

1. Provide a transparent, reproducible, open source, and flexible implementation of site suitability analysis
2. Evaluate different parameters used in site suitability analysis

2 Literature review

This section first provides an overview of the usual methodology for site suitability analysis. Next, it goes in-depth into the evolution of the field and the issues still present in the literature. It identifies the debates regarding standardization of the criteria, namely the use of buffers versus continuous functions. Next, it critically evaluates the weighting and standardization steps, namely the lack of transparency in the justification of the values chosen, explaining transparency and reproducibility as a central aim of this IGS. It situates the IGS as an original project that takes a more open perspective to knowledge production.

The methodology for site suitability and selection analysis are broadly similar across numerous studies around the world (See Sanchez-Lozano et al. (2014), Arán Carrión et al. (2008), Al-Yahyai et al. (2012), Baban et al. (2001), Sun et al. (2013), Domínguez Bravo et al. (2007)). The criteria that is important to various stakeholders are selected. Criteria are based on the region in question, regulations, and the actual data that is available. There are two types of criteria, constraints and factors. Constraints are binary criteria that is either satisfied or not. For example, a site cannot be built in a military base by commercial firms, so all pixels representing said area must not be considered at all. They are usually excluded from the study area before the rest of the analysis. The factors are non-binary criteria whose suitability depends on the exact values of the layer. These raster layers have different units, so to make them comparable they are standardized into a certain range, usually from 0 to 1, so that a pixel with a value of 0 means it is completely unsuitable, and a value of 1 means it is suitable. A variety of methods are used for standardization, and this is explored later as it is a central focus of this IGS. The raster layers numerically represent not just individual suitability of a factor, but also with the correct polarity or direction. For example, a site might benefit by being closer to transmission lines but further away from residential areas. The raster layers are then weighted according to their importance, and then summed or multiplied to form a final suitability map.

Research has evolved throughout time, improving on each other by using more accurate methodologies. However, there remains unresolved problems in existing studies. First, the standardization of raster layers into suitability scores used discrete classification, which was insufficiently rigorous, leading to the development of continuous functions based on fuzzy values.

Second, there is an issue in the lack of transparency. The weighting of the criteria is not transparent enough because the evidence and reasoning that was used to form the weights were not published. The selection of parameters used for the constraint buffers, and discrete classification or standardization functions, also did not have sufficiently clear justification. Transparency is therefore the first aim of this IGS, in conjunction with developing visualizations that demonstrates that an open source implementation that is clear and transparent is possible.

2.1 Discrete classification to continuous standardization

A vast majority of studies used a simple threshold based buffer for constraints, but this is vulnerable to uncertainties in the data collected, debates on the exact thresholds, and does not properly reflect the continuous nature of the world. For example, most studies excluded slopes of greater than 3% to 5%, but a slope of 6% might be suitable as well (and certainly more suitable than slopes of 10%). There was a noticeable lack of discussion and evaluation on this, as well as vagueness. Instead of using hard borders as buffers, Asakereh et al. (2017) used a continuous piecewise linear function of distance and even experimented with a Gaussian function. This blurs the distinction between constraints and factors because traditional buffers are essentially factors but with a binary function rather than a continuous function. Using a piecewise function moves the debate to where should the intervals be, and what the easing function should be (for example a linear or a sigmoidal function). This IGS would follow the latest developments in literature and use continuous functions. To mitigate concerns of complexity, a range function is also used, as well as any arbitrary expression for discrete standardization functions.

Continuous functions has also replaced standardization of factors. Broadly speaking, standardization is a transformation from the raw values into dimensionless quantities that can be compared. To standardize the factors, many studies classified values into an interval scale and assigned discrete scores (e.g. Ali et al. (2019)). However, like constraint buffers, it loses information and does not reflect the continuous nature of the world. Studies like Watson et al. (2015) have instead standardized factors by their range to give a continuous linear function, resembling Equation 1. Mondino et al. (2015) used a reciprocal function for average rainfall. Suh et al. (2016) used a monotonic S-shaped sigmoid function, determined by Equation 2. They still recognized that factors such as temperature and distance to transport links are better modelled with a continuous linear function.

$$range\ x = \frac{x - min(x)}{max(x) - min(x)} \tag{1}$$

Equation 1: Range function, where x is the value of a raster cell. A reverse range function is 1 subtracted by this expression

$$sigmoid\ m\ s\ x = \frac{1}{1 + (\frac{x}{m})^s} \tag{2}$$

Equation 2: Sigmoidal function. Note that the spread s should be negative for higher-is-better relationships (an increasing function)

The Gaussian function is used by Asakereh et al. to standardize solar radiation, but they did not publish the parameters used and only gave a figure showing a monotonically increasing function. Assuming they used Equation 3, the standard equation of the Gaussian function (Athanasiou et al., 2017), the parameters could be calculated based on the figure: a solar radiation of 6 kWh/m^2/day received a score of 1, zero radiation received a score of 0, and 4 kWh/m^2/day was given a score of 0.5. a is the height of the peak (the maximum suitability score), which is equal to 1. b is the x-coordinate of the peak, which is equal to 6. c can then be calculated by substitution of the midpoint $(4, 0.5)$. For the end user, it would be easier to ask for the x-coordinate of the midpoint instead of the standard deviation, which not as intuitive. Therefore, the calculation of c is handled internally. The parameters of the Gaussian function for end users is in the form of Equation 4.

$$gaussian\ a\ b\ c\ x = a \times \exp\left(-\frac{(x-b)^2}{2c^2}\right) \tag{3}$$

Equation 3: Gaussian function, where a is the height of the peak, b is the x-coordinate of the maximum, and c is the standard deviation. x is the amount of solar radiation (kWh/m^2/day) and the output is the standardized score for solar radiation.

$$gaussian\ b\ m\ x = \exp\left(\frac{(\ln 0.5)(x-b)^2}{(m-b)^2}\right) \tag{4}$$

Equation 4: Gaussian function, where b is the x-coordinate of the maximum and m is the x-coordinate of the midpoint

$$linearL\ min\ max\ x = \begin{cases} 0 & \text{if } x < min \\ 1 & \text{if } x > max \\ \frac{x-min}{max-min} & \text{if } min \leq x \leq max \end{cases} \tag{5}$$

Equation 5: Increasing clamped linear function (larger is better)

$$linearS\ min\ max\ x = \begin{cases} 1 & \text{if } x < min \\ 0 & \text{if } x > max \\ \frac{-x+max}{max-min} & \text{if } min \leq x \leq max \end{cases} \tag{6}$$

Equation 6: Decreasing clamped linear function (smaller is better)

This IGS will continue from the latest developments and use continuous functions to standardize factors, either by their range, a Gaussian, or sigmoid as appropriate. These functions are visualized in Figure 1. The limitations are again complexity to the end user, so range standardization should always be available as a last resort. Equations 5 and 6 show the linear functions used by Asakereh and Suh, which is a linear interpolation between two values with clamps. For example, in the increasing case, anything larger than max is assigned 1, anything less than min is assigned 0, and anything in between is a linear interpolation between max and min that maps to the range $[0, 1]$ on the domain $[min, max]$.

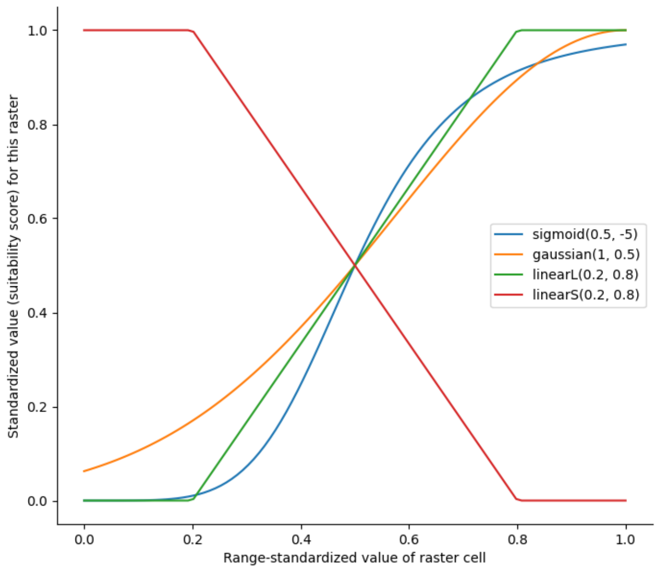

Figure 1: A graph of the standardization functions used

2.2 Polarity

The literature disagrees on whether some criteria should be a positive or negative factor. On one hand, authors like Al Garni et al. (2017) and Merrouni et al. (2018) viewed that solar farms should be closer to residential areas because shorter distance to travel for power lines means lower costs. On the other hand, studies like Wiguna et al. (2016) thought that the proximity to residential areas will have negative environmental and visual impacts. Even Al Garni et al. recognized that looking for sites further away was a method to avoid not-in-my-backyard (NIMBY) opposition. Ali et al. found that the locals around an existing solar farm were affected by noise pollution, visual disturbance and moonlight reflection at night-time. Recognizing this contradiction, Giamalaki et al. (2019) used the VIEWSHED tool in ArcGIS to classify potential sites based on their visibility to various important areas, while still being able to be close to residential areas. Considering that the primary sources collected on the field indicates that the negatives outweigh the positives, proximity to residential areas will be treated as a negative, but existing solar farms should be compared to see if they match the research.

2.3 Weighting of criteria

Studies have relied on experts to determine the weighting of criteria. Because multiple responses are collected, a multi-criteria decision making (MCDM) technique is used to combine the different answers into one set of weights. The most common MCDM technique is AHP (Al Garni et al., 2017). First developed by Thomas Saaty in 1980, AHP involves a pairwise matrix to weight each criteria against each other and calculate an overall set of criteria weights. Geographers has applied this pairwise matrix to calculate weights for site suitability factors.

As a relatively early study, Uyan's 2013 study did not mention how the pairwise matrix was completed. Many studies, including some disappointingly recent, credited "experts" for their pairwise matrix but did not elaborate further (Tahri et al., 2015) (Asakereh et al., 2017) (Shorabeh et al., 2019). Later studies began to offer more transparency, discussion, and evaluation. While Merrouni et al. also invited local experts with knowledge of the area, there was no further elaboration or development until Ali et al. based their unique study on regional experts and public opinion. Instead of relying purely on literature, they also visited an existing solar farm in the study area and collected public opinion about disturbances the farm caused and then based their constraints on that information. The experts were specifically regional experts, with the ability to give advise specific to the area, such as a larger buffer around forests because of past flooding in the study area.

Watson et al. (2015) also provided more detail about their process. The experts were identified to be consultants, project and technology managers working in the field, justified with a standard based on Xiang et al. (1994). The expert's opinions were analyzed and critically evaluated against literature. Likewise, Suh et al. (2016) invited ten experts working as solar PV professionals and analysts. Giamalaki et al. used a formal survey for local experts but also included stakeholders including policymakers, power suppliers, and environmental groups.

The increasing details are good because of better transparency and reproducibility. Yet, studies have inadequately evaluated the use of expert opinion. Different stakeholders are ultimately going to have different priorities, and combining their opinions into one set of weights is dubious. A consistency ratio has been used to ensure the final weights are consistent, but through my personal experimentation, it is difficult for a single person to give a consistent response, let alone responses from a wider group of people. Every study claimed that their matrix had a good consistency ratio, so there was little consideration of what should happen if the consistency ratio was too high. Tahri et al. says the matrix would be “rejected and revised”, and Merrouni et al. says they have to be “recalculated”, but it was unclear what calculations were done, and if it was needed. Were the experts asked to modify their answers? Were the numbers processed without the expert’s knowledge? Giamalaki et al. explicitly said they would use the Maximum Deviation Approach from Gastes et al. (2012), but no study reported that they had to fix the matrix to reduce the consistency ratio. Furthermore, having a plan to reduce the consistency ratio misses the point. There should be a discussion on the implications of adjusting the answers, because there would be an underlying reason why the answers were inconsistent. Adjusting it admits that the respondents do not have a consensus, or are divided about the priorities, which should be addressed in the paper rather than numerically bypassed or resolved during interviews beforehand.

2.4 Standardization parameters

The parameters of site suitability analysis refers to the exact method used in constraint buffers and factor standardization. For example, the choice of a 1 kilometre buffer away from protected areas. Many studies justify their choices with “based on previous studies”, but either does not elaborate further and did not explain how they were “considered” (Aydin et al., Asakereh et al.), or did not justify their own choices and explain the methodology they used to decide, like Watson et al. Ali et al. admits that there was “no specific rule” in deciding the distances, citing the variation in previous studies, and ended up contributing another number with no justification.

Studies like Giamalaki et al. indicated a better attempt at increasing inclusiveness during the decision making process, but did not make clear what conclusion was reached. They hosted a workshop that included policymakers, power supplies, and environmental groups, but it was unknown what conclusions were reached, if any, nor the evidence and reasoning that was used to arrive at those conclusions. It is not known if a consensus could be found or if different stakeholders had irreconcilable differences. Interviews, workshops, and focus groups appeared to give the community a new voice independent of the technical engineer, but if the results are not publicly available, then it was just another private event with no transparency.

Even when studies did explain the factors behind increasing or decreasing a value, it was unclear why a specific value was chosen. The experts in Suh et al. considered the island's lack of flat land and the relative unimportance of roads into forming the weights, but it was still unclear how they gave exact value for the weights and their reasoning. For example, if they had to raise the tolerance for steeper slopes, how did they decide how steep would a site be allowed to be in? If they used research showing that solar farms are uneconomical or mechanically dangerous beyond a certain point, they should cite it for understanding. Ali et al. involved an interview with experts, who suggested that the buffer to wetland should be 400m because of previous flooding in the area. This is a welcome advance but still falls short, because why exactly was 400m chosen? Did it match the extent of inland flooding? There was still insufficient context to that number.

Some studies cited government guidance or regulations (Ali et al.). Taking recommendations from the government could be paradoxical because governments may use academic literature to base their recommendations on. There is still a small feedback loop for academic literature, but advances in the field should propagate faster than government regulations. Furthermore, regulations are based on certain studies or reports in the past, so following government regulations is indirectly following the evidence that formed the basis of the regulation. It would be better to directly cite the underlying evidence and forgo the indirection, because it could be outdated and newer research could provide more accurate analysis and more critical evaluation. It is not hard to justify the choices. Asakereh et al. cited the amount needed for solar panels to be economically feasible (at least 4.5) to decide their standardization function for solar radiation.

3 Methodology

The second aim of this IGS is to evaluate different parameters used in constraint buffers and standardization functions. A more straightforward approach is taken, by adapting previous studies to a single study area and compares their different methodologies. The goal is not necessarily to recommend specific sites or repeat an analysis on another region, so adapting previous studies ensures that there is a realistic basis for the criteria choice, while not sacrificing mathematical rigour. This section first introduces the study area of Arizona as a case study, and the two studies to be adapted. The datasets used, the weights, and the standardization functions are described and justified. The analysis techniques that will be used to evaluate the analysis algorithms as per the two other aims are also previewed.

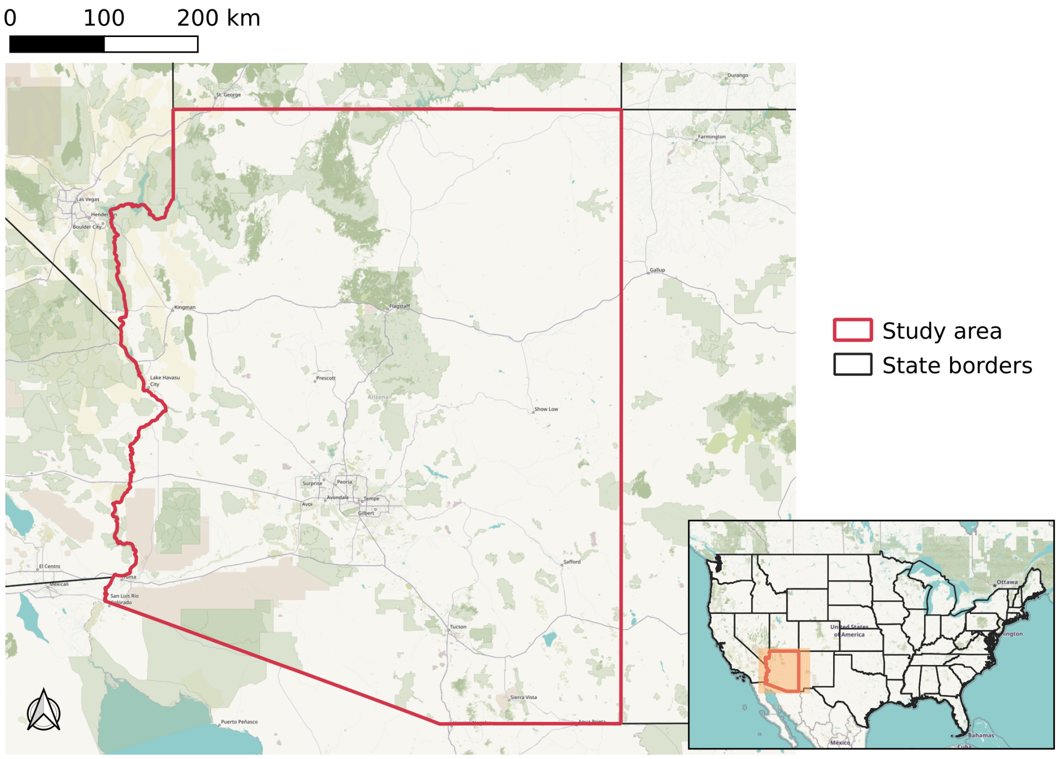

Figure 2: Study area

Figure 2 shows the study area, which includes the entire state of Arizona in the United States. Arizona is not particularly a special place, but it already has high solar energy production due to its south-west location with high solar radiation, so it has a solid foundation to expand upon because existing solar farms are used to compare the analysis with. The aims of this IGS is location-agnostic, so analysis should not be unnecessarily dependant on the particular situations of the study area.

Two studies are chosen based on the diversity of their methods, the availability of data, and the overall quality of the study. The major determinant for quality is the use of a continuous standardization function, and only a few did so. The number

Criteria	Type	Raw or calculated	Resolution	Source
Insolation	Raster	Raw	1 km	Fick et al. (2020)
Temperature	Raster	Raw	1 km	Fick et al. (2020)
Elevation	Raster	Raw	30 m	CGIAR (2018)
Slope	Raster	Calculated from DEM	30 m	-
Aspect	Raster	Calculated from DEM	30 m	-
Residential	Vector polygon	Calculated from land use	30 m	USGS (2019)
Protected	Vector polygon	Raw	-	UNEP-WCMC et al. (2021)
Roads	Vector line	Raw	-	OpenStreetMap (2021)
Power lines	Vector line	Raw	-	HIFLD (2021)

Table 1: Metadata of the criteria used

of comparisons has to be limited to keep the analysis focused. The two studies are Asakereh et al. and Suh et al. Henceforth they are referred to by the first author's last name.

Table 1 shows all the data used (for both constraints and factors), their metadata, and the sources. The criteria were from those studies, but some criteria has been removed due to unavailable data. The weights, as given in Table 2, are subsequently adjusted to ensure all factors sum to 1, in a manner that reflects the similarities of the factors. The weights for the modified parameters are in Table 6 in the appendix. The insolation data is in kJ/m^2/day, but Asakereh et al. and Suh et al. gives the parameters in kWh/m^2/day. The data is thus converted into kWh/m^2/day first before standardizing. The sunshine data was not available, so Suh et al.'s weights for sunshine is added to insolation. Asakereh et al. only gave a weight for "environmental constraints", so it was divided by three to give the weights for slope, residential, and protected areas.

All layers are converted into raster data, with each cell representing either the value or distance from a feature. Vector polygons can be used as both a constraint or with proximity analysis, while vector lines can only be used with the latter. Proximity analysis is vulnerable to edge effects, therefore the Arizona state border will be buffered by 100 km first, then applying the proximity calculation, then clipping it with the actual borders.

Table 3 shows the constraints used. Asakereh et al. was vague on how roads were

Criteria	Asakereh et al.	Suh et al.
Insolation	0.539	$0.3889 + 0.2682 = 0.6571$
Temperature	-	0.0838
Slope	$\frac{0.291}{3} = 0.097$	0.0799
Residential	$\frac{0.291}{3} = 0.097$	-
Protected	$\frac{0.291}{3} = 0.097$	-
Roads	0.17	0.0641
Power	-	0.1151
Sum	1	1

Table 2: Weights of the factors for each study

Criteria	Asakereh et al.	Suh et al.
Insolation	-	-
Temperature	-	-
Slope	-	-
Elevation	-	$<$ 492m
Aspect	-	-
Residential	$>$ 0m	$>$0m
Protected	$>$ 0m	$>$0m
Roads	$>$ 100m	-
Power	-	-

Table 3: Constraints used and their threshold/distance. 0m means only the polygons are excluded and no buffering is done.

Criteria	Units	Asakereh et al.	Suh et al.
Insolation	kWh/m^2/day	Gaussian 6 4	sigmoid 2 -5
Temperature	Degrees Celsius	-	linearS 20 27.3
Slope	Angular Degrees	linearS 3 10	sigmoid 9 3
Residential	Metres	linearL 1000 5000	-
Protected	Metres	linearL 100 400	-
Roads	Metres	linearS 1000 4000	linearS 5 1600
Power	Metres	-	linearS 0 3000

Table 4: Standardization functions for each study

excluded; being a vector line it has a width of zero (even though in reality it has a width), so a buffer of 10m is used.

For the factors, Table 4 show the studies and their parameters for the standardization functions. For example, gaussian 6 4 means a Gaussian function with the peak at 6 kWh/m^2/day and midpoint at 4 kWh/m^2/day, and x represents a cell in the raster layer. The functions are applied to every cell in the layer.

Applying these studies to Arizona is still an approximation of their study because some data is not available for Arizona. Asakereh et al. had separate functions for urban and rural areas; but only the urban area function is used for residential areas. Rivers, conservation areas, shrubberies and swamps used linearL 100 400. This function is used for the protected areas layer with the assumption that protected areas include those features.

After standardization, the factors are multiplied by their weights. This is because a low value in one layer should lead to a low value in the final result. Addition implies the criteria are independent of each other, and penalises low values less severely, while multiplication recognizes that sometimes site suitability analysis is about excluding unsuitable areas more than including suitable areas. The weighted factors are multiplied and the final result range-standardized. This was then clipped by the constraints layer, if any.

Results are compared based on zonal statistics of existing sites. The suitability scores of existing sites is the main measure of model quality, not the distribution of the scores themselves. More pixels with higher scores are not necessarily better, because it leads to the extreme result of all pixels having a perfect score. More high scores in general also increases the number of potential sites to individually consider, making selecting specific sites harder. The goal is not to model existing sites, so models should not overfit them. This means that methods for spatial modelling such as OLS are not used, because this is not a modelling exercise. Furthermore, OLS would still be inappropriate because the some standardization functions like the gaussian and the sigmoid are not linear.

A sensitivity analysis will be conducted to examine the effects of different definitions of residential areas. Only a few studies attempted to do a sensitivity analysis, but all of them were about weights, which included an "all-even" scenario where all factors are weighted equally. The subsequent scenarios were just variations of increasing weights for a type of factors, excluding some types of factors, or only including some types of factors. Their primary purpose was to evaluate their result and discuss the impact of their weighting choices, as a way to improve analysis quality and critical awareness. As this IGS does not aim to repeat analysis but evaluate the existing practices in the field, the primary purpose of sensitivity analysis is to extend the existing evaluation.

4 Results of open source implementation

This section presents the results targeting the first aim of the IGS, which is to provide a transparent, reproducible, open source, and flexible implementation of site suitability analysis. It first focuses on the abstract big picture, and explains what open source really means and some general limitations regarded open source and transparency. The concrete repository is then described with a flowchart. It then explains the separation between code and configuration and how that improves transparency and reproducibility. The next section will target the second aim; it presents the analysis results and evaluates the algorithm parameters with new visualizations.

The source code ("repository") is available at Github with an open source license[1]. This means anyone can run the program, read the source code, make changes to it, re-distribute it themselves, for personal or commercial use. Free software is fun-

[1] `github.com/akazukin5151/site-suitability`

damental to ensuring transparency and reproducibility, as anyone can inspect the source code to ensure it does what it says, and can reproduce an analysis given its parameters.

It should be recognized that it is unlikely that this implementation will receive attention from commercial industry and the spatial analyst community, because is little financial incentive for stakeholders to contribute directly to the repository and no marketing is done. Furthermore, software engineering is also not the focus, so the execution speed is not fully optimized. These are real limitations on the project overall, limiting its effectiveness to businesses and policymakers, but the aims of this IGS is prioritized first. The primary focus is to be a demonstration for academics and researchers. The key contribution is to show that a more transparent process is possible, even when using more sophisticated algorithms.

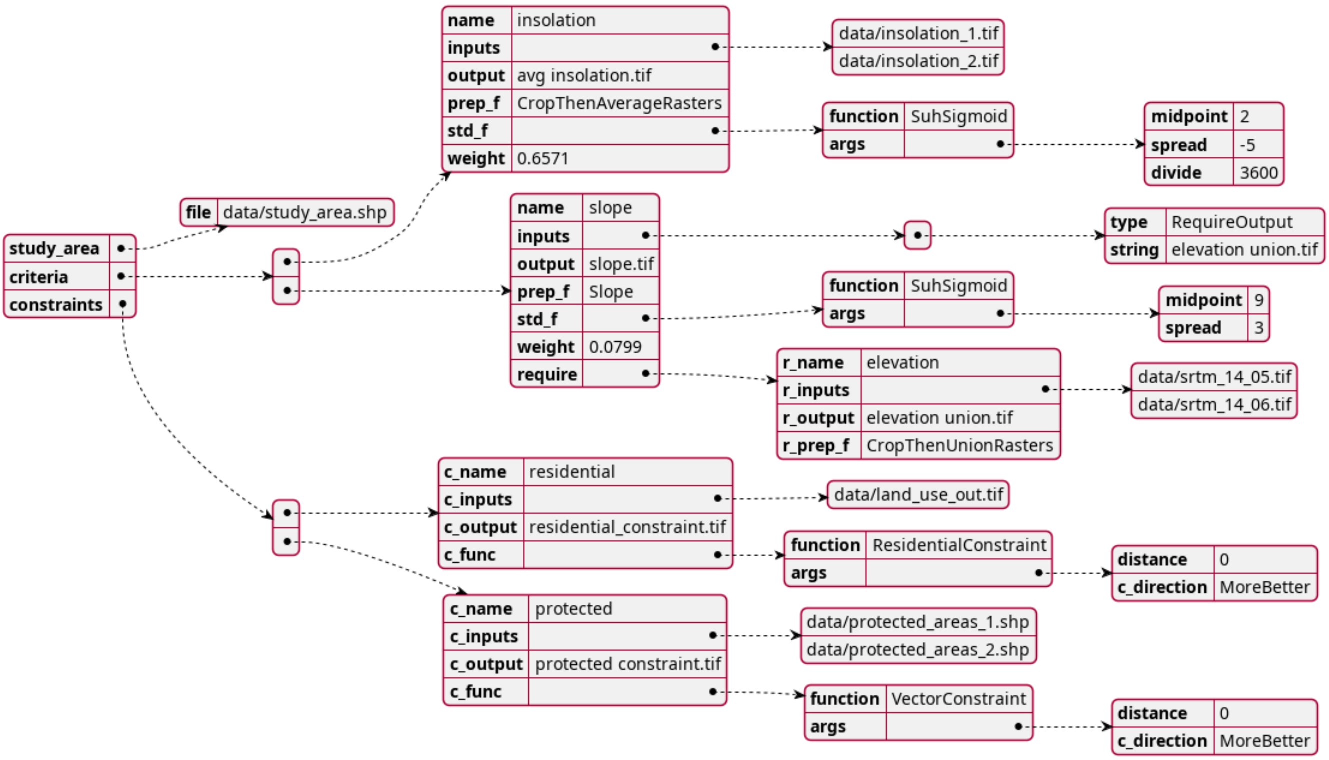

Figure 3: A diagram of an example JSON configuration file

On the aim of transparency and reproducibility, there is a fine difference between reproducibility and code auditing. On one hand, open source code is readily available for anyone to review and ensure the code does what it is supposed to do. This enhances transparency because people can verify the analysis. On the other hand, a study is truly reproduced if it was independent. Reproducing a study is not simply copying the code and running it on another machine. It is designing an independent research that answers the same question, then independently collect and analyse the data. Therefore, while open source code increases transparency, it does not necessarily increase reproducibility. This is mitigated by separating the input specification and the calculation. In contrast to loading inputs, processing them, and calculating them all in one file, the input layers are specified in a separate JSON configuration file. For visual reference, Figure 3 shows a diagram of an example file.

The configuration file is publicly documented. Every layer has to specify a standardization function and weight. The configuration file is not code and does not execute anything, so it is trivial to generate such a visualization. It only describes what inputs there are, how important they are, and how they should be standardized. The JSON format is a well known format that non technical people can directly read and understand without prior knowledge, so the separation has clear benefits in widening the audience that can benefit from this IGS. The program reads the given configuration files and run the calculations. This flexibility enables analysts to be more transparent because they can release the configuration files used without releasing the program that does the calculations. The configuration files will reveal everything that a reproducing study would need to know – what inputs were used, how they were standardized, and how important they were. The exact technique used to calculate the result from the inputs must be written independently.

This still does not address where the parameters in the standardization functions and the weights came from. The configuration file enables one to perfectly reproduce a given study, but does not shed light on why these values were chosen. As the reproducing study should have an independent collection of data, including consultation with experts, it is likely that it would decide on a slightly different set of values, yielding a different result. Indeed, this is not a software problem but a publishing problem, so it is not something that open source can be solved. The biggest contribution of this IGS is therefore not the open source implementation given but the idea and the argument that this field needs more transparency. Transparency on standardization function parameters, weights, and constraint buffers can only be obtained through publishing, and this is what researchers reading this must grasp from this IGS.

Figure 4 shows a flowchart of the entire repository. First, the configuration files

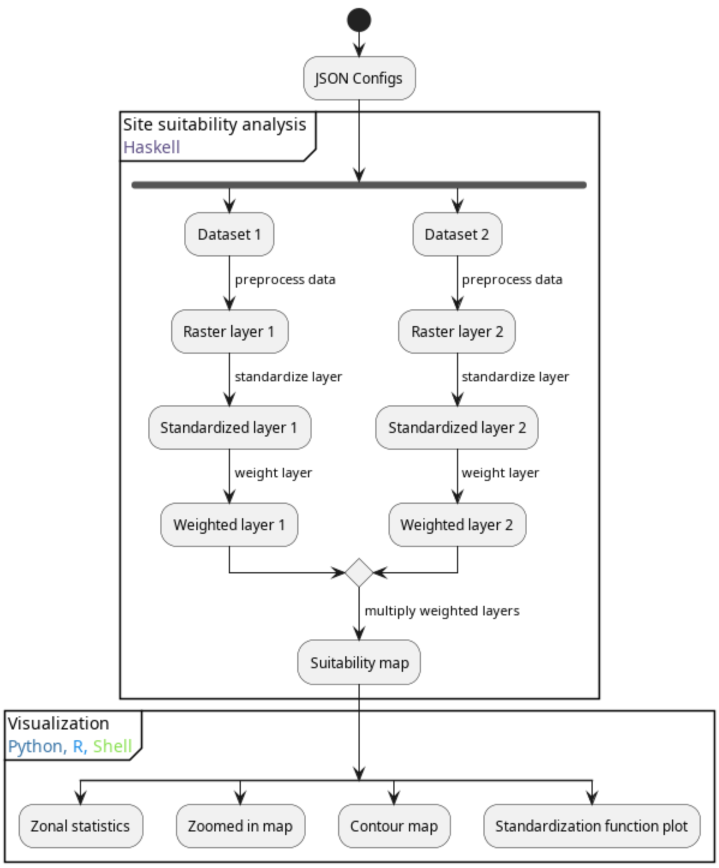

Figure 4: A flowchart explaining the structure of the entire repository

are parsed and read. For every configuration file, every dataset it describes is preprocessed, standardized, and weighted. The layers are multiplied to give the overall suitability map. The site suitability analysis stage keeps the intermediate computations, including all the standardized layers separately, so that they can be visualized and analysed individually in the visualization stage. The site suitability analysis state is automated based on the configuration files, but the visualization stage needs to be individually adjusted and ran.

The current iteration can be more user friendly, for example, using a graphical user interface (GUI) that streamlines the initial site suitability analysis with the subsequent visualizations this IGS used. The configuration file can also be represented through a GUI, broadening the amount of people who can edit and view them. Because JSON is a well known format, and the configuration by itself does not require technical knowledge, this should not be a hard extension. However, more advanced uses of standardization functions would involve editing the source code directly. There are only six built-in preprocessing functions that transforms the raw data into single raster layers. More specialized or uncommon datasets may require a composition of multiple preprocessing functions, or an entirely different pipeline. For example, the study area borders and the land use data was preprocessed with shell scripts before passing it to the main program. The limitation is mainly because there is an infinite ways in which the datasets can be organized, making it hard to be fully flexible. There are five built-in standardization functions, but the sixth allows for an arbitrary expression to be used, so even discrete classification schemes can be used. Editing the Haskell source code is not for beginners, but writing a custom expression for the standardization function also requires some basic programming knowledge.

The source code of the main program is written in Haskell, but the functionality does not depend on any Haskell features because it is essentially a wrapper around shell commands, and can be implemented in any programming language. Broadly speaking the most needed improvements is support for more general data preprocessing, but altogether the code does require some technical knowledge to modify. The subsequent visualizations are either written in R and Python or manually created using QGIS. Therefore, even though open source software lowers the barriers of entry, it does not eliminate the need for expertise in the field.

5 Analysis and Discussion

This section is organized as follows. First, the literal, near unmodified application of the two studies to Arizona is done. The maps shows that there is a need to further adjust the parameters, in particular for some maps the roads and power lines layers are too strict and narrow, which significantly constrained the overall output and does not resemble existing sites. While studies with expert opinion about the standardization functions do not necessarily need to fine tune parameters in this matter, the subsequent discussion in this IGS is still relevant for the evaluation of their results. A range function is used as a replacement standardization function for this IGS, but future research will benefit from better visualizations to increase transparency and improve justification for their methodology. Contour maps are used to illustrate the impact of different parameter choices with a concrete representation of the area. Contour maps are not a novel concept, but studies have not used this intuitive visualization to increase transparency. It demonstrates that it is not hard for future research to publish intuitive justification that might have been used by experts and stakeholders alike before the analysis. The range function used is also evaluated by plotting the function overlaid on the histogram of the actual values of the layer. Again, this is not a revolutionary idea but it is very helpful for readers to understand the implications of the chosen function and whether the expert's opinion was accurate. It again demonstrates another visualization that future research can take advantage of to increase transparency.

The process of evaluating the quality of the modified output is discussed, because it is a subjective process. Comparing to existing solar farms, if possible, is tautological if the aim is to construct a new farm. The use of distance or proximity to features also does not always fully reflect the desired qualities. Using one of the maps as an example, a site scored high because it was close to a major road, but the actual travel time by car to that major road involved a detour using minor roads. Finally, a sensitivity analysis is done by varying the definition of "residential areas".

5.1 Adaption of parameters and how contour maps can aid the process

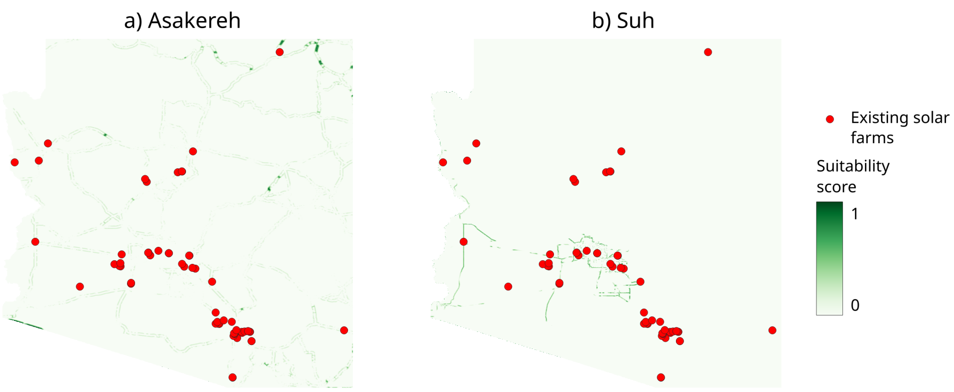

Figure 5: Site suitability maps based on the parameters from the two studies

Figure 5 shows the site suitability maps for parameters based on the two studies. Table 5 shows the zonal statistics of the suitability maps for polygons of existing solar farms. The literal application of methodology from different studies onto Arizona,

Name	Mean of sum	Mean of mean	Mean of median	Standard deviation of mean
Asakareh	0.194	0.000220	0.00000700	0.00146
Asakareh improved	4.54	0.00193	0.00136	0.0105
Suh	3.045	0.0429	0.0434	0.136
Suh improved	320	0.457	0.448	0.368

Table 5: Zonal statistics of suitability scores for polygons of existing solar farms

needs to be modified to suit the study area. Asakareh and Suh's maps have very few suitable areas, and most of the state essentially has a score of 0.

Asakereh's parameters gave a very limited places with non-zero suitability, and the few places that have high suitability bears no resemblance to existing sites. Figure 17 in the appendix is evidence that most pixels have very low values. The average suitability score of existing site polygons are 0.000220, indicating that it does not model existing sites well. The largest sites are in the corner of the state and very remote, sticking closely to roads and avoiding residential areas. The roads are visible in the form of two narrow, parallel lines. This makes sense because the residential areas layer is slightly more expansive than the roads layer: the standardization function for the former clamps at 5000 m, while the roads are clamped at 4000.

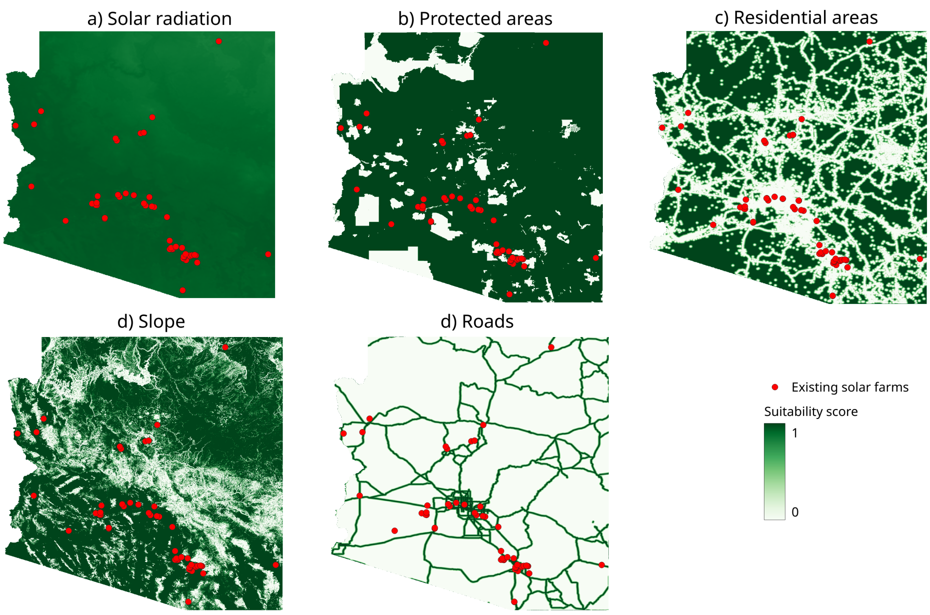

Figure 6: Maps of the standardized raster layers used in Asakereh's parameters

Specifically, places further than 4000 metres away from roads were given a score of 0. Compared to other the layers in Figure 6, it is a clear anomaly and the primary reason why the map looks sparse. This problem is due to the different study area. The area of Asakareh's study area was 64,055 km^2, while Arizona is 295,234 km^2. For simplicity, the clamped linear function would be replaced by a range function like Watson et al., which simply standardizes the values based on their state-wide range.

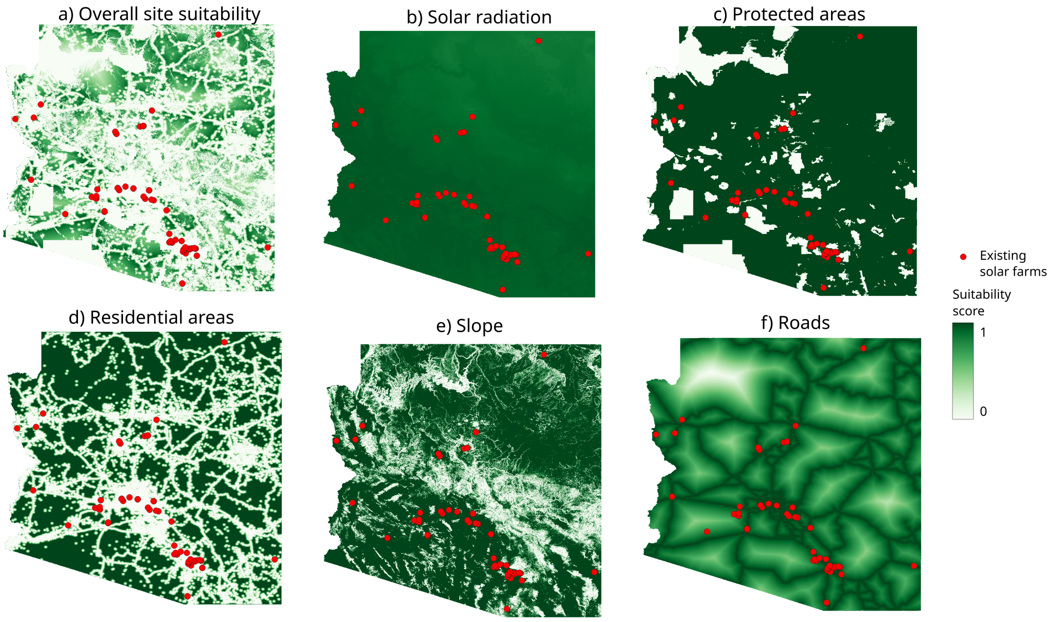

Figure 7: Site suitability and factor layers of Asakereh's parameters, using a range function for roads.

Figure 7 shows the suitability map and the factor layers, after modifying Asakereh's parameters. The area-average of suitability scores within existing site polygons are now 0.00193, an eight-fold improvement. The layer that constrained out the most places is now the residential areas layers, which shows that in reality, solar farms in Arizona were closer to residential areas than what research shows. It does not necessarily mean the literature was incorrect, but just solar farms in Arizona did not prioritise keeping a distance away from residential areas. It does not indicate that residents does not have visual pollution with solar farms being nearby.

Suh's suitability map has the highest area-average score compared to the other one, despite visually looking very limited. Figure 8 shows the suitability map in the southern part of the state, overlaid with a points layer of existing solar sites. The polygon layer is too small to been seen at that scale, so the points are used with the radius

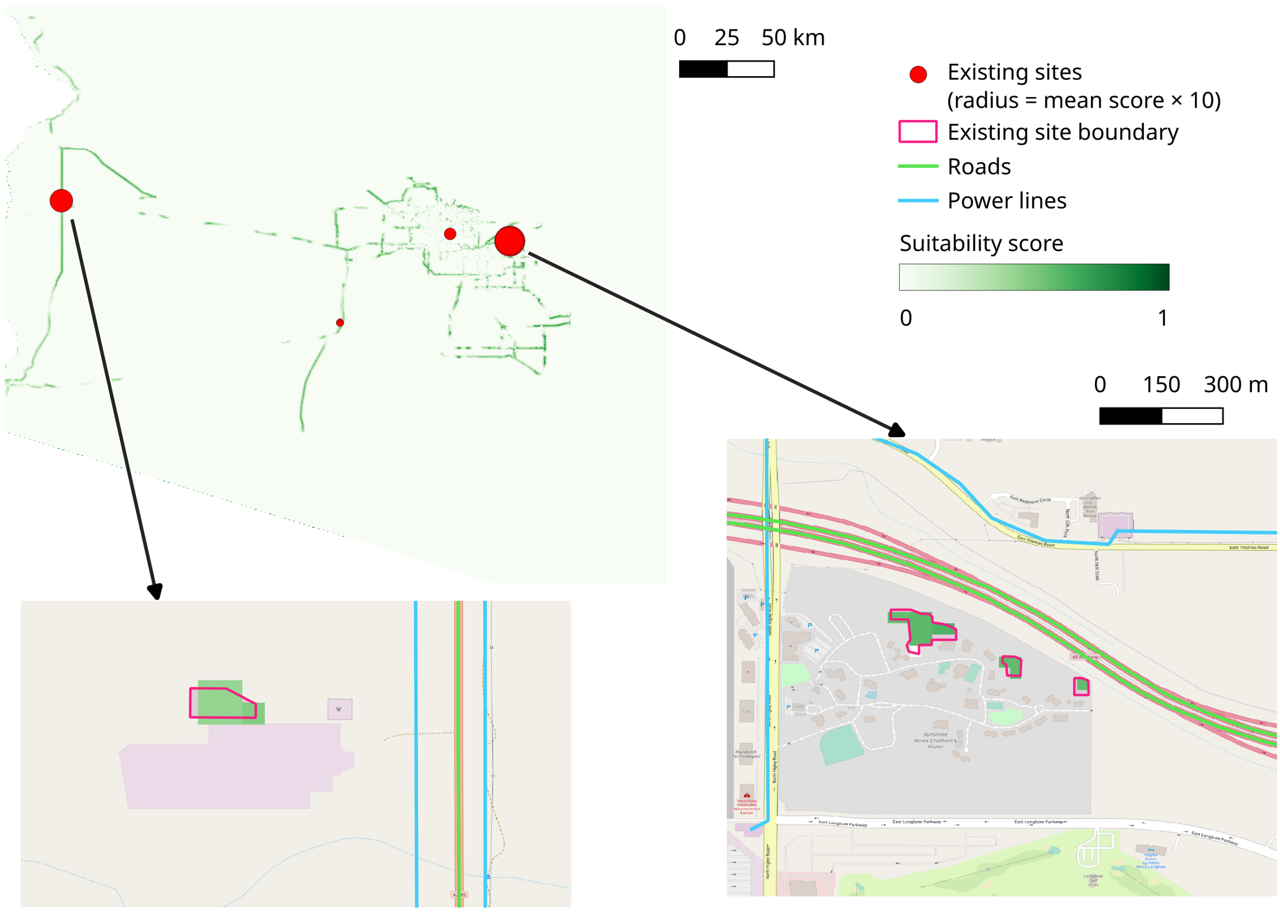

Figure 8: Detailed zoom of existing sites in Suh's map

scaled according to the average suitability score within in the site boundary. The two farms with the highest suitability scores are zoomed in. Both share the common feature of being near roads and power lines, which is the main factor boosting their suitability score. Those two factors boosting a site is unsurprising, but like Asakareh the two factors are defined very narrowly. Places further than 1600 metres away from roads are given and 3000 metres away from power lines are given a score of 0, and it is significant compared to the other layers used as shown in Figure 9. More importantly, this map reveals a limitation in using proximity from a feature to measure a factor. Factors, or more precisely qualities that stakeholders want to have in a solar farm, are descriptive and qualitative. Raster layers translate it into something quantitative and directly measurable from the real world. If stakeholders want sites to be closer to roads, they likely mean in terms of practical travel time by car to the site, and not the literal proximity to a vector line. The map on the right shows the site is close to the major road, but to actually travel there by car involves a U-shaped detour from the south. This does not seem to be recognized by literature; although some have explored the use of view-shed analysis, travel time was not considered, probably because the increase in computational intensity and research design is not worth the small increase in accuracy. However, it is still a limitation that future research should be willing to tackle.

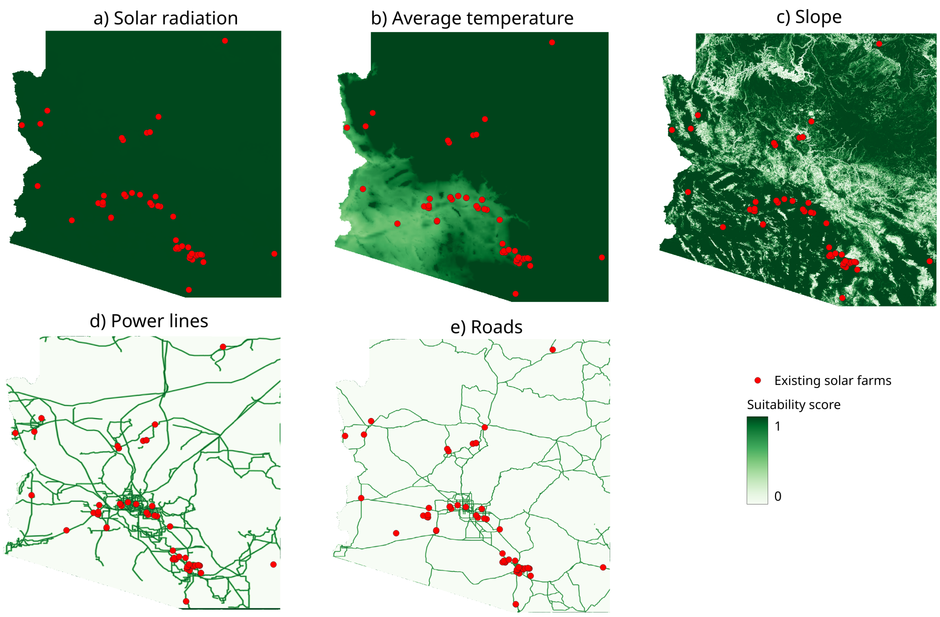

Figure 9: Maps of the standardized raster layers used in Suh's parameters

Suh's parameters are clearly only designed for their study area in mind, which was an island east of South Korea. The area of the island was only 72 km^2. A range function would be used instead, for both power lines and roads. These two factors are not especially remarkable by itself, it is only the narrow range used that gives it outsized influence in the final map.

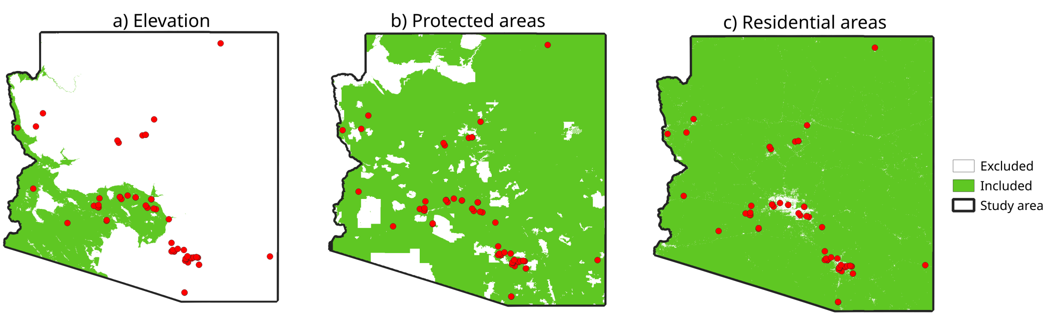

Figure 10: Maps of the constraints used with Suh's parameters

Figure 10 shows the third reason why Suh's map are mostly unsuitable: the elevation constraint excluded places higher than 492 m, due to South Korean law. Following local regulations was not a goal for this IGS, so laws and zoning restrictions are not considered, which is an important limitation. It is not hard to add or modify layers to reflect relevant regulations. Furthermore, focusing too much on a case study comes at the cost of less flexibility and applicability. Nevertheless, the elevation constraint would be removed as the primary aim is to evaluate the algorithm parameters used by site suitability analysis in general.

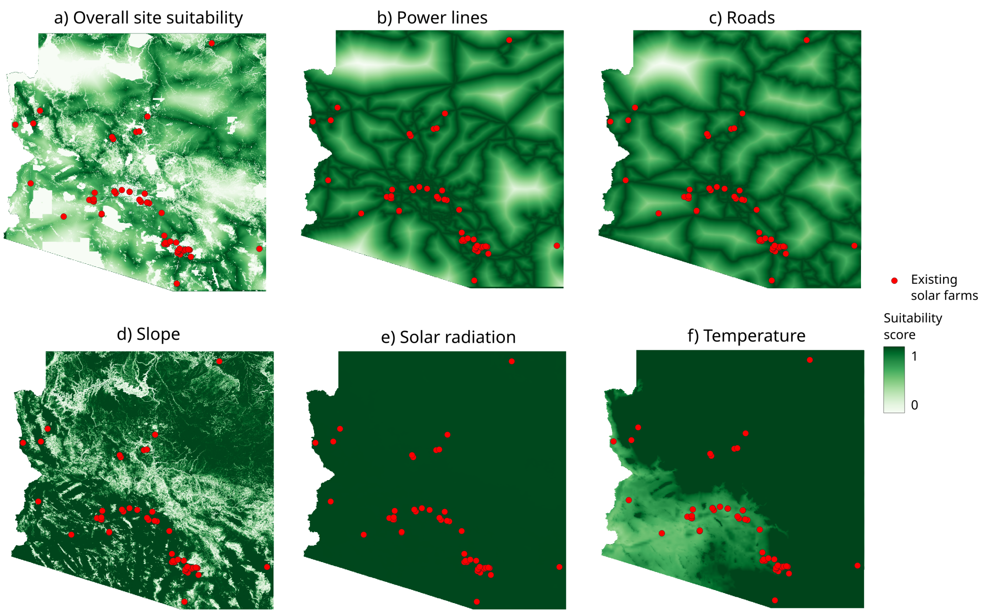

Figure 11: Maps of the constraints used with Suh's parameters, modified

Figure 11 shows the suitability map and factors, using parameters modified from Suh: removing the elevation constraint, and using a range function for roads and power lines. The mean-of-mean has improved by ten-fold to 0.457, indicating that slightly more than half of the existing solar farms got a suitability score of less than 0.5: in other words, a small majority of existing farms are only half as ideal. The improvement of the mean came at a cost of higher scores overall, meaning it gave areas without solar farms higher scores. The average sum of scores within existing sites is 320, a few magnitudes higher than the rest. Pixels with low scores do not increase the sum, so there is no penalty for low scores within sites. Instead, even low score counts towards increasing the total score for the sites. On the other hand, the mean and median measures will penalize pixels with low scores, as long as they are inside an existing solar site, because that represent places in a solar farm that was rated as unsuitable. The mean and median are therefore more appropriate measures to evaluate the models as they give some consideration to false negatives.

On the other hand, neither will penalize false positives, which is when cells outside existing sites are given high suitability scores but in reality they are not suitable. Neither would penalize a raster filled with all ones. This is a difficult problem to solve because the absence of existing sites does not necessarily indicate the site is unsuitable, but the presence of existing sites also does not necessarily indicate the site is suitable. Site suitability analysis therefore cannot be the only step in the site selection process, which should include detailed on-ground investigation. It is useful to filter out the most obvious unsuitable sites and provide indicators of where the most suitable sites might be, based on the constraints and factors layers. However, relatively few studies did go further, presumably because the focus was on site suitability analysis and not the site selection process in general, and the aim was to contribute to the technical and mathematical aspects of analysis or repeat analysis for a different area. Suh et al. used satellite imagery to explore the potential sites, finding that one of them was not suitable because it was used as the island's harbour and that another site was partially occupied by a building. While solar panels can be installed on top of buildings, other topics such as wind farms might not be able to, so different topics will have a different level of acceptability regarding on-ground investigation. This reflection concedes that the IGS is also focused on the technical and mathematical aspects of site suitability analysis, rather than the wider site selection process. Detailed analysis of the constraints and factor layer, plus the standardization functions used, are perhaps more useful than the final suitability map at the end. Therefore, this IGS still have made important contributions to the site selection process.

A range function was used primarily to avoid the need to calculate, debate, or interview experts, because the range function is entirely dependant on the data within the study area, and the aim of the IGS is to evaluate the algorithm parameters, not to repeat site suitability analysis to Arizona. However, that cannot be said for future studies that do aim to repeat site suitability analysis to new places. One major contribution of this IGS is to demonstrate the viability of contour maps in justifying parameters, evaluating the result, and providing transparency. The following analysis is on Suh's unmodified map, but the same technique will work for any other maps. Figure

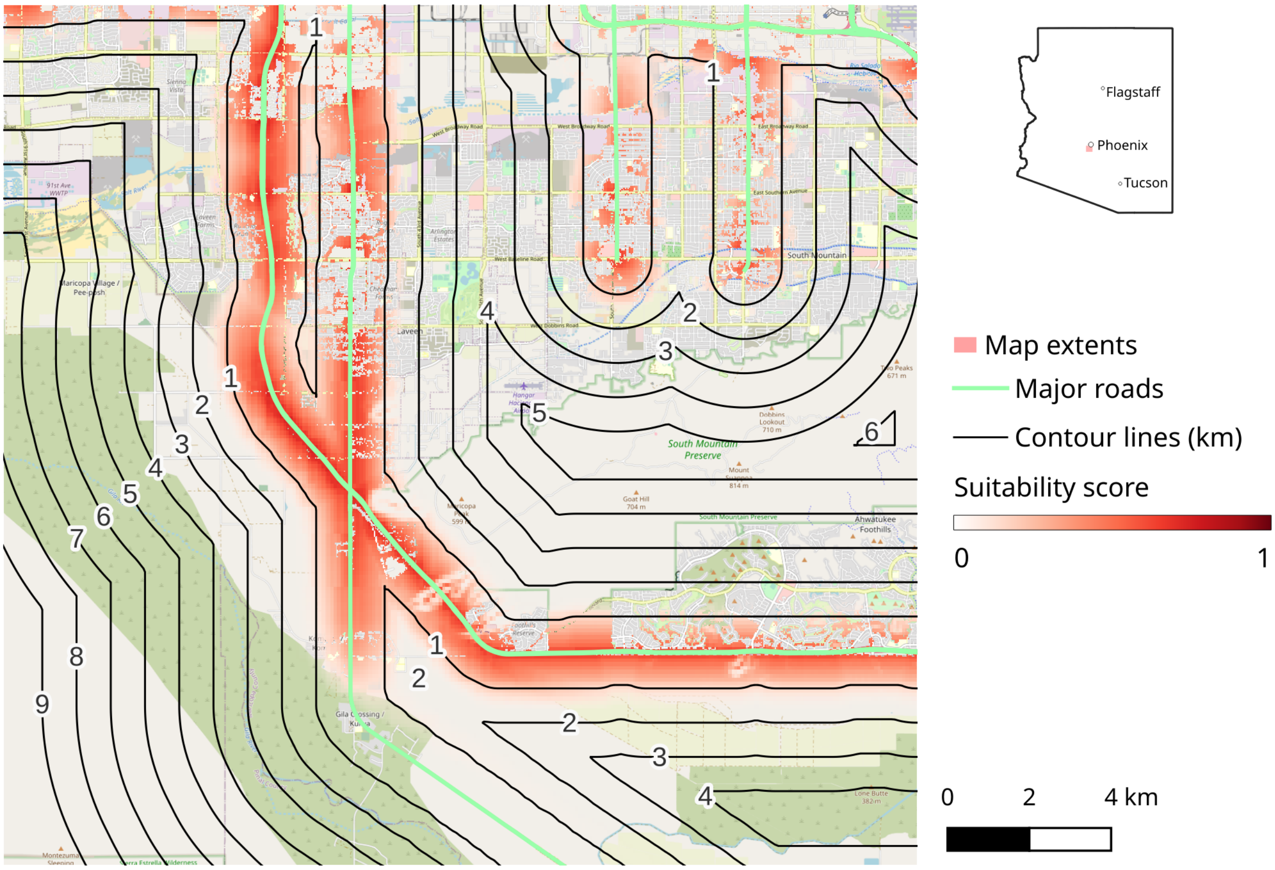

Figure 12: Map of major roads near Phoenix, and the contour lines of proximity to roads, overlaid on the suitability map using Suh's parameters. Zero suitability is not shown

12 shows the contour map for roads near Phoenix, Arizona. The black lines surrounding the roads (in green) are the contour lines. Points on a contour line are always a certain distance from any major road. It confirms that all suitable areas are within 1600 metres from roads, as specified in the configuration. The background base map enables users to see exactly what would happen if a different parameter was used. For example, if it was extended to 2 km, exactly what places would it now include? Alternatively, stakeholders may be interested in when a particular place would be included or excluded depending on the cut-off used. The background base map can assist Although this is a continuous function, the problem of deciding where the parameters should be, and what the cut-off should be, has not been eliminated. Therefore, this contour map is equally capable to visualize constraint buffers or discrete classifications. Although this might look obvious, there is no other study that has

mapped out different buffer distances; they all relied on “experts” or sourced their numbers from “literature”. The key advantage of using contour maps is that it is able to relate abstract numbers back to the real world, for the specific factor and study area. This contour map is repeated for Asakareh in the Appendix (Figure **??**), but the most suitable area in the state was in the northern border, and the area is very remote. Nevertheless, that contour map also worked to confirm that all suitable areas are within 4000 metres from the roads, which is the maximum clamp Asakareh used for their roads standardization function.

5.2 Visualization of standardization functions relative to distribution of data

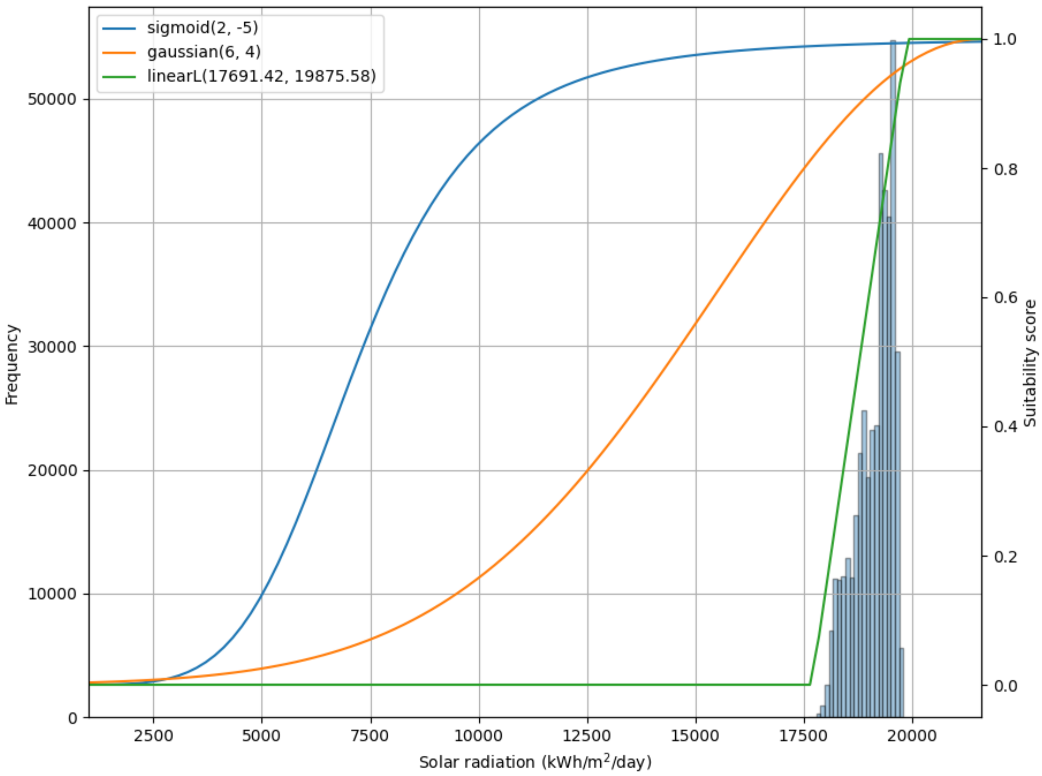

Figure 13: Comparison of the standardization functions used for solar radiation, compared to the distribution of values

Figure 13 explains why Suh’s sigmoid function for insolation gave the state a uniformly a good score, and Asakereh's gaussian function had little variance. They anticipated much lower amount of solar radiation in their respective study areas, which was not adjusted for Arizona's climate. This is another reason why roads and power lines appear to have an outsized influence – because solar radiation was essentially irrelevant when every pixel has roughly the same value. On one hand, it is a good thing that the entire state has an excellent amount of solar radiation compared to South Korea and Iran, making the overall state more suitable. On the other hand, it does not help distinguish between potential sites within Arizona. Therefore, the choice of standardization function depends on purpose of the user, and whether they want to compare domestically or internationally. The range function, which was used as the replacement, considered the entire range of values in the state, and was able to avoid this problem, at the cost of being more localized to the study area, reducing the generalizability of the function in other regions, and being more prone to small changes in the study area, such as zooming into specific sites. This is a form of the modifiable areal unit problem. As the standardization function essentially represents the most desired values, ostensibly justified by scientific evidence or local requirements, which should not vary significantly based on study area. For example, Asakareh justified using 4 kWh/m^2/day as the midpoint of their Gaussian function for solar radiation with the US National Renewable Energy Laboratory's classification as a "good" amount for solar panels (Pohekar et al., 2004), and that solar power needs at least 4.5 kWh/m^2/day to be economically feasible (Aydin et al., 2013). While it does varies for different regions in the world (Arnette et al., 2011), (Sánchez-Lozano et al., 2013), (Charabi et al., 2011), it does not make sense to have one range of values state-wide but a smaller range for a part of the state. The range function were chosen as a replacement in the improved maps as a supposedly "neutral" function, but it is not necessarily neutral or represents stakeholders accurately. This is a limitation in the methodology of this IGS, because proper studies that ask for a range of inputs from stakeholders and consult region-specific values would be able to keep the standardization function fixed in place.

The purpose of this graph is able to highlight in detail the shortcomings of the range function and the methodology used in this IGS. However, none of the studies used this as an aid to decision making, explanation of their results, or evaluation of their methods. At most they would include only the line plot like Asakereh et al. and Suh et al. did, but those graphs lack the context of the actual distribution of the data. The plot is conceptually quite simple, as it is just a line plot with a histogram overlaid and sharing a common x-axis. Even for non-continuous methods, studies assigning discrete scores by classifying ranges of the data would benefit from understanding the naturalness of the breaks. It is likely that a similar visualization was used behind the scenes, but if they are not published then the reader would not know the full context of their choices, or the implications of the functions chosen, or any unintended effects. This graph was clearly helpful to the discussion in this IGS, and this would apply to any other study. This is important because the reader could be another researcher planning an improved study, or a government updating their regulations.

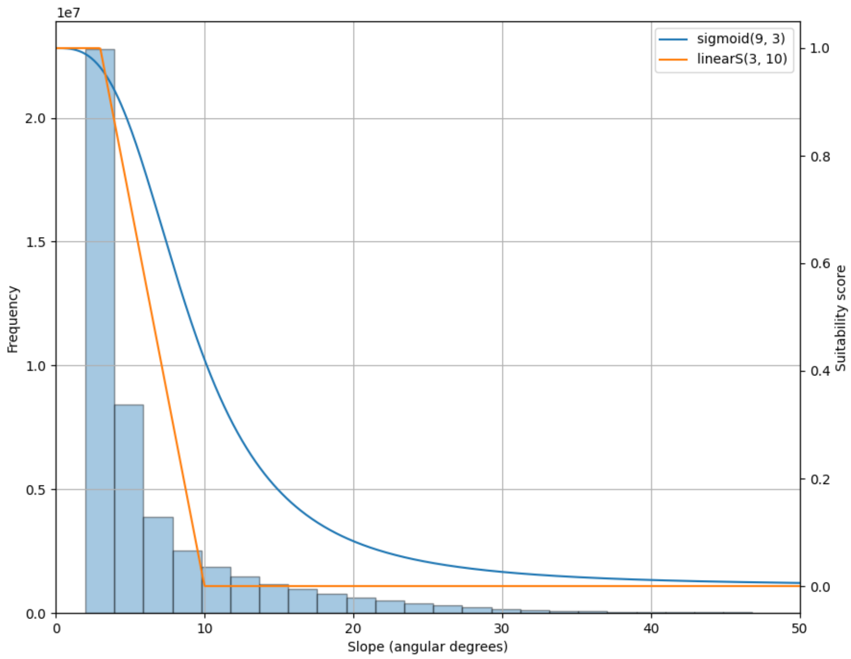

Figure 14: Comparison of the standardization functions used for slope, compared to the distribution of values

Figure 14 shows that Suh's sigmoid is more forgiving of large slope values than Asakareh's linear function. For most slope values, the sigmoid gives a higher standardized score for the slope than the linear function. Asakareh's function gives all slope values above 10 degrees zero, but the sigmoid still gives a score of 0.42. The result is that Suh's slope layer has a narrower range of suitability score values.

5.3 Definition of residential areas

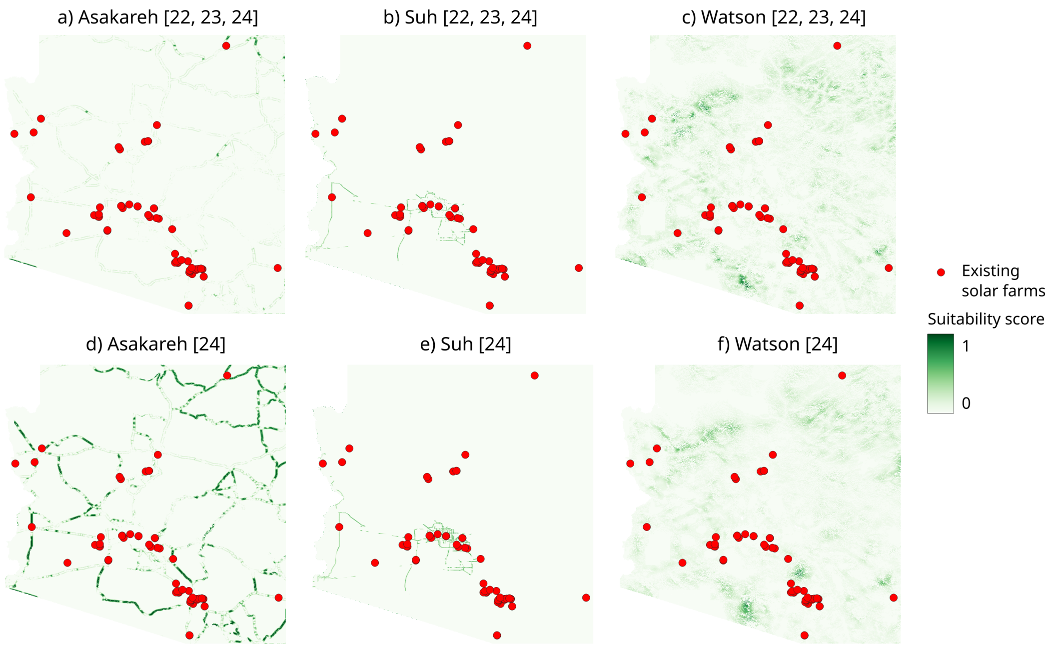

Figure 15: Suitability maps with a more precise definition of "residential areas"

"Residential areas" were defined to be "developed land" of any intensity (high, medium and low) in the CONUS land use dataset. Because of American suburbia, it is not necessarily clear that low intensity urban land are not residential areas. The above analysis included all three, but Figure 15 explores an alternative where the residential areas layer is defined by only high intensity developed land. While not a direct evaluation of the parameters of standardization functions or buffers, it is still a discussion of the analysis done above. Deciding on the definition is not an objective decision so it has to be recognized here to avoid omission by feigning ignorance. Figures 15a shows Asakareh's unmodified parameters, which appear to be slightly larger compared to the original definition of residential areas, but not large enough for residential areas and other factors to be visually impactful, indicating that the roads and

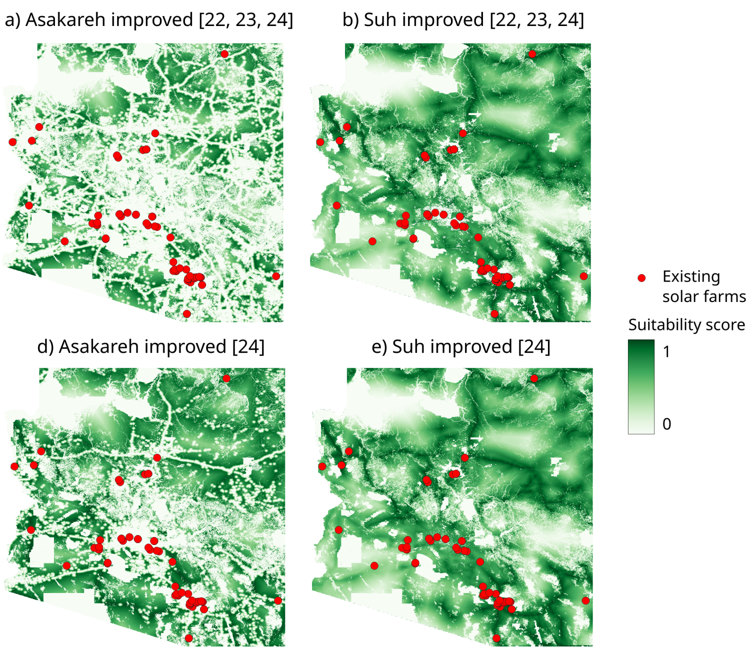

Figure 16: Suitability maps with a more precise definition of "residential areas" for the modified maps

power lines layer was still too narrowly defined. Even though both maps has a higher mean-of-mean, Suh's map showed even less difference compared to the original, indicating that the roads layer was even more predominant than Asakareh's.

Figure 16 shows the modified Asakareh and Suh maps with a narrower definition of residential areas. As noted in Figure 7, the residential areas layer constrained out the most area, so the more precise definition expanded potentially suitable sites considerably. However, it suffers from perhaps a bit too much. One of the biggest uses of site suitability analysis is to exclude sites that are suitable, so the expansion of suitable sites is not necessarily good because it increase the potential area needed for later steps to investigate. The Suh improved map also has this problem, but the difference was not very visible because it was not used as a factor but as a constraint, which is essentially a factor layer with a buffer of 0 m.

6 Conclusion

The first aim of this IGS was to provide a transparent, reproducible, open source, and flexible implementation of site suitability analysis. The open source implementation is transparent because users can upload their parameters for everyone to audit. It is reproducible because the configuration allows them to separate specification from implementation. It is flexible because it can be applied to other areas and topics without large structural changes. The implementation also complements the second aim, which is to evaluate different parameters used in site suitability analysis. It provides visualization tools, including contour maps and plots of the standardization functions, to visualize the impact of abstract numbers into the concrete world.

Appendix

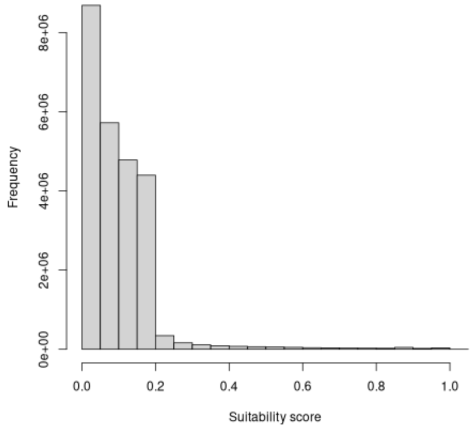

Figure 17: Histogram of all suitability scores using Asakereh's parameters

Criteria	Asakareh modified	Suh modified
Insolation	0.539	0.6571
Temperature	-	0.0838
Slope	$\frac{0.291}{3} = 0.097$	0.0799
Residential	$\frac{0.291}{3} = 0.097$	-
Protected	$\frac{0.291}{3} = 0.097$	-
Roads	0.17	0.0641
Power	-	0.1151
Sum	1	1

Table 6: Weights of the factors for each study that was further modified

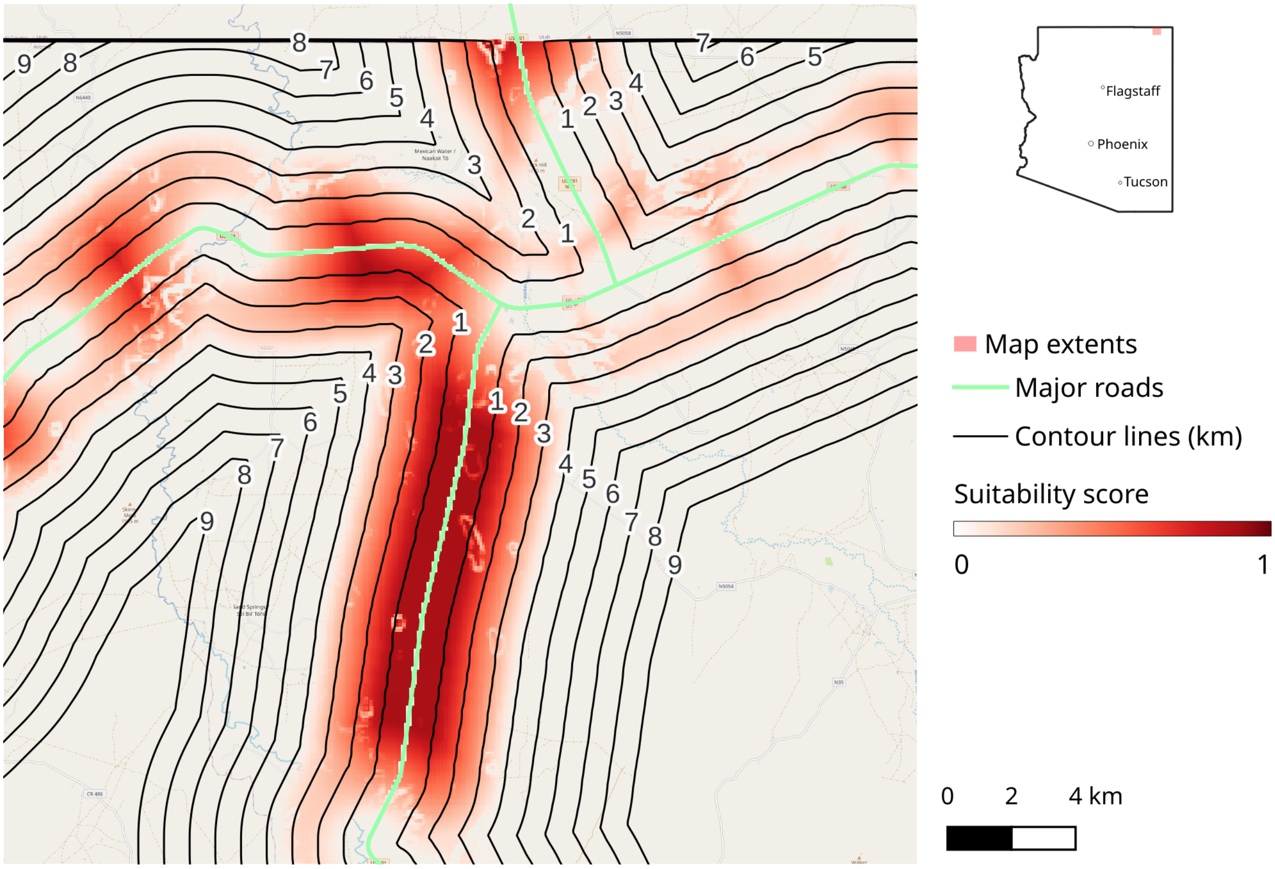

Figure 18: Map of major roads in the northern border, and the contour lines of proximity to roads, overlaid on the suitability map using Asakareh's parameters. Zero suitability is not shown

Research Ethics
Office

Franklin Wilkins Building
5.9 Waterloo Bridge Wing
Waterloo Road
London SE1 9NH
Telephone 020 7848 4020/4070/4077
rec@kcl.ac.uk

KING'S College LONDON

15/05/2021

Yik Ching Tsui

Dear Yik Ching

Site suitability analysis of photovoltaic solar farms in Arizona using GIS and AHP

Thank you for submitting your Minimal Risk Self-Registration Form. This letter acknowledges confirmation of your registration; your registration confirmation reference number is MRSU-20/21-23415

IMPORTANT CORONAVIRUS UPDATE: In light of the COVID-19 pandemic, the College Research Ethics Committee has temporarily suspended all primary data collection involving face to face participant interactions, unless the data collection fall under one of the exemptions and fulfils the criteria outlined by CREC at the link below:

https://internal.kcl.ac.uk/innovation/research/ethics/applications/COVID-19-Update-for-Researchers

Ethical Clearance

Ethical clearance for this project is granted. However, the clearance outlined in the attached letter is contingent on your adherence to the latest College measures when conducting your research. Please do not commence data collection until you have carefully reviewed the update and made any necessary project changes.

Ethical clearance is granted for a period of **one year** from today's date and you may now commence data collection. However, it is important that you have read through the information provided below before commencing data collection:

As the Minimal Risk Registration Process is based on self-registration, your form has not been reviewed by the College Research Ethics Committee. It is thereforeyour responsibility to ensure that your project adheres to the Minimal Risk Guiding Principles **and the agreed protocol does not fall outside of the criteria for Minimal Risk Registration. Your project may be subject to audit by the College Research Ethics Committee and any instances in which the registration process is deemed to have been used inappropriately will be handled as a breach of good practice and investigated accordingly.**

Record Keeping:

Please be sure to keep a record of your registration number and include it in any materials associated with this research. It is the responsibility of the researcher to ensure that any other permissions or approvals (i.e. R&D, gatekeepers, etc.) relevant to their research are in place, prior to conducting the research.

In addition, you are expected to keep records of your process of informed consent and the dates and relevant details of research covered by this application. For example, depending on the type of research that you are doing, you might keep:

- A record record of all data collected and all mechanisms of disseminated results.
- Documentation of your informed consent process. This may include written information sheets or in cases where it is not appropriate to provide written information, the verbal script,or introductory material provided at the start of an online survey.
  Please note: For projects involving the use of an Information Sheet and Consent Form for recruitment purposes, please ensure that you use the KCL GDPR compliant Information Sheet & Consent Form Templates
- Where appropriate, records of consent, e.g. copies of signed consent forms or emails where participants agree to be interviewed.

Audit:

You may be selected for an audit, to see how researchers are implementing this process. If audited, you and your Supervisor will be asked to attend a short meeting where you will be expected to explain how your research meets the eligibility criteria of the minimal risk process and how the project abides by the general principles of ethical research. In particular, you will be expected to provide a general summary of your review of the possible risks involved in your research, as well as to provide basic research records (as above in Record Keeping) and to describe the process by which participants agreed to participate in your research.

Remember that if you at any point have any questions about the ethical conduct of your research, or believe you may have gained the incorrect level of ethical clearance, please contact your supervisor or the Research Ethics Office.

Data Protection Registration

If you indicated in your minimal risk registration form that personal data would be processed as part of this research project, this letter also confirms that you have also met your requirements for registering this processing activity with King's College London in accordance with the General Data Protection Regulation (GDPR).

More information about how the GDPR affects researchers can be found here: https://internal.kcl.ac.uk/innovation/research/Research-Governance/how-does-GDPR-affect-research/How-does-GDPR-affect-research

Please note that any changes to the storage, management, or type of personal data being collected should also be included in a modification request.

We wish you every success with your project moving forward.

With best wishes,

The Research Ethics Office

On behalf of the College Research Ethics Committee

References

Al Garni, H. Z. and A. Awasthi (2017). “Solar PV power plant site selection using a GIS-AHP based approach with application in Saudi Arabia”. *APPLIED ENERGY* 206, pp. 1225–1240.

Ali, S. et al. (2019). “GIS based site suitability assessment for wind and solar farms in Songkhla, Thailand”. *RENEWABLE ENERGY* 132, pp. 1360–1372.

Arán Carrión, J. et al. (2008). “Environmental decision-support systems for evaluating the carrying capacity of land areas: Optimal site selection for grid-connected photovoltaic power plants”. *Renewable and Sustainable Energy Reviews* 12.9, pp. 2358–2380.

Arnette, A. N. and C. W. Zobel (2011). “Spatial analysis of renewable energy potential in the greater southern Appalachian mountains”. *Renewable Energy* 36.11, pp. 2785–2798.

Asakereh, A., M. Soleymani, and M. J. Sheikhdavoodi (2017). “A GIS-based Fuzzy-AHP method for the evaluation of solar farms locations: Case study in Khuzestan province, Iran”. *SOLAR ENERGY* 155, pp. 342–353.

Athanasiou, L. S., D. I. Fotiadis, and L. K. Michalis (2017). “5 - Plaque Characterization Methods Using Optical Coherence Tomography”. *Atherosclerotic Plaque Characterization Methods Based on Coronary Imaging*. Equation 5.3. Oxford: Academic Press, pp. 95–113.

Aydin, N. Y., E. Kentel, and H. Sebnem Duzgun (2013). “GIS-based site selection methodology for hybrid renewable energy systems: A case study from western Turkey”. *Energy Conversion and Management* 70, pp. 90–106.

Baban, S. M. and T. Parry (2001). “Developing and applying a GIS-assisted approach to locating wind farms in the UK”. *Renewable Energy* 24.1, pp. 59–71.

CGIAR (2018). “SRTM 90m DEM Digital Elevation Database”. `srtm.csi.cgiar.org/download/`.

Charabi, Y. and A. Gastli (2011). “PV site suitability analysis using GIS-based spatial fuzzy multi-criteria evaluation”. *Renewable Energy* 36.9, pp. 2554–2561.

Domínguez Bravo, J., X. García Casals, and I. Pinedo Pascua (2007). “GIS approach to the definition of capacity and generation ceilings of renewable energy technologies”. *Energy Policy* 35.10, pp. 4879–4892.

Fick, S. and R. Hijmans (2020). "WorldClim 2: new 1km spatial resolution climate surfaces for global land areas". *International Journal of Climatology* 37 (12). `worldclim.org/data/worldclim21.html`, pp. 4302–4315.

Gastes, D. and W. Gaul (2012). "The Consistency Adjustment Problem of AHP Pairwise Comparison Matrices". *Quantitative Marketing and Marketing Management: Marketing Models and Methods in Theory and Practice.* Ed. by A. Diamantopoulos, W. Fritz, and L. Hildebrandt. Wiesbaden: Gabler Verlag, pp. 51–62.

Giamalaki, M. and T. Tsoutsos (2019). "Sustainable siting of solar power installations in Mediterranean using a GIS/AHP approach". *RENEWABLE ENERGY* 141, pp. 64–75.

HIFLD (2021). "Electric Power Transmission Lines". `protectedplanet.net`.

Merrouni, A. A. et al. (2018). "Large scale PV sites selection by combining GIS and Analytical Hierarchy Process. Case study: Eastern Morocco". *RENEWABLE ENERGY* 119, pp. 863–873.

Mondino, E. B., E. Fabrizio, and R. Chiabrando (2015). "Site Selection of Large Ground-Mounted Photovoltaic Plants: A GIS Decision Support System and an Application to Italy". *INTERNATIONAL JOURNAL OF GREEN ENERGY* 12.5, pp. 515–525.

OpenStreetMap (2021). "highway='motorway' or highway='trunk' or highway='primary'". `overpass-turbo.eu/`.

Pohekar, S. and M. Ramachandran (2004). "Application of multi-criteria decision making to sustainable energy planning—A review". *Renewable and Sustainable Energy Reviews* 8.4, pp. 365–381.

Saaty, T. L. (1980). *The Analytic Hierarchy Process.* New York: McGraw-Hill.

Sanchez-Lozano, J. M. et al. (2014). "GIS-based photovoltaic solar farms site selection using ELECTRE-TRI: Evaluating the case for Torre Pacheco, Murcia, Southeast of Spain". *RENEWABLE ENERGY* 66, pp. 478–494.

Sánchez-Lozano, J. M. et al. (2013). "Geographical Information Systems (GIS) and Multi-Criteria Decision Making (MCDM) methods for the evaluation of solar farms locations: Case study in south-eastern Spain". *Renewable and Sustainable Energy Reviews* 24, pp. 544–556.

Shorabeh, S. N. et al. (2019). “A risk-based multi-criteria spatial decision analysis for solar power plant site selection in different climates: A case study in Iran”. *RENEWABLE ENERGY* 143, pp. 958–973.

Suh, J. and J. R. S. Brownson (2016). “Solar Farm Suitability Using Geographic Information System Fuzzy Sets and Analytic Hierarchy Processes: Case Study of Ulleung Island, Korea”. *ENERGIES* 9.8.

Sun, Y.-w. et al. (2013). “GIS-based approach for potential analysis of solar PV generation at the regional scale: A case study of Fujian Province”. *Energy Policy* 58, pp. 248–259.

Tahri, M., M. Hakdaoui, and M. Maanan (2015). “The evaluation of solar farm locations applying Geographic Information System and Multi-Criteria Decision-Making methods: Case study in southern Morocco”. *RENEWABLE & SUSTAINABLE ENERGY REVIEWS* 51, pp. 1354–1362.

UNEP-WCMC and IUCN (2021). “Protected Planet: The World Database on Protected Areas (WDPA) and World Database on Other Effective Area-based Conservation Measures (WD-OECM)”. `protectedplanet.net`.

USGS (2019). “National Land Cover Database”. `usgs.gov/centers/eros/science/national-land-cover-database`.

Uyan, M. (2013). “GIS-based solar farms site selection using analytic hierarchy process (AHP) in Karapinar region, Konya/Turkey”. *RENEWABLE & SUSTAINABLE ENERGY REVIEWS* 28, pp. 11–17.

Watson, J. J. W. and M. D. Hudson (2015). “Regional Scale wind farm and solar farm suitability assessment using GIS-assisted multi-criteria evaluation”. *LANDSCAPE AND URBAN PLANNING* 138.SI, pp. 20–31.

Wiguna, K. A., R. Sarno, and N. F. Ariyani (2016). “Optimization Solar Farm Site Selection Using Multi-Criteria Decision Making Fuzzy AHP and PROMETHEE: Case Study in Bali”. *PROCEEDINGS OF 2016 INTERNATIONAL CONFERENCE ON INFORMATION & COMMUNICATION TECHNOLOGY AND SYSTEMS (ICTS)*. International Conference on Information & Communication Technology and Systems. IEEE, pp. 237–243.

Xiang, W. N. and D. L. Whitley (1994). “Weighting land suitability factors by the PLUS method.” *Environment and Planning B: Planning and Design* 21, pp. 273–304.

Al-Yahyai, S. et al. (2012). “Wind farm land suitability indexing using multi-criteria analysis”. *Renewable Energy* 44, pp. 80–87.