Mental Health and Land Cover: A Global Analysis Based on Random Forest

Abstract

Nature features and processes in living environments can help to reduce stress and improve mental health. Different land types have disproportionate impacts on mental health. However, the relationships between mental health and land cover are inconclusive. Here, we show the complex relationships between mental health and percentages of eight land types based on the random forest method and Shapley additive explanations. The accuracy of our model is 93.09%, while it is normally no more than 20% in previous studies. According to the analysis results, we estimate the average effects of eight land types. Shrubland, wetland, and bare land have the highest effects on mental health due to their scarcity in living environments. Cropland, forest, and water could improve mental health in the high population-density areas. The impacts of urban land and grassland increases are tiny compared with other land types. Due to scarcity values, the current land cover composition influences people's attitudes toward a certain land type. This paper provides insights to formulate better land-use policies to improve mental health and eventually to achieve a sustainable society based on a machine learning case study.

Keywords:

Mental Health; Land Cover; Environmental Valuation; Random Forest; SHAP

Introduction

Natural land cover in people's living environments has positive impacts on human well-being and mental health (Alcock et al. 2014; Bratman et al. 2019; Diaz et al. 2019; Diener et al. 2018; MacKerron and Mourato 2013; Malek and Verburg 2020), which is mainly driven by ecosystem service (Costanza et al. 1997; Costanza et al. 2014; Diaz et al. 2018). Environments with natural land cover provide a variety of ecosystem services that can benefit people (Carpenter et al. 2009; Chaplin-Kramer et al. 2019; Diaz et al. 2018; Diaz et al. 2019), such as recreational activities (MacKerron and Mourato 2013), air pollution abatement (Chameides et al. 1988; Eitelberg et al. 2016; Mendoza-Ponce et al. 2018), and the creation of aesthetic, artistic, and scientific values for human beings (Felipe-Lucia et al. 2018; Seresinhe et al. 2015). However, 2.7% of global semi-natural or natural land has been continuously converted to other land types, specifically cropland and built-up area, from 1992 to 2015 (OECD 2017). With a decrease in natural land cover, the estimated aggregate value of ecosystem services from 1997 to 2011 has slashed by $4.3 trillion globally and annually (Costanza et al. 2014). As the benefits of the natural land cover are vast and enormous (Costanza et al. 2014; Diaz et al. 2018), there is a need to clarify the effect of land cover change on mental health. Because the world continues to develop and urbanize (OECD 2017), the share of natural land cover in people's living environments will likely keep decreasing. Due to the trade-off between development and the desire for natural land, there is a further need to detect how much alteration of land cover composition affects future mental health and where the effects of a certain land type change are the highest.

To estimate the impacts of land cover change on mental health, relatively precise relationships between land cover and mental health are desired. This study employs 100,956 observations in an international survey from 37 countries and applies a non-parametric machine learning method, random forest, to obtain a high-fit model. However, because the random forest is typically model-agnostic, we must use effective tools to make the results understandable. A well-developed theory, Shapley value, can directly illustrate the feature value's contribution to mental health (Lundberg et al. 2020; Lundberg and Lee 2017). Based on the attribute of random forest, we create a new way to make the result more understandable and reliable. According to explainable and accurate results, our research provides more information to formulate sustainable land-use policies to improve residents' mental health.

Literature Review

The relationship between land cover and mental health has been long investigated (Bratman et al. 2019; Diener et al. 2018; Hartig et al. 2014; Wang et al. 2021). Natural environments could reduce air pollution and stressor exposure and increase physical activity and social contacts, which eventually improve health and well-being (Bratman et al. 2019; Dzhambov et al. 2020; Hartig et al. 2014). Furthermore, they are also the primary mediator in attenuating the negative impacts of air pollution on human health (Li and Managi 2021a). A reduction in air pollution significantly improves human well-being, especially in metropolitan areas (Li and Managi 2021c). Some findings display that blue-green spaces are critical to maintaining better mental health during the COVID-19 pandemic lockdown since they cut down stressor exposure (Pouso et al. 2021). A controlled laboratory study shows that the impacts of natural sounds and images on stress and mental status are positive (Berman et al. 2012). Substantial and significant evidence shows that people in natural environments experience higher well-being and better emotions (MacKerron and Mourato 2013; Seresinhe et al. 2019). An empirical study indicates that individuals have significantly better mental health if they move to greener areas, and the effects last several years (Alcock et al. 2014). Furthermore, environmental degradation and the absence of green spaces are causal factors of mental health issues, according to a well-designed casual mechanisms study (Wang et al. 2021). Green space disproportionately affects human health among different socio-economic and demographic groups, so those variables must be carefully considered (Wu and Kim 2021). On the other hand, urban land cannot be simply regarded as a negative factor. People desire more urban land to support a better life if the cities are too crowded (Li and Managi 2021b). Although the relationship between land cover and mental health has long been detected and discussed, the detailed impacts remain elusive. In other words, making the value of land cover change understandable and comparable is needed to achieve a sustainable society, maintain human mental health, and lead public policy.

To probe the comparable value of land cover, the quantitative land cover data play a distinct role in the empirical analyses. Land cover data are various in previous studies, which are shares of one or several land types in each defined area (Alcock et al. 2014; Kopmann and Rehdanz 2013; Krekel et al. 2016; Tsurumi et al. 2018; White et al. 2013), or greenery index, mainly normalized difference vegetation index (NDVI) (Dzhambov et al. 2018; Wang et al. 2019). Land cover data include several land types, which are more straightforward, but their temporal resolution is longer, at least one year (Gong et al. 2019). The NDVI can be obtained every eight days with the highest temporal resolution but only depicts greenery. Although these two types of data are widely used in previous studies, the land cover data are more suitable for the research, which does not only concentrate on greenery. Monetary values of land covers can be estimated (Frey et al. 2010; Tsurumi and Managi 2015). For example, residents in German are willing to pay 23 Euro for a 1-ha increase in green urban areas within 1000 meters of their houses (Krekel et al. 2016). The monetary value estimation is based on the marginal substitute rate (MSR) between land cover and income. The MSR strongly relies on the marginal effects of land cover and income on well-being or mental health indicators from the statistical models (Frey et al. 2010), and most previous studies follow this method. Obviously, the accuracy of statistical models dramatically affects the reliability of the estimated monetary value, and the assumption of the models is vital. Currently, the linear assumption is still widely applied.

The advantage of machine learning methods is their high accuracy. The goodness of fit in previous studies is no more than 20%. The relationship between mental health and land cover is mainly assumed to be linear (Alcock et al. 2015; Tsurumi et al. 2018; White et al. 2013), quadratic, or logarithmic (Kopmann and Rehdanz 2013; Krekel et al. 2016; Tsurumi and Managi 2015). On the one hand, the linear relationship is straightforward and has a clear attitude towards a certain land type. These models are based on a simple assumption that amounts of certain land types always have the same effect on mental health, no matter the current land cover status. In this case, people should live in the environment only with this land type that has the most positive effect on mental health. This is the main shortcoming of this assumption, and it is far from reality. On the other hand, the non-linear relationship is more in line with reality. Preferences for certain land types depend on the current land cover allocation (Kopmann and Rehdanz 2013; Krekel et al. 2016; Tsurumi and Managi 2015). It is the fundamental idea to build a non-linear model. There are two types of non-linear models. One assumes the relationship between the coverage of land types, and well-being is logarithmic (Kopmann and Rehdanz 2013), and the other assumes the relationship is quadratic (Krekel et al. 2016; Tsurumi and Managi 2015). In the logarithmic relationship assumption, when certain coverage keeps increasing, the effect of this land type on well-being or mental health gets down, but the direction of this attitude does not change (Kopmann and Rehdanz 2013). In the quadratic relationship assumption, when the share of land cover changes, the intensity of effects on mental health will vary and even may alter the direction of the impact. Though these non-linear assumptions are better than linear ones, they still have low accuracy. The accuracy of machine learning methods, such as random forest, typically exceeds 70% (Kim and Kim 2022; Wang et al. 2021). High accuracy means that the relationships estimated by the trained model are closer to the actual situation. To make the policies based on the analysis results reliable, we must guarantee assumptions similar to the real world. Machine learning has fewer assumptions than previous methods (Kim and Kim 2022). Therefore, using machine learning methods to analyze the relationships is necessary.

Materials and Methodology

Materials

Survey Information

Our study is based on an international survey conducted by Kyushu University from July 2015 to March 2017, covering 37 countries, including both developed and developing countries. The investigation periods for each country are generally less than one month. Moreover, to guarantee the reliability of the survey, the same questionnaires are used, while currency-related questions are based on local currencies. The population and GDP of these countries account for 68.58% of the global population and 82.67% of the worldwide GDP in 2017, respectively (**Appendix Table A1**). This survey was conducted to obtain individual mental health and several other demographic and socio-economic characteristics. The total number of observations that were recorded is 100,956. However, due to a lack of geographical location or records, 95,571 observations are kept. In addition, because some individuals did not provide income information, 89,273 observations are used in the calculations (Descriptive statistics of the features shown in **Appendix Table A2**).

Mental Health

We include the twelve-item General Health Questionnaire (GHQ-12) in the survey to assess individual mental health. As a validated and reliable instrument, the GHQ-12 has been widely used in previous mental health studies, e.g. (El-Metwally et al. 2018; Hankins 2008; Quek et al. 2001). The GHQ-12 comprises 12 items to assess the individual mood states. Each item of the GHQ-12 has four potential answer options, encoded by the Likert method as 0, 1, 2, and 3, from negative to positive. The mental health assessment score is computed as the summed score of all 12 items. Thus, the output variable of our study is a discrete numeric variable ranging from 0 to 36. The current random forest method is designed to execute either regression or classification. The algorithm would perform the classification task using the discrete output variable, assuming the output is categorical. However, adjacent scores of the mental health assessments are related, i.e., they are ordinal rather than categorical. **Figure 1** illustrates the statistical distribution of the mental health assessment score. Most people get 24 points in the assessment, and significantly more people score 24 to 30 points than others. In this situation, if we perform the classification random forest, the classification accuracy for the people with lower or higher scores would be extremely low due to the unbalanced output distribution. Thus, we assume that the mental health assessment score is continuous.

Global Land Cover Data

As for the land cover, we use remote sensing data compiled by Tsinghua University (http://data.ess.tsinghua.edu.cn/), because, to the best of our knowledge, it is the dataset with the highest global resolution, at approximately 30 meters. This dataset is the 2017 global land cover. It classifies land cover into ten categories: cropland, forest, grassland, shrubland, wetland, water, tundra, urban land, bare land, and snow/ice (Gong et al. 2019). We calculated areas of each land type surrounding our survey respondents with these data. To estimate the impact of land cover in our analysis, we use the percentages of each land type within a radius of 5,000 meters around each respondent following the previous study (Krekel et al. 2016). Eight land types are taken as the land cover data because little tundra and snow/ice are in the most analyzed area. After the random forest analysis, we will estimate the Shapley values of each land type. To make the estimation more efficient, we reduce the ranges of the percentages of shrubland, wetland, water, and bare land. The range of shrubland is set from 0 to 40% because only 282 respondents live with more than 40% shrubland and some even exceed 60%. Similarly, the wetland, water, and bare land ranges are set from 0 to 3%, from 0 to 50%, and 0 to 20%, respectively.

Individual Income Data

Converting the impacts of land cover on mental health into monetary values is an effective way to make them understandable to the public without professional knowledge. Moreover, individual income is also an essential factor affecting mental health. In this study, we use the difference between individual income and GDP per capita in the respondent's country (DIG) as the income variable because the income in the survey is based on local currency. Additionally, the same amounts of money have different effects in different countries. For example, the ability and sense of 100 USD in the U.S. and Sri Lanka are not actually the same for the local people. The DIG is calculated as follows:

$$DIG_i = \frac{Inc_i - GDPPC_i}{GDPPC_j} \tag{1}$$

where DIG_i is the DIG of the respondent i, Inc_i is the individual income of the respondent i, and $GDPPC_j$ is the GDP per capita of the respondent i's country in the surveyed year. It must be noted that the units of Inc_i and $GDPPC_i$ are the current price USD. In the survey, the income data is based on local currencies. To unify the income data, we convert the local currencies into USD. We employed the official annual average exchange rate of the year during the conversion process when the survey was conducted in that country. Moreover, the survey questionnaires required the respondents to select their gross household income range rather than to ask them to report their exact gross household income. Thus, we take the midpoint of the range selected by the respondent as the gross household income. For instance, in the U.S., if respondents reported that their gross household income ranged from 50,000 to 60,000 USD per year, their household income is 55,000 USD per year in the analysis. According to previous studies (Mackerron and Mourato 2009; Yuan et al. 2018), the calculation of the annual gross individual income is as follows:

$$Inc_i = \frac{GHI_i}{(Adu_i + 0.7Chi_i)^{0.5}} \tag{2}$$

where Inc_i is the annual gross individual income of the respondent i, GHI_i represents the gross annual household income of the respondent i, Adu_i represents the number of adults in the respondent i's household, and Chi_i represents the number of children in the respondent i's household. Limited by the household size and the maximum value of the selection interval, individual income rarely exceeds three times GDP per capita in the respondent's country. To improve the estimation accuracy, we set the range of the difference from -1 to 3. **Figure 2** demonstrates the statistical distribution of the DIG in the respondent's country.

Other Control Variables

We add several other control variables because mental health status may differ according to people's socio-economic and demographic characteristics: age, gender, employment, educational background, emotion in the surveyed week, children number, self-reported health, self-reported personality, and evaluation of living environment. Among these control variables, employment, educational background, and self-reported personality are categorical. We use the one-hot encoding method to convert them into a series of dummy variables. Thus, in the analysis, every respondent has 49 features and one output variable. The descriptions of the features are listed in **Appendix Table A3**.

Methodology

To detect influential factors on mental health and confirm the relationship between mental health and land cover, linear regression methods, such as ordinary least square (OLS) and ordered logit regression (OLR), is widely applied, e.g., (Akpinar et al. 2016; Alcock et al. 2015; Alcock et al. 2014; Li and Managi 2021b). The studies evaluate the monetary values of land cover through OLS estimation because the OLS is straightforward to explain. Additionally, the investigations employing the OLR are theoretically more reasonable since mental health evaluation is a discrete variable rather than a quantitative and continuous variable in most studies (Alcock et al. 2015; Alcock et al. 2014). The OLR is a typical classification function based on logistic regression. However, these two models rely on linear assumptions and cannot directly illustrate the importance of predictors on the outcome variable. Putting another way, based on the linear assumption, a 1-unit increase in a certain land type always has the same effect on mental health, whatever the status quo. This is not consistent with the actual situation. Random forest could smoothly grasp the non-linear relationship, which bundles many different decision tree algorithms as an improved boosting method (Breiman 2001). The decision tree algorithms are non-linear and closer to real-world situations (Czajkowski and Kretowski 2016). Moreover, according to the out-of-bag (OOB) test, each feature's importance on the output variable is estimated.

The Regression Decision Tree

A single decision tree is the fundamental element of the random forest method. There are two types of trees, namely, the decision tree for either classification or regression (Breiman 2001; Liaw and Wiener 2002). **Figure 3** shows a simple example of a three-layer regression decision tree. To complete the prediction in the example tree, the algorithm passes three internodes and makes three judgments at most. As the example illustrated in **Figure 3**, we assume only three features, the unemployed dummy variable, self-reported health, and the DIG, affect the output variable, mental health. The rules of each judgment and feature range splits are a critical part of machine learning training. A large amount of data is put into the algorithm to train the decision tree to decide the rules of each judgment and feature range splits. We employ a greedy approach to train regression decision trees (Breiman et al. 2017). This approach chooses the features and splits their ranges to minimize the residual sum of squares (RSS) as follows:

$$RSS = \sum_{l \in leaves} \sum_{i \in C_l} (y_i - \bar{y}_{C_l})^2 \tag{3}$$

where l is a leaf, C_l is the cases in leaf l, y_i is the observed value and $\bar{y}_{C_l}$ is the average observed value in leaf l. Unless the RSS is smaller than the defined threshold or the number of remaining cases in the end leaf is less than the defined threshold, the number of internode of trees will keep increasing (Breiman et al. 2017).

Random Forest

In most cases, a single regression decision tree is insufficient to fit the output variables and usually causes an over-fitting analysis. To solve this issue, the random forest ensemble a bundle of decision trees and let them vote for the results (Breiman 2001). The voting strategy for regression is taking the average value of all individual predictions as the random forest prediction. Bagging and bootstrapping are performed to guarantee the accuracy and reliability of the random forest (Liaw and Wiener 2002). Bootstrapping is the sampling technique of the random forest. Firstly, we set the number of trees in our random forest as N_{tree}. We extract N_{tree} samples with replacement from the original data and the sample sizes are 2/3 data of the total sample. Every decision tree utilizes the bootstrapped dataset. However, only N_{try} random features are used in a single decision tree, rather than all. For regression tasks, N_{try} is roughly one third of the total number of features. After training, the random forest can predict the output variable by aggregating the votes from each tree. Using the bootstrapped dataset and the aggregate of votes, this process is terminologically called "bagging". Additionally, roughly 1/3 of the total sample is not employed in the training process named the OOB dataset. The OOB dataset is applied to test the accuracy of the random forest through the OOB score, which is the proportion of OOB observations correctly predicted by the trained random forest. Due to the bagging technique, cross-validation is unnecessary. The reliable trained models have a relatively high OOB score.

In the random forest, most parts are built randomly, while only three critical parameters must be decided by the users, specifically, the minimum number of remaining observations in end leaves, N_{tree} and N_{try}. Firstly, the minimum number of observations in end leaves decides where the split stops because our random forest follows the greedy approach. Moreover, the random forest accuracy will increase to some extent when more trees are included. However, the cost of infinitely increasing N_{tree} is a dramatic increment of calculation power and calculating time. Moreover, when N_{tree} exceeds a particular value, the marginal effect of increasing the number is minimal. Accordingly, considering the size of our dataset and computing ability, the number of trees is set to 1,000. Moreover, the number of features used in the decision trees, N_{try}, is another vital factor. A large N_{try} might lead to overfitting, while a small N_{try} might cause underfitting. Previous studies indicate that roughly one third of the total number is recommended (Breiman 2001; Breiman et al. 2017; Liaw and Wiener 2002). Thanks to our relatively sufficient computing ability of a high-performance computer, we test the most possible N_{try} values and the number of remaining cases in the end leaf based on 10-fold cross-validation. According to the test, the goodness of fit peaks when N_{try} value is nine, and the number of remaining cases in the end leaf is two. In plain language, each decision tree would randomly pick nine features from the dataset, and each end leaf at least includes two observations.

Variable Importance

The random forest could estimate the importance of each feature on the output variable. The basic idea of importance estimation in the random forest is to calculate the reduction in accuracy between before and after excluding a specific feature (Breiman 2001). The reduction in the accuracy of a particular feature would be higher when it is more important to successfully predict the output variable compared with other features. This reduction is similar to the partial R^2 in the OLS algorithm. There is no need to select the features in the random forest algorithm since the issues, such as multicollinearity, do not influence the accuracy of the random forest. Yet, multicollinearity is a fatal problem in the OLS.

Shapley Additive Explanations (SHAP)

Although the accuracy of random forest is high, it is challenging to understand and explain the results (Friedman 2001; Greenwell 2017; Lundberg et al. 2020). SHAP is an advanced approach explaining the contributions of each feature locally based on the game theoretically optimal Shapley values (Lundberg and Lee 2017). To explain the contributions of features, each feature of the observation is a "player" in a game, and the prediction value is the payout. Shapley values help us to fairly distribute the payout among the players (Lundberg and Lee 2017; Molnar 2020). The Shapley value of a feature value is estimated as follows:

$$S_{jx} = E[\frac{1}{p!}\sum_{J} g^{j|\pi(J,j)}(x)] \tag{4}$$

where x represents a specific observation of interest, j represents a particular feature of interest, S_{jx} represents the Shapley value of the feature j of the observation x, J represents a permutation of the set of indices $\{1, 2, \ldots, p\}$ corresponding to an ordering of p features included in our random forest model, $\pi(J, j)$ represents the set of the indices of the features contained in J before the j-th variable, and $g^{j|\pi(J,j)}(x)$ represents the estimated contribution value of feature j of the observation x with a specific permutation. $g^{j|\pi(J,j)}(x)$ is calculated as follows:

$$g^{j|\pi(J,j)}(x) = E\left(f(X)\middle|X^1 = x^1, \dots, X^{j-1} = x^{j-1}, X^j = x^j\right) - E\left(f(X)\middle|X^1 = x^1, \dots, X^{j-1} = x^{j-1}\right) \tag{5}$$

where X represents a matrix of random values of features, $f()$ represents our trained random forest model, $E\left(f(X)\middle|X^1 = x^1, \dots, X^{j-1} = x^{j-1}, X^j = x^j\right)$ is expected value of the predictions of X, when we set $X^1 = x^1, \dots, X^{j-1} = x^{j-1}, X^j = x^j$, and $E\left(f(X)\middle|X^1 = x^1, \dots, X^{j-1} = x^{j-1}\right)$ is expected value of the predictions of X, when we set $X^1 = x^1, \dots, X^{j-1} = x^{j-1}$. It must be noted that generally, the random values are deemed to have no explaining ability. In real computation, the random dataset X is not randomly generated but randomly picked up from our dataset. In our analysis, we set the dataset size of X as 900, approximately 1% of the total dataset. A bigger dataset size here would definitely increase the computation time. To estimate the Shapley values efficiently, we use ten random permutations of all features. Of course, more permutations lead the estimated values to the real values, but the computing time is not affordable.

The connection between Features' Values and Their SHAP Values

SHAP value is the feature value's contribution to each observation's current mental health status. For example, in one observation's living environment, the urban land takes 99.60%, and its SHAP value is -0.056. This observation's living environment is monotonous and full of urban land, which might negatively affect mental health. In another instance, the observation has 73.98% of urban land, and its SHAP value is 0.044. The impacts of a certain feature on mental health might be associated with the current status. We employ linear regression to probe the relationship between the feature value and its contribution to mental health. However, since this research is global, a huge spatial extent makes the globally unified relationship suspicious. Estimating the relationship locally is more rational.

Building a series of local datasets is the critical point. In the model training process, the location information is also included, which is the longitude and latitude of the observation. Some decision trees pick up these features. These trees divide the global extent into several zones. The observation's location belongs to zones divided by different trees. We obtain a bag of boundaries. The medians of the boundaries in each direction are regarded as the dividing lines. Every observation is surrounded by a rectangle of dividing lines, and others within one observation's zones are the neighbors. The neighbor zones differ by location. The local relationship is estimated based on one observation and others located in its neighbor zone, so the relationship coefficients also spatially vary. The estimated process is as follows:

$$S_{jX} = \alpha_{jx} X_x^j + \beta_{jx} \tag{6}$$

where α_{jx} and β_{jx} are the slope and the intercept of the local relationship between feature j's value and its SHAP value based on x's neighbor zone, X_x^j is a vector of the feature j's values in x's neighbor zone, and S_{jX} is a vector of the SHAP values corresponding to X_x^j. According to the local relationship coefficient, we could interpret the marginal contribution of an increase in a certain feature to mental health.

Monetary Values of Land Cover

To make the impacts of land cover change on mental health understandable, we estimate the monetary values of land cover. We take the marginal substitution rate (MSR) of land cover and income as the monetary values, and it is estimated as follows:

$$MSR_{jx} = \frac{\alpha_{jx}}{\alpha_{INCx}} \tag{7}$$

where MSR_{jx} is the MSR of feature j in the observation x's location, and α_{INCx} is the local relationship coefficient between the income value and its SHAP value based on the observations in x's neighbor zone. In this equation, we require the coefficients α_{jx} and α_{INCx} are significant (p value < 0.1), or the MSR would be set to zero. Based on this equation, the monetary values can be explained by how much DIG changes equals a 1% increase in a specific land cover.

Results

In this study, the trained random forest employs 1000 trees. Nine features are randomly chosen in the bootstrapped datasets to train each tree. Every end leaf must have at least two observations. The accuracy of the random forest is 93.09%, whereas the accuracy of the OLS is only 42.69%. Moreover, the root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) of the random forest are 1.66, 2.74, and 1.24, respectively, while the RMSE, MSE, and MAE of the OLS are 4.77, 22.75, and 3.64. In terms of accuracy, the random forest in this study significantly exceeds the linear regression. However, the random forest has its own issue: overfitting. The OOB score of our model is 49.05%. Additionally, the average 10-fold cross-validation score is 41.24%.

Although the cross-validation score is not as perfect as the accuracy, this model has the highest cross-validation score in 270 potential models with different N_{try} and the number of remaining cases in the end leaf, respectively. Our model is the best in terms of accuracy, OOB score, and cross-validation score. It must be noted that the linear model, e.g., OLS, is more serious overfitting. **Figure 4** demonstrates the relationship between predicted and measured mental health scores. The slope of the fit line between predicted and measured mental health scores is lower than 1. The random forest rarely exactly predicts extreme values, such as 0 and 36. Putting another way, the random forest's prediction is closer to the mean value of the output variable. As shown in **Figure 1**, extreme values are rare, so the status of the random forest is acceptable.

Figure 5 demonstrates the importance of each feature. Emotions, including sadness, pleasure and smile, and self-reported health, affect mental health most. For example, if we do not employ the feature, "Sadness", in the model, the accuracy will decrease by 30.10%. The income and land cover in respondents' living environment significantly influence mental health. The accuracy will slash by 6.15% without the DIG in the model. Moreover, the importance values of grassland, forest, cropland, shrubland, water, urban land, wetland, and bare land, are equal to 4.22%, 4.25%, 4.55%, 3.58%, 3.05%, 3.94%, 4.14%, and 3.25% reduction in the accuracy, respectively.

Figure 6 illustrates nine maps of spatially average SHAP value of income (DIG) and land cover features. In order to make the SHAP spatial distribution readable, we use a spatially average value. We mean all the values in each cell with a 2.5-arc-degree side length. The observation numbers in each cell are different. **Figure 6.a** displays the spatially average SHAP value of DIG. In most areas, current income, DIG, negatively contributes to mental health. A lower DIG value is the main reason for negative contributions. Previous studies indicate that increased income improves human self-evaluation and emotional well-being, although some note there is a threshold for further improvement (Jebb et al. 2018; Killingsworth 2021). To most people, generally, mental health could benefit from increased income. Based on **Figure 6.a** and the current status of DIG (**Appendix Figure A1.a**), income positively affects mental health. However, it must be emphasized that the SHAP value of the land cover feature represents the attitudes toward current feature values. **Figures 6.b-i** demonstrate the SHAP values of the land cover feature. The observation's low mental health score due to land cover in their living environment might be various. The living environment with too more or too less a certain land type might negatively impact mental health status. For example, in terms of urban land feature, too high urban land percentage means a monotone scene of the living environment, but too low value indicates a totally rural area without convenient urban services. In other words, based on the SHAP values, we can only judge whether the current land cover status positively impacts mental health, but never know that the negative status is due to insufficiency and overplus here. Referring to the current urban land in the living environment (**Appendix Figure A1.i**), the urban land's contribution to good mental health is marginal (**Figure 6.i**), even harmful, when the living environment has more than 50% urban land. As for other land types, the maps of spatially average SHAP values indicate where the specific land type is suitable.

The connection between the current land feature value and its SHAP value is desired since the SHAP value cannot inform us that increasing or decreasing specific features would improve mental health. **Figure 7** demonstrates nine maps of spatially average local coefficient of DIG value and land cover features on mental health, based on **Equation 6**. If a local dataset's coefficient is insignificant, it would be set to zero. According to **Figure 7.a**, in most zones, a high DIG value is associated with a high SHAP value. The coefficients of income on mental health are either positive or insignificant. An increase in individual income would enhance its contribution to mental health. Putting another way, the rise in income improves mental health. **Figure 7.b** shows that people in metropolitan areas, including New York, Paris, London, Shanghai, Hong Kong, Tokyo, among others, prefer to have more cropland. Additionally, the positive coefficient on the Eastern coast of Australia is obvious. The reason for people's preference is due to scarcity value. Less cropland is in their living environment compared with people in other places (**Appendix Figure A1**). Other land features also have this attribute, as shown in **Figures 7.c-i**. The local coefficients of certain land types are larger if this land type is scarce.

Figure 8 illustrates the spatially average monetary values of eight land types, according to **Equation 7**. As shown in **Figure 8.a**, the monetary values of cropland are higher in metropolitan areas such as New York, London, Paris, Hong Kong, Shanghai, Tokyo, among others. Forest's and water's monetary values are also higher in the big cities. Although grassland's monetary values are positive, they are relatively lower than other natural land types. In most places, their monetary values are favorable due to wetland's and shrubland's scarcity values. It must be noted that wetland, shrubland, bare land are very rare in most living environments (as shown in **Appendix Figure A1**). A slight increase in wetland, shrubland, or bare land is difficult. This is the reason for their extraordinary monetary value, which is consistent with previous studies (Costanza et al. 2014). Moreover, **Figure 8.i** demonstrates the monetary values of urban land, but more areas are insignificant. Referring to the current status and coefficients of urban land and DIG, we find the coefficients of urban land and DIG are unmatched. The people who prefer to have more urban land are less impacted by income, while people highly affected by income care less about the percentage of urban land in their cities.

Discussion

Our main findings are that mental health and land cover relationships are non-linear and various. Increases in each land type positively impact mental health when the percentages of these land types are low. Accordingly, it could be implied that people who prefer to live in environments with high diversity and extremely monolithic landscapes might lead to poor mental health. Furthermore, it is the first study that uses SHAP and random forest to grasp the relationship between land cover and mental health. To make the results understandable, we employ geographically local technology to connect the current land cover status to its SHAP values. Based on the links between SHAP value and current status, the monetary values of land cover are estimated, although the numbers of significant monetary values of land cover are limited. Our results show that a slight increase in shrubland, wetland, and bare land in most regions could improve people's mental health. Cropland, forest, and water are mainly desired in metropolitan areas and places with too less cropland, forest, and water. It is worth noting urban land is also desired due to relatively crowded environments in metropolitan areas. Moreover, the model's accuracy is relatively high, indicating the reliability of the results. The accuracy, RMSE, MSE, and MEA are 93.09%, 1.66, 2.74, and 1.24, respectively, exceeding most previous studies.

Previous studies focus more on the impacts of green space on human well-being or mental health in the city (Alcock et al. 2015; Alcock et al. 2014; Krekel et al. 2016; Tsurumi et al. 2018; White et al. 2013). The coverage percentage of green space positively affects mental health (Alcock et al. 2015; Alcock et al. 2014; White et al. 2013). In our study, all land types are positively related to mental health when their percentages are low. These results are consistent with previous studies because they mainly concentrate on the urban area with less coverage of other land types. The impact of grassland is not so high as other land types, which is counterintuitive. Two reasons cause this problem. Firstly, this research is based on remote sensing data. In the remote sensing process, grassland is easier to be misclassified especially when close to cropland and shrubland (Gong et al. 2019). Especially, sporadic grass is more likely to be misclassified, so the low accuracy of grassland in urban areas might mislead the model's results. Secondly, grassland is for grazing rather than improving mental health in rural areas. Therefore, the grassland's effects are not so high. Wetland is the most preferred, as it provides the most ecosystem service (Costanza et al. 1997; Costanza et al. 2014), and it is scarce in the living environment. Bare land's average SHAP values and monetary values are high. According to the figure in the data provider's article (Gong et al. 2019), the large area of bare land is generally desert, while it might be sports play yards in the cities. Shrubland is similar to wetland and bare land, and it positively impacts mental health when they are scarce. Forest and cropland's effects are various. In metropolitan areas, increased cropland and forest percentage would improve mental health. People cannot enter the large aggregated forest to have various nature experiences, and they are also associated with the possibility of crime (Lee and Maheswaran 2011; Markevych et al. 2017). The high percentage of urban land is associated with mental health. Living in cities naturally is necessary (Douglas et al. 2017; Hartig and Kahn 2016; White et al. 2017). However, the adverse effects of large amounts of non-urban land types on mental health indicate that people living in rural areas are likely to have mental disorders and need more assistants. Therefore, in the land use, the percentage of urban land should be carefully treated and balanced.

There are several limitations and issues worthy of note. First, the land cover variables are the percentages of eight land types in the buffers with a 5-km radius surrounding the living locations of respondents. There is an assumption that the quality of land cover does not influence the effects of those land types on mental health. For example, there may be no difference between a well-designed urban park and grassland in the pasture. Furthermore, the impacts of the distance to a certain land type are ignored. Secondly, this study only uses global cross-sectional data, so it cannot detect the difference within people when land cover changes. Global research using panel data to probe the effects within individuals is still desired. Finally, the number of respondents in each country is not the same or proportional to the country's population. The countries with more respondents have more substantial impacts on the results. Thus, the results might be prejudiced, though this database is one of the biggest databases in this field. In future studies, the long panel data should be used to investigate the impacts of land cover within individuals. Moreover, because the current computing ability is insufficient, the calculations of SHAP use relatively fewer random paths, which might reduce the accuracy of Shapley values. Effective explaining methods and tools should be developed to make the machine learning results understandable.

Conclusion

The relationships between land cover in living environments and mental health are more complex than linear assumptions. An unsuitable increase in a specific land type might not improve residents' mental health. Among eight land types, shrubland, wetland, and bare land have the highest effects on mental health due to their scarcity in living environments. Cropland's, forest's, and water's impacts are high, mainly in metropolitan areas. The impacts of urban land and grassland increases are tiny compared with other land types. Our study illustrates the heterogeneity of the effects of eight land types on mental health to provide more information for governments and the public. This critical knowledge helps draft sustainable land-use policies to improve mental health.

Data Availability

The fully reproducible codes are publicly available at https://github.com/MichaelChaoLi-cpu/MentalHealthAndLandCover . Data are available from the corresponding author on reasonable request.

Acknowledgment

This research was supported by the following funding agencies: JSPS KAKENHI (Grant No. JP20H00648), the Environment Research and Technology Development Fund of the Environmental Restoration and Conservation Agency of Japan (Grant No. JPMEERF20201001), and also JST SPRING (Grant No. JPMJSP2136).

Figure:

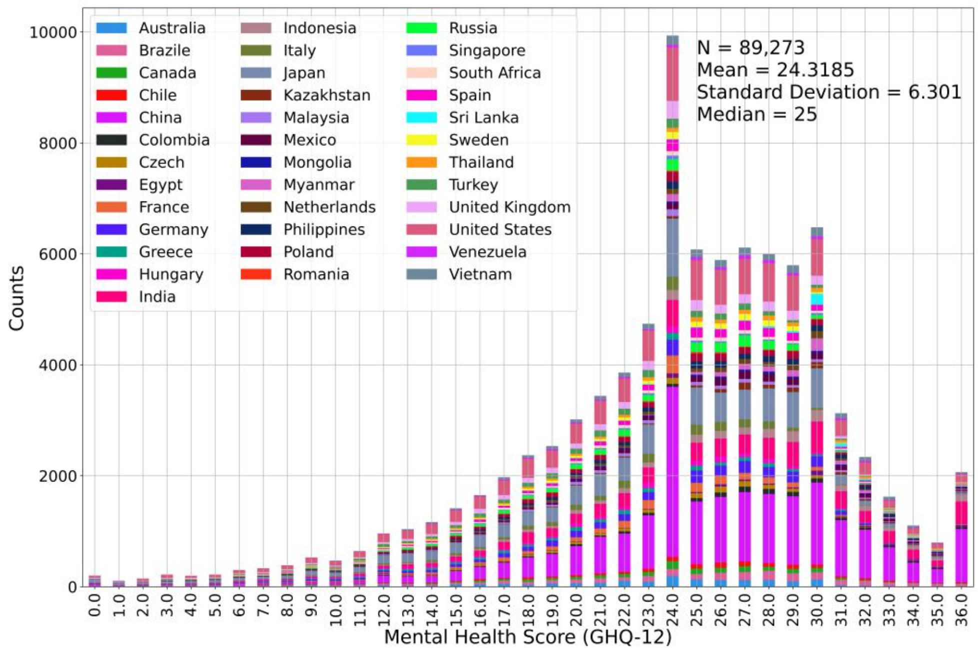

Figure 1: The Statistical Distribution of Mental Health Assessment

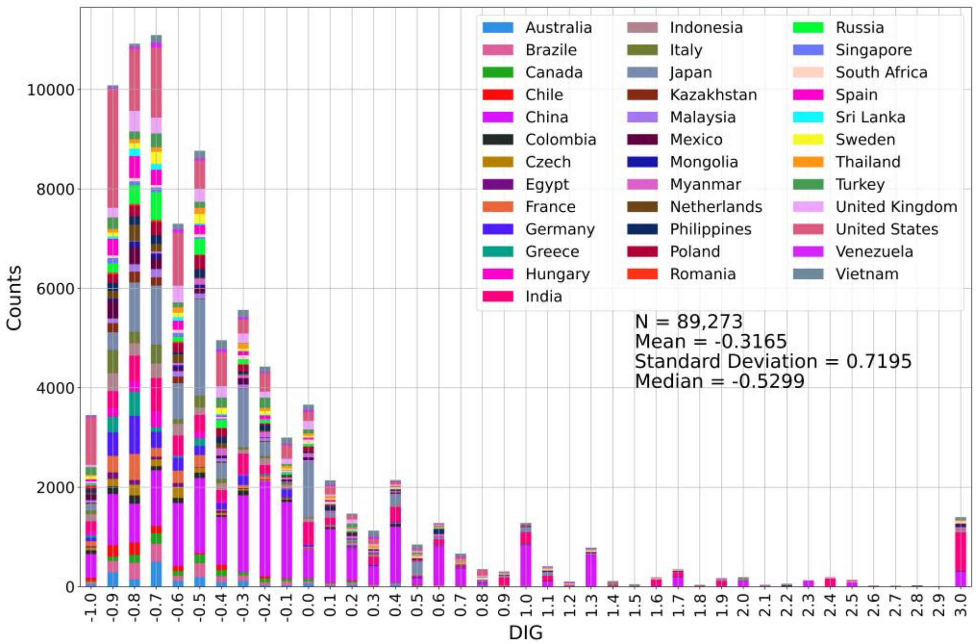

Figure 2: The Statistical Distribution of DIG

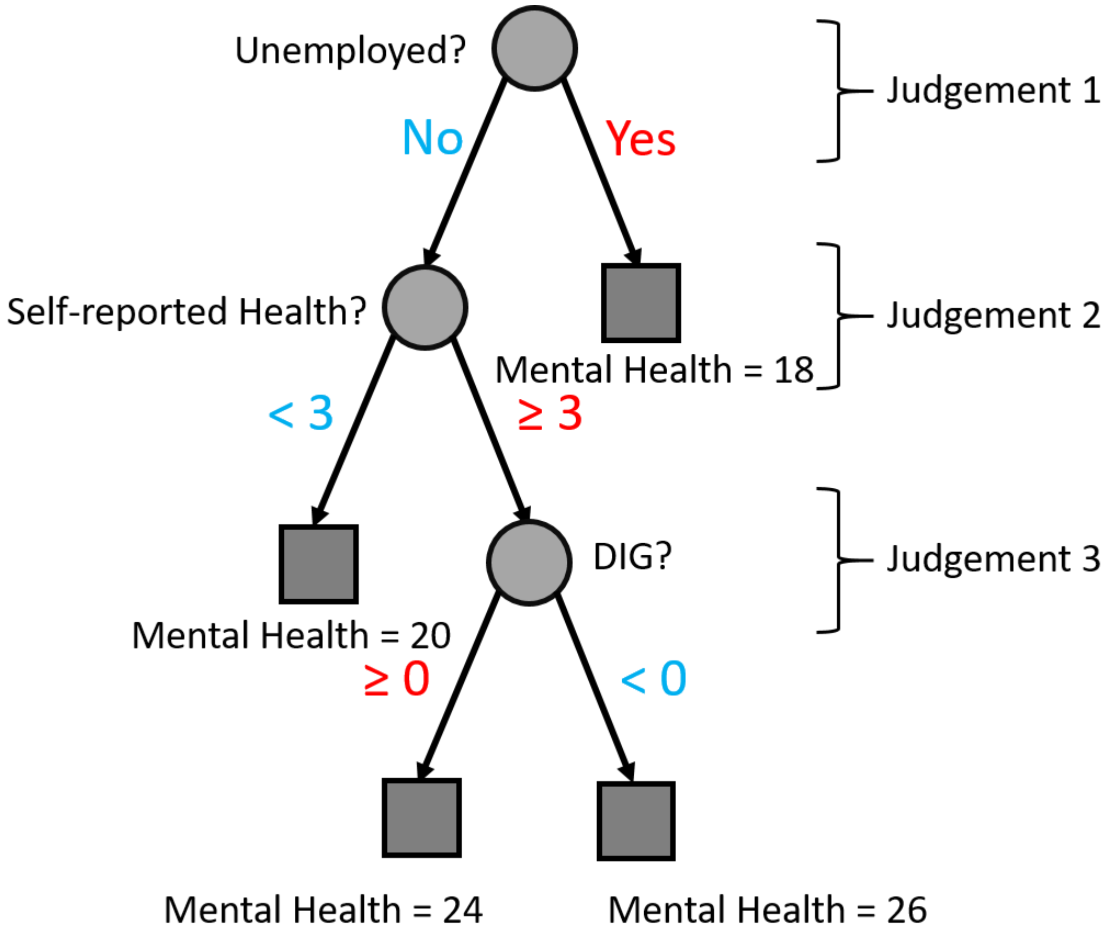

Figure 3: Example of a Regression Decision Tree

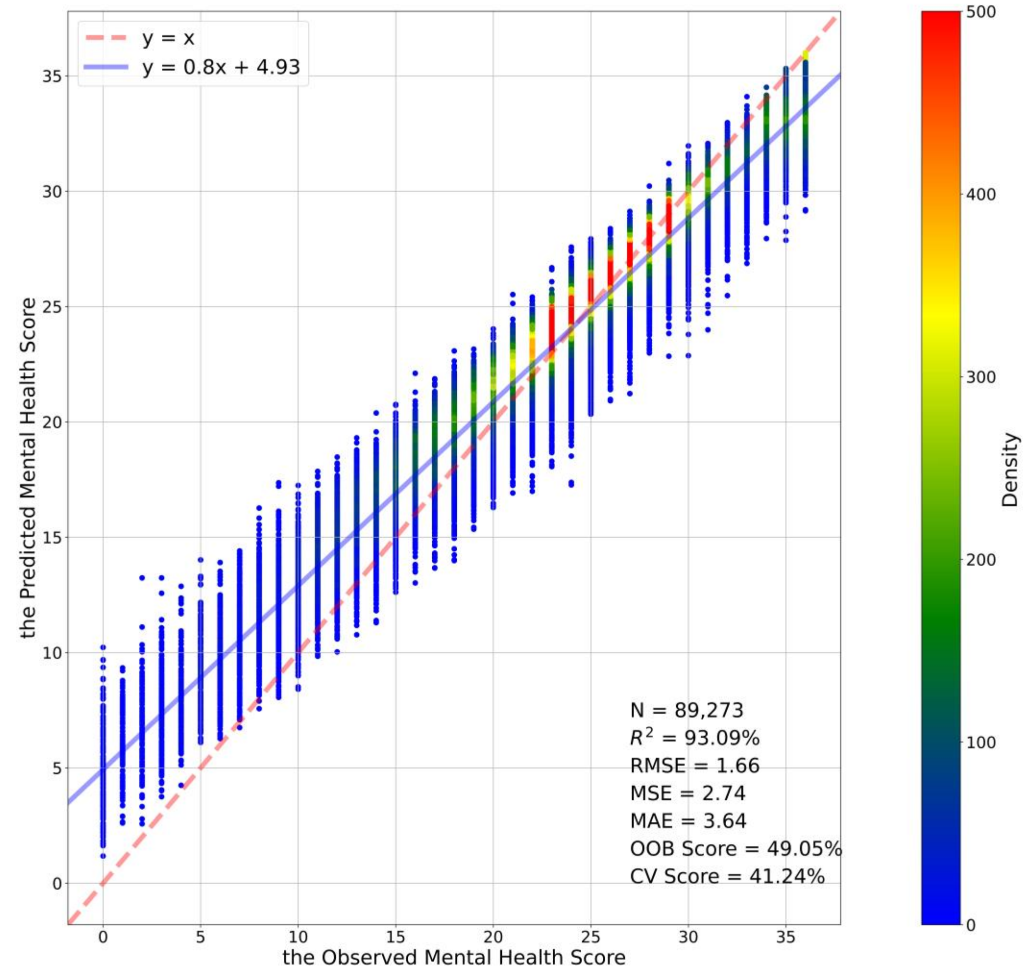

Figure 4: The Density Plots between the Measured and Predicted Mental Health Score

(The red dashed line is the 1:1 line. The blue line is the regression line.)

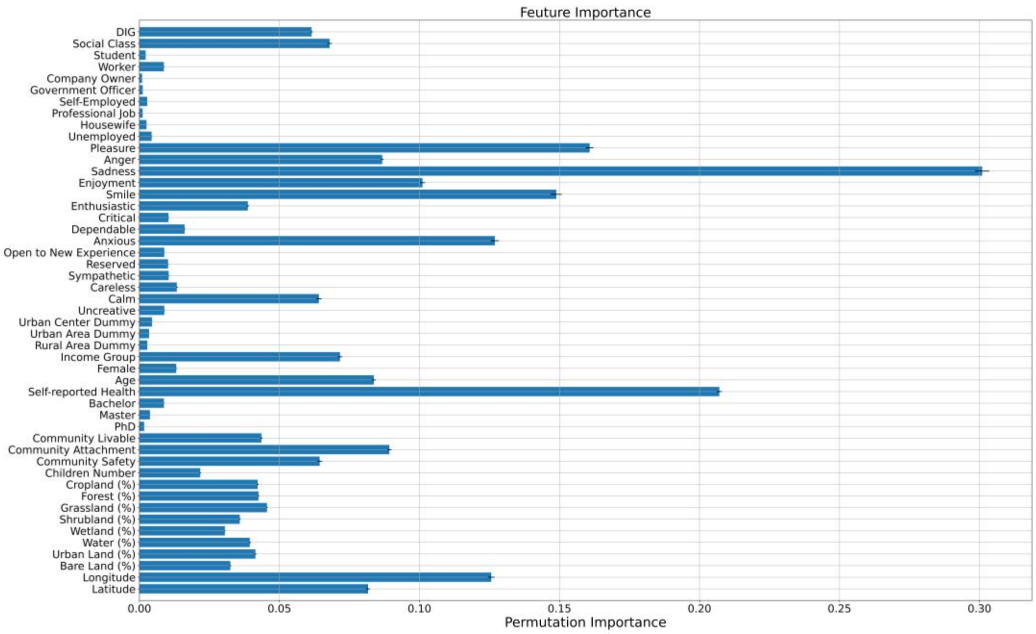

Figure 5: Feature Importance

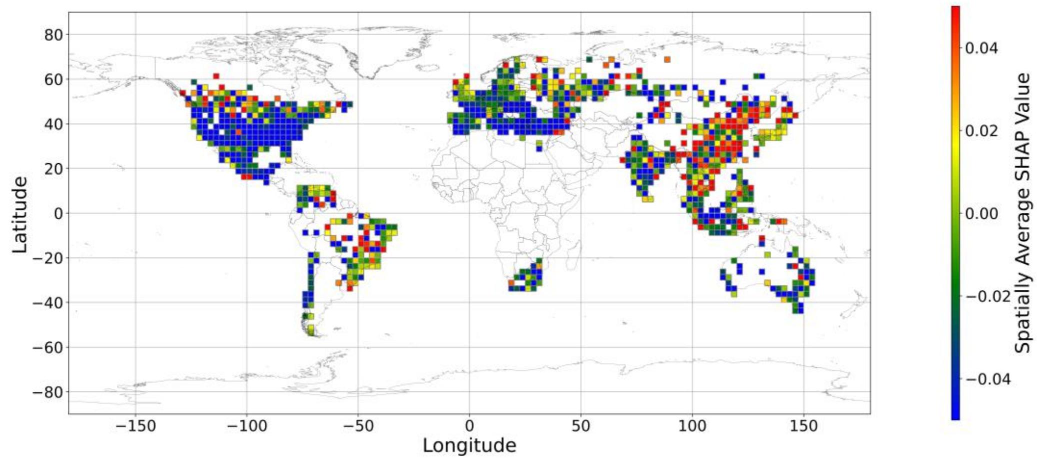

Figure 6.a: The Spatially Average SHAP Values of Income

(Note: Cell size is 2.5° × 2.5°)

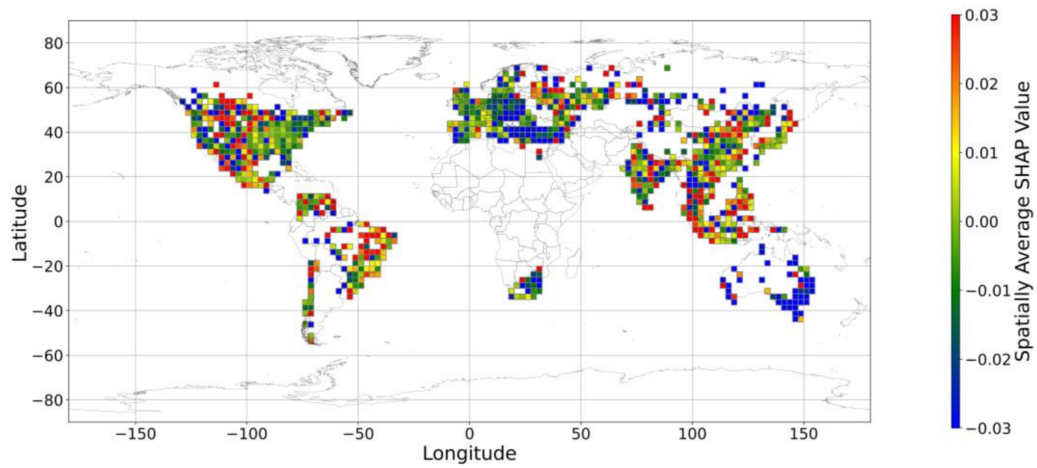

Figure 6.b: The Spatially Average SHAP Values of Cropland

(Note: Cell size is 2.5° × 2.5°)

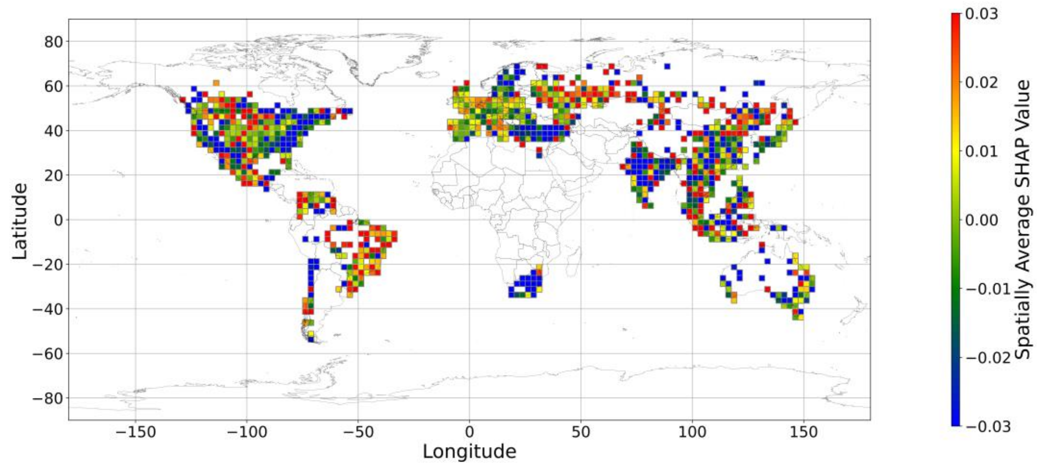

Figure 6.c: The Spatially Average SHAP Values of Forest

(Note: Cell size is 2.5° × 2.5°)

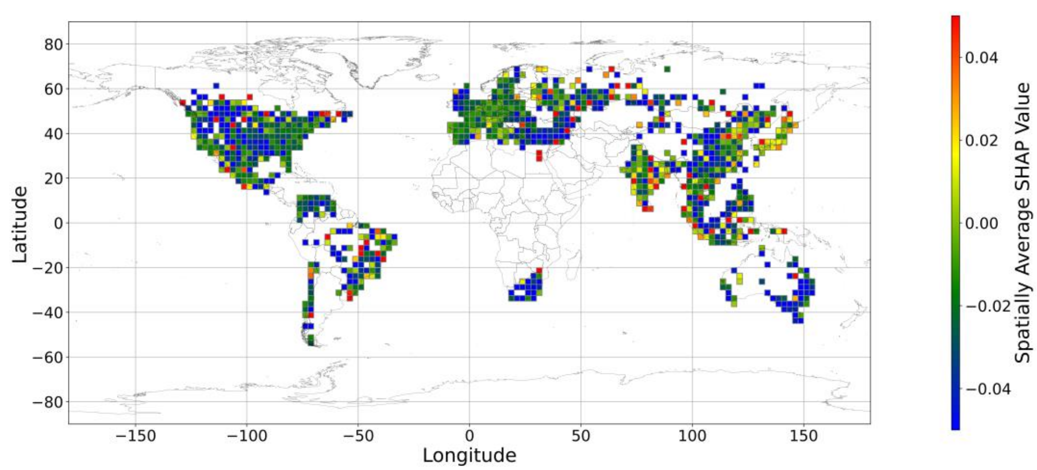

Figure 6.d: The Spatially Average SHAP Values of Grassland

(Note: Cell size is 2.5° × 2.5°)

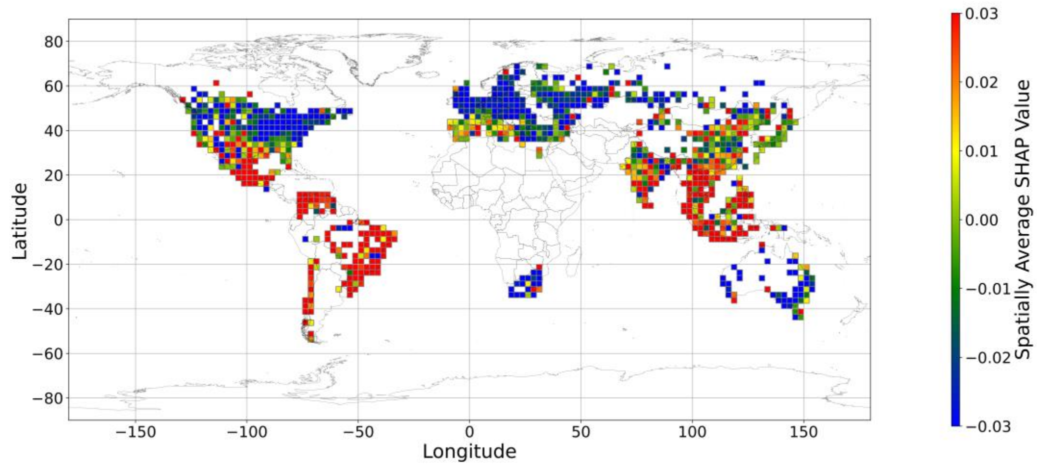

Figure 6.e: The Spatially Average SHAP Values of Shrubland

(Note: Cell size is 2.5° × 2.5°)

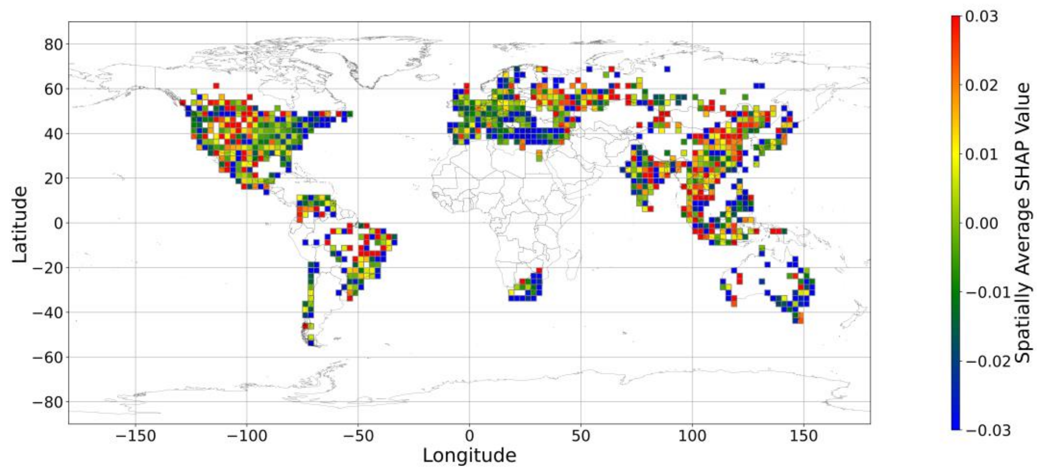

Figure 6.f: The Spatially Average SHAP Values of Water

(Note: Cell size is 2.5° × 2.5°)

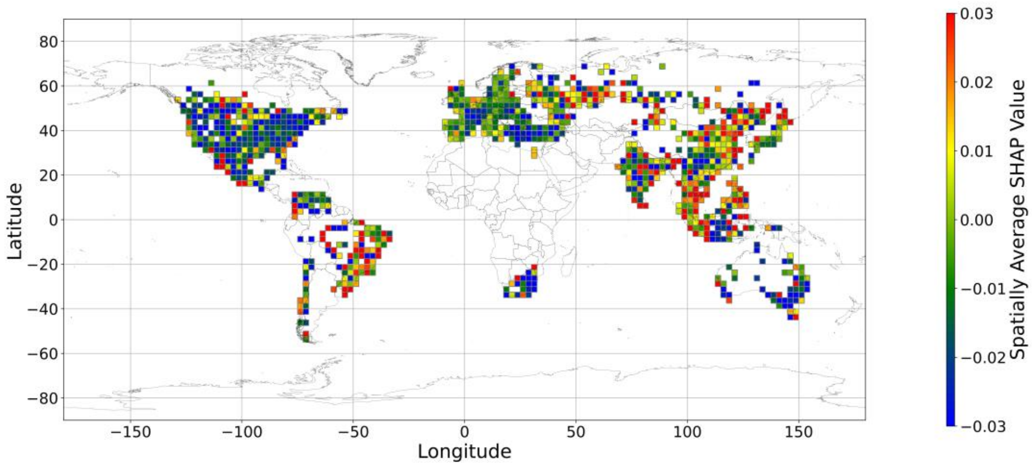

Figure 6.g: The Spatially Average SHAP Values of Wetland

(Note: Cell size is 2.5° × 2.5°)

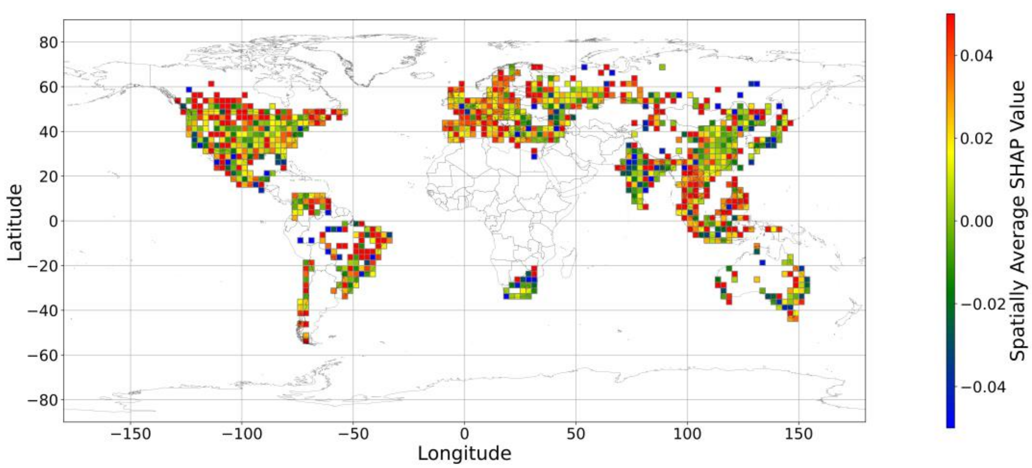

Figure 6.h: The Spatially Average SHAP Values of Urban Land

(Note: Cell size is 2.5° × 2.5°)

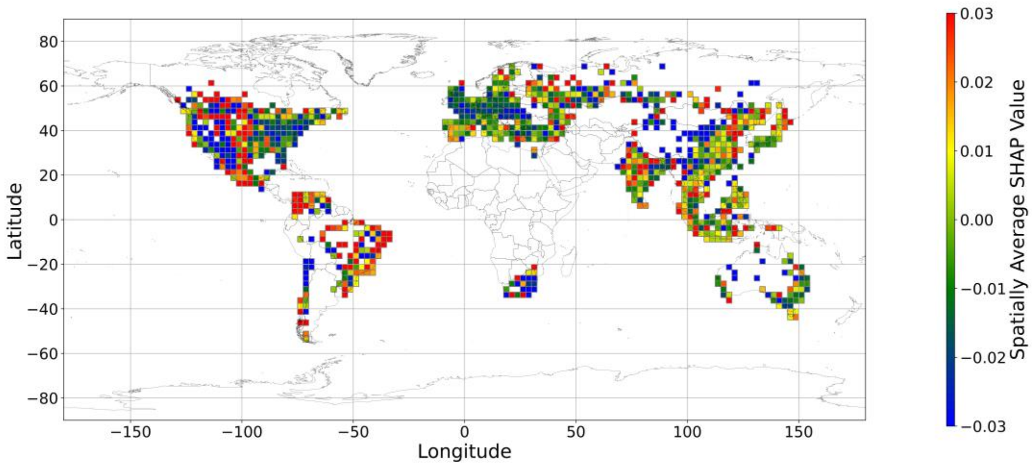

Figure 6.i: The Spatially Average SHAP Values of Bare Land

(Note: Cell size is 2.5° × 2.5°)

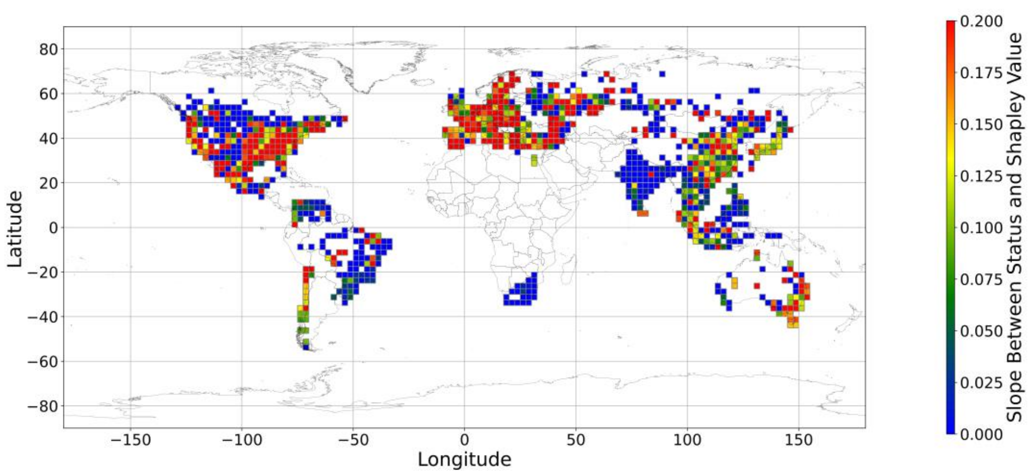

Figure 7.a: The Spatially Average Coefficients of Income

(Note: Cell size is 2.5° × 2.5°)

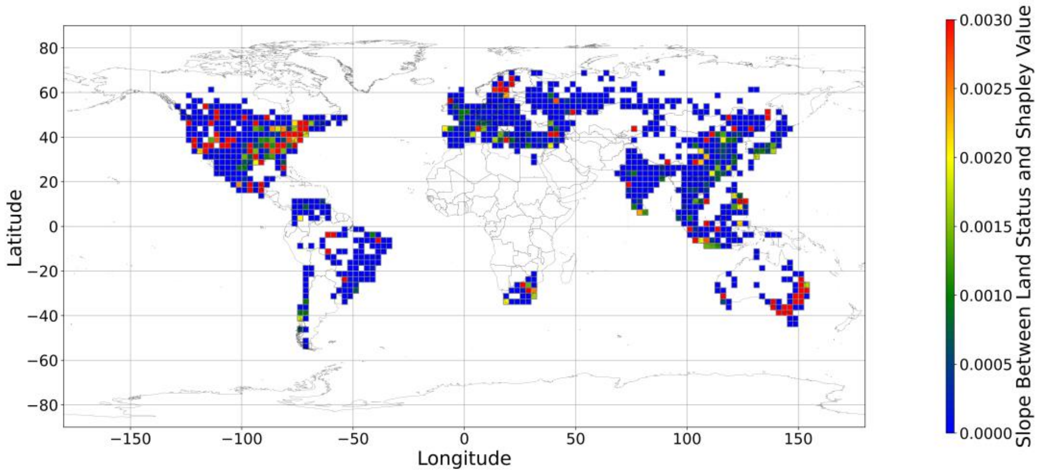

Figure 7.b: The Spatially Average Coefficients of Cropland

(Note: Cell size is 2.5° × 2.5°)

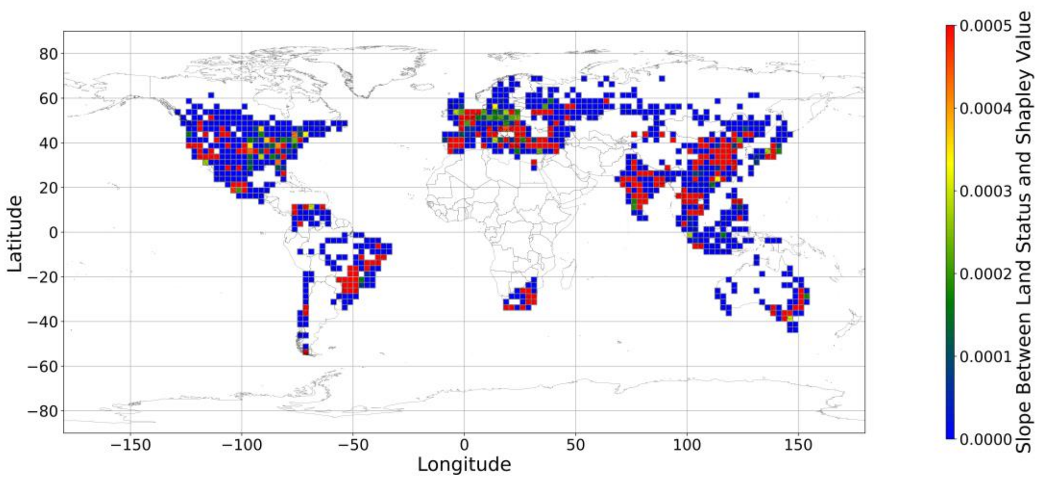

Figure 7.c: The Spatially Average Coefficients of Forest

(Note: Cell size is 2.5° × 2.5°)

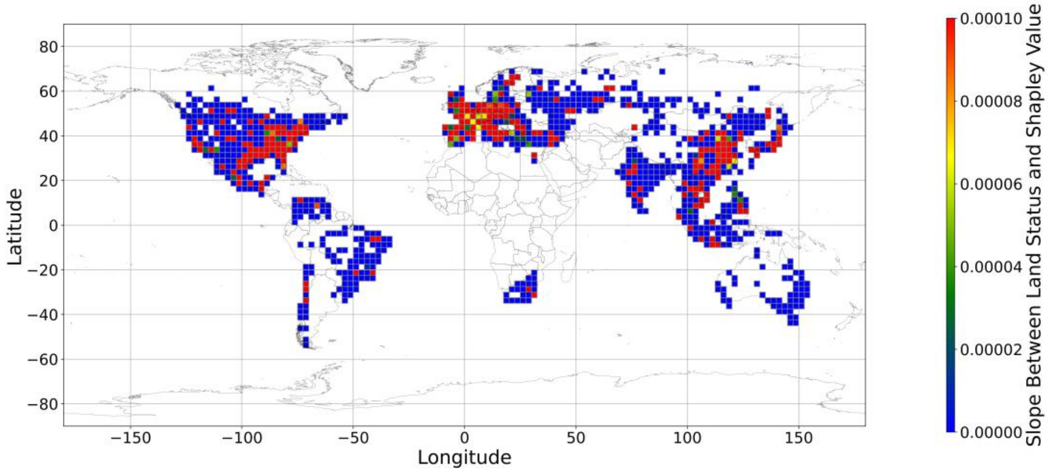

Figure 7.d: The Spatially Average Coefficients of Grassland

(Note: Cell size is 2.5° × 2.5°)

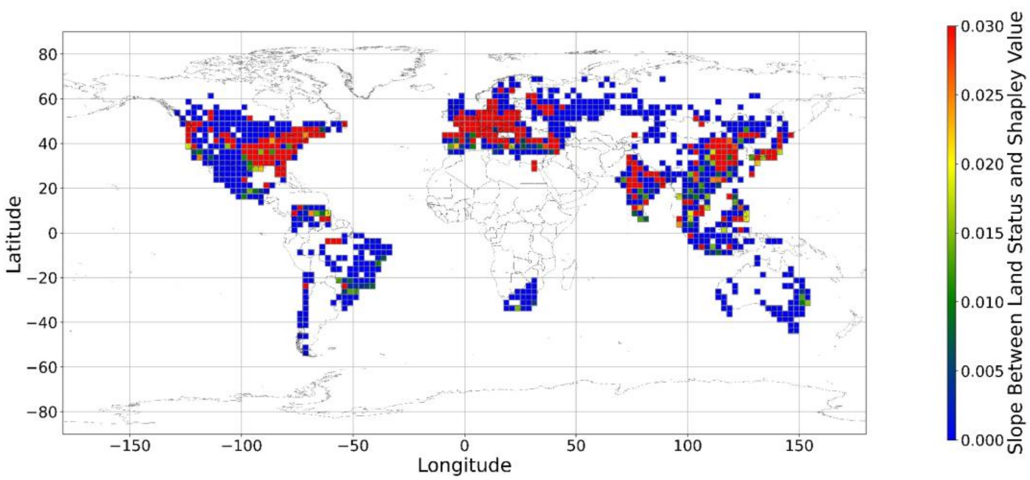

Figure 7.e: The Spatially Average Coefficients of Shrubland

(Note: Cell size is 2.5° × 2.5°)

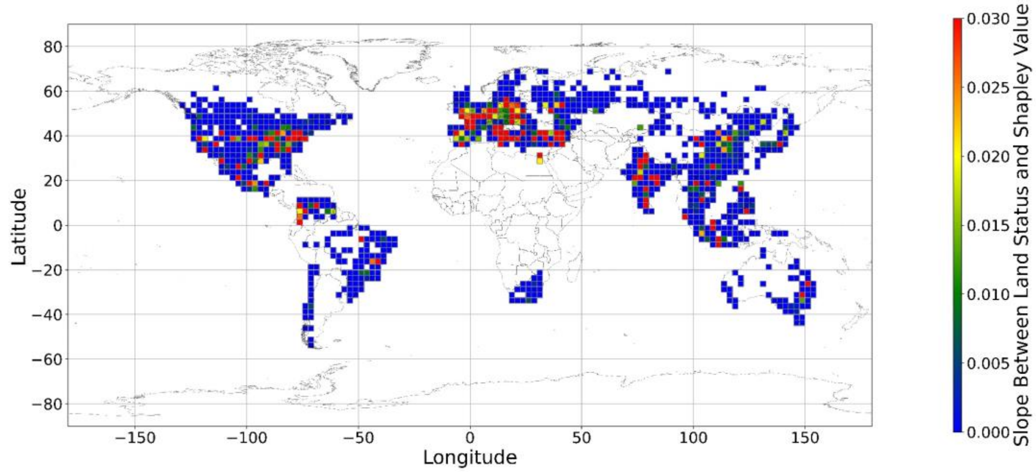

Figure 7.f: The Spatially Average Coefficients of Water

(Note: Cell size is 2.5° × 2.5°)

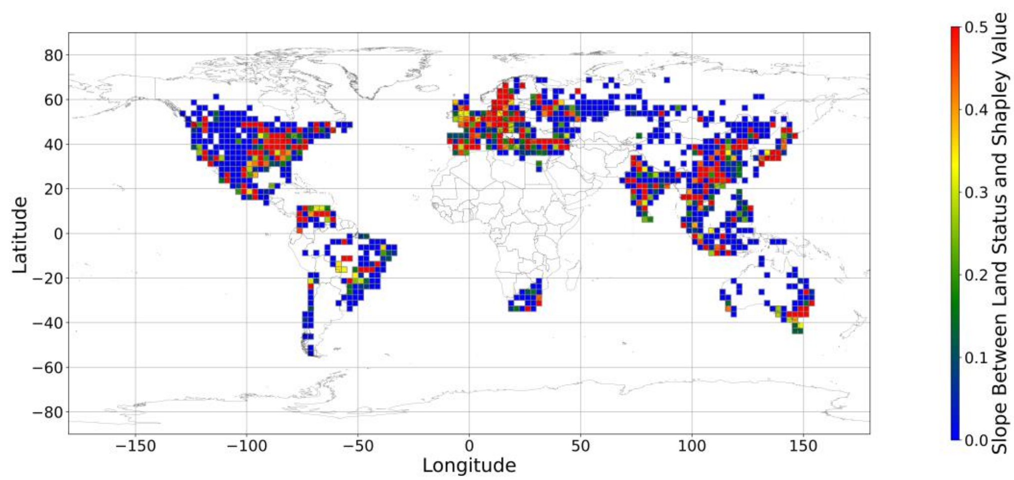

Figure 7.g: The Spatially Average Coefficients of Wetland

(Note: Cell size is 2.5° × 2.5°)

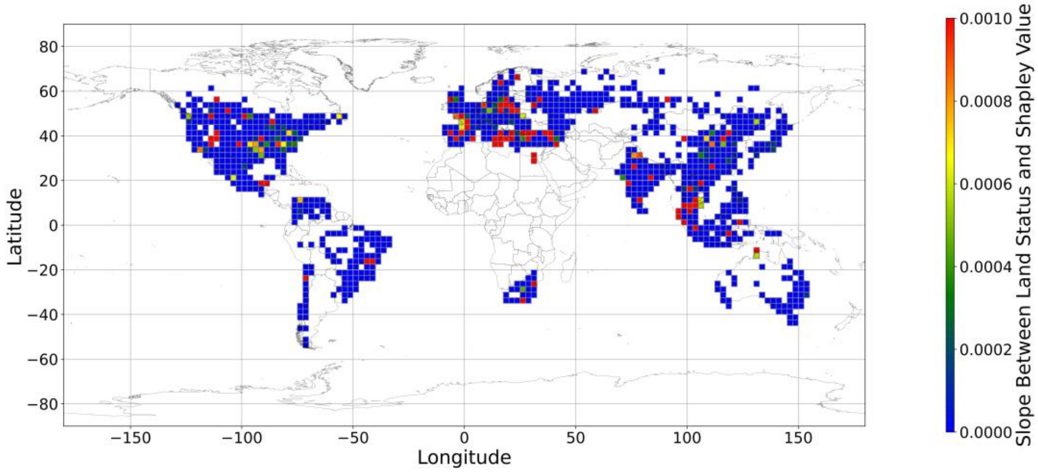

Figure 7.i: The Spatially Average Coefficients of Urban Land

(Note: Cell size is 2.5° × 2.5°)

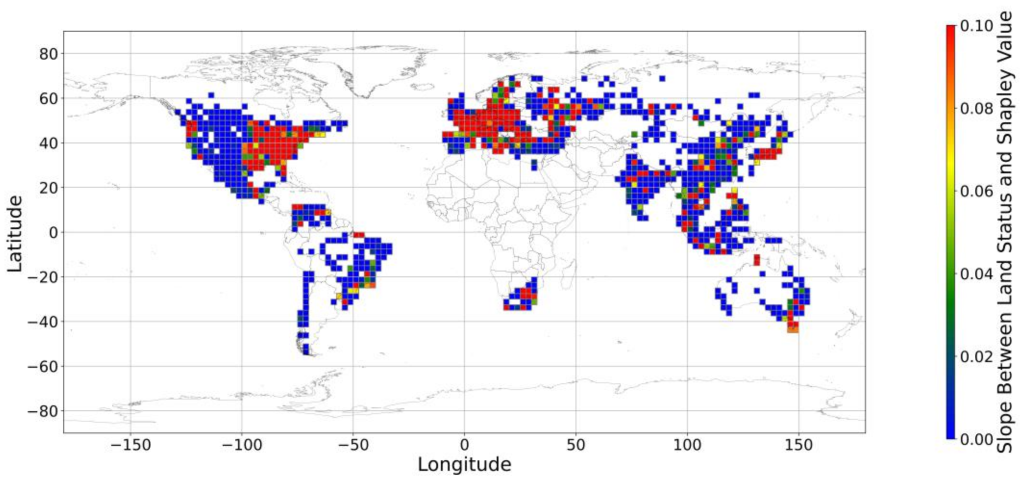

Figure 7.i: The Spatially Average Coefficients of Bare Land

(Note: Cell size is 2.5° × 2.5°)

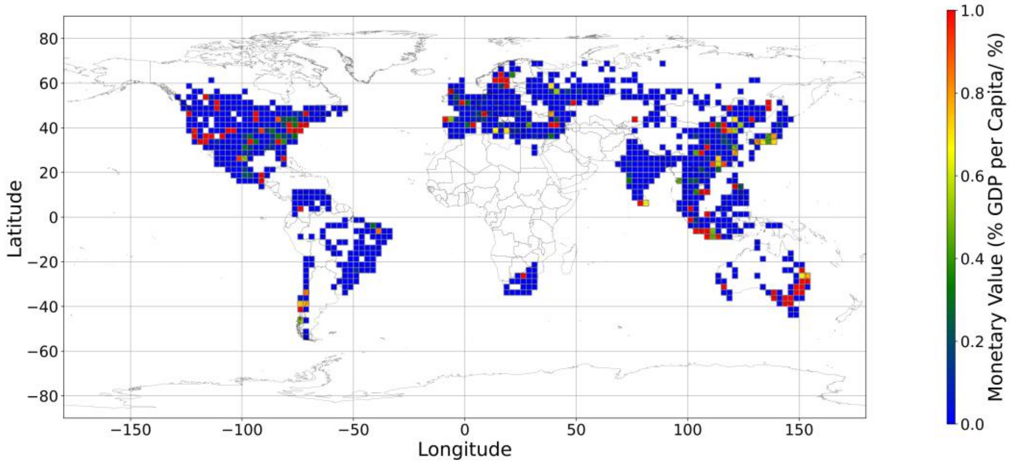

Figure 8.a: The Spatially Average Monetary Value of Cropland

(Note: Cell size is 2.5° × 2.5°)

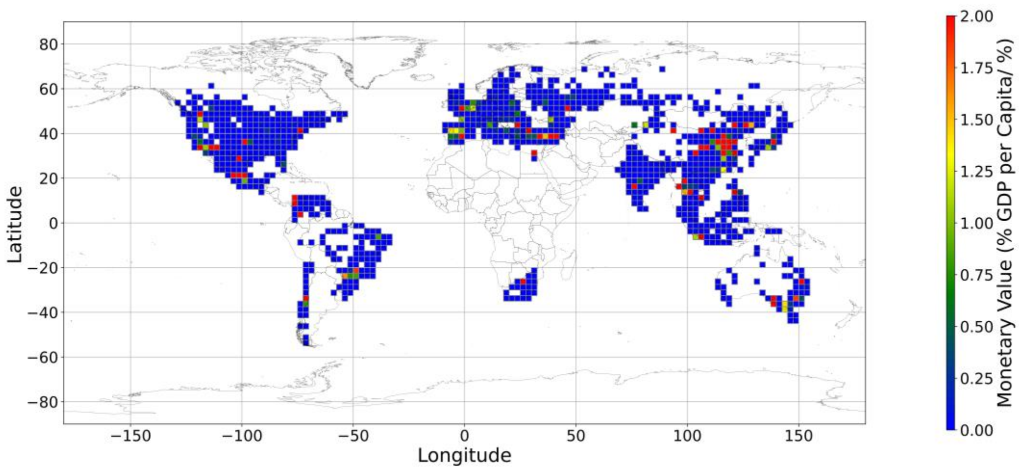

Figure 8.b: The Spatially Average Monetary Value of Forest

(Note: Cell size is 2.5° × 2.5°)

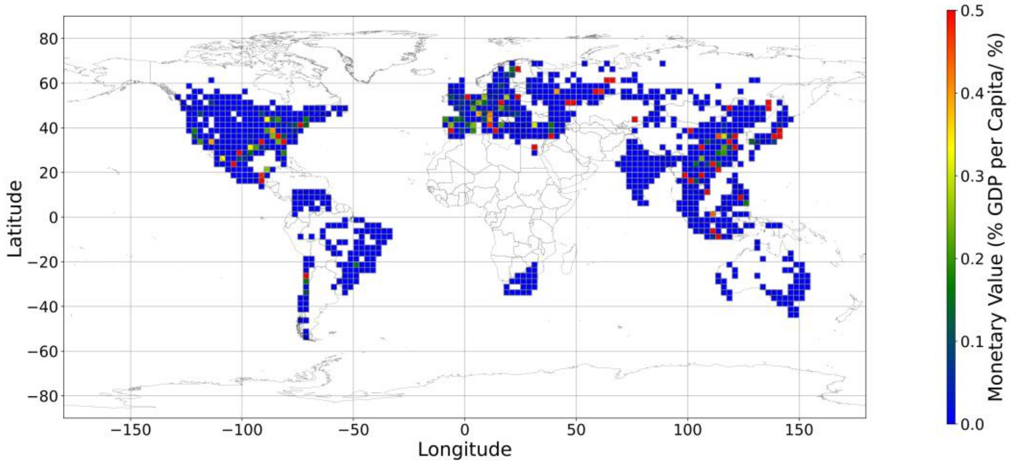

Figure 8.c: The Spatially Average Monetary Value of Grassland

(Note: Cell size is 2.5° × 2.5°)

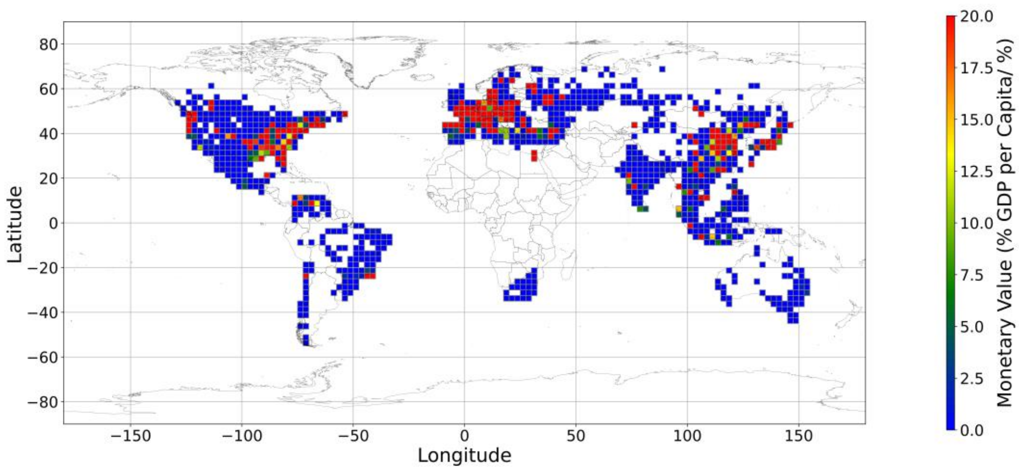

Figure 8.d: The Spatially Average Monetary Value of Shrubland

(Note: Cell size is 2.5° × 2.5°)

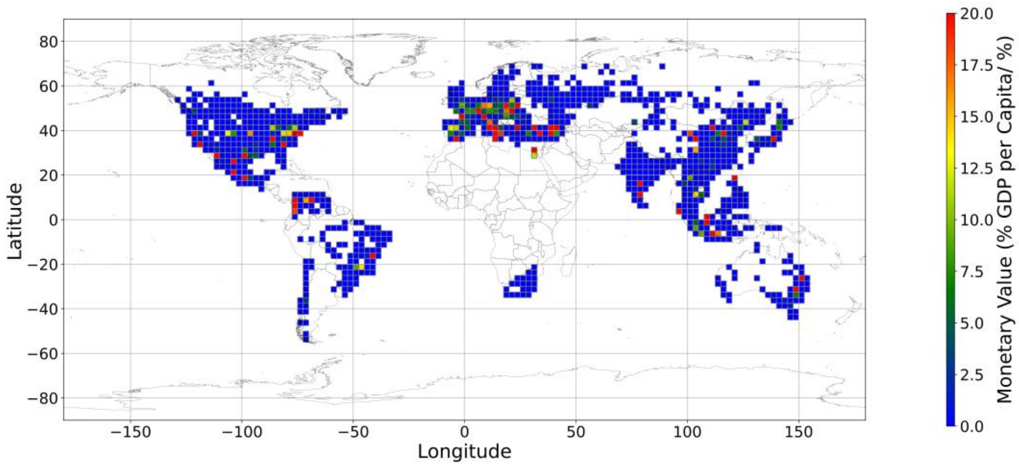

Figure 8.e: The Spatially Average Monetary Value of Water

(Note: Cell size is 2.5° × 2.5°)

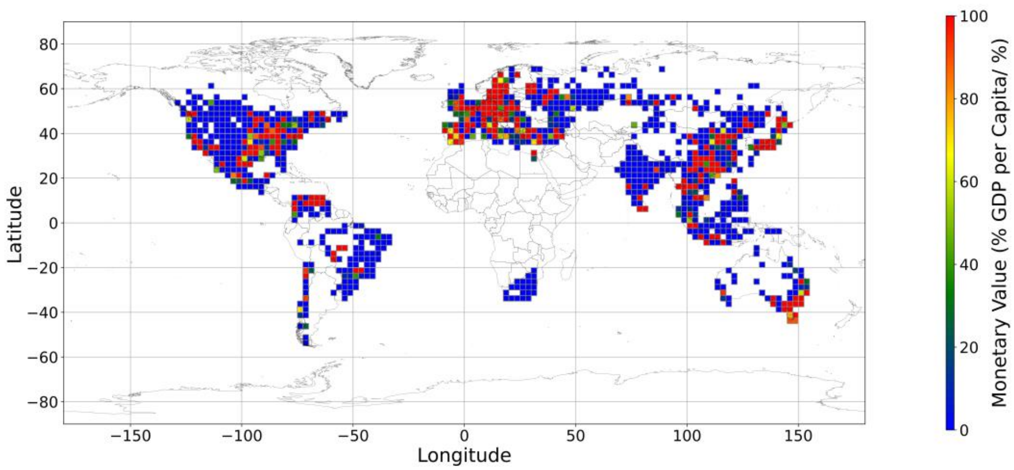

Figure 8.f: The Spatially Average Monetary Value of Wetland

(Note: Cell size is 2.5° × 2.5°)

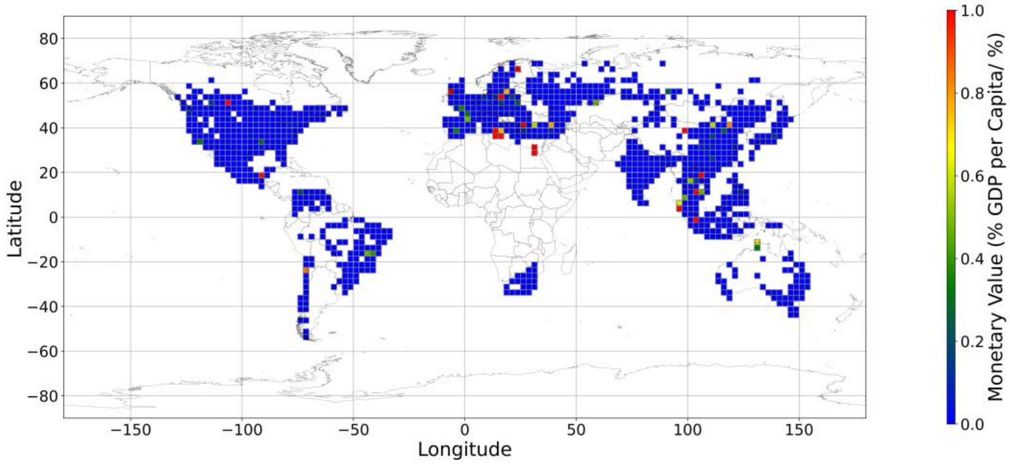

Figure 8.g: The Spatially Average Monetary Value of Urban Land

(Note: Cell size is 2.5° × 2.5°)

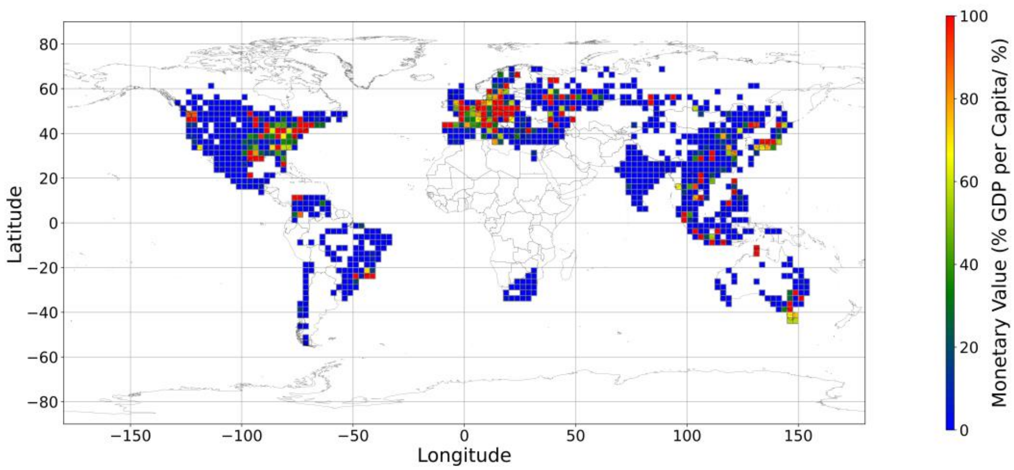

Figure 8.h: The Spatially Average Monetary Value of Bare Land

(Note: Cell size is 2.5° × 2.5°)

Reference:

Akpinar, A., Barbosa-Leiker, C., & Brooks, K.R. (2016). Does green space matter? Exploring relationships between green space type and health indicators. *Urban Forestry & Urban Greening*, *20*, 407-418

Alcock, I., White, M.P., Lovell, R., Higgins, S.L., Osborne, N.J., Husk, K., & Wheeler, B.W. (2015). What accounts for 'England's green and pleasant land'? A panel data analysis of mental health and land cover types in rural England. *Landscape and Urban Planning*, *142*, 38-46

Alcock, I., White, M.P., Wheeler, B.W., Fleming, L.E., & Depledge, M.H. (2014). Longitudinal Effects on Mental Health of Moving to Greener and Less Green Urban Areas. *Environmental Science & Technology*, *48*, 1247-1255

Berman, M.G., Kross, E., Krpan, K.M., Askren, M.K., Burson, A., Deldin, P.J., Kaplan, S., Sherdell, L., Gotlib, I.H., & Jonides, J. (2012). Interacting with nature improves cognition and affect for individuals with depression. *Journal of Affective Disorders*, *140*, 300-305

Bratman, G.N., Anderson, C.B., Berman, M.G., Cochran, B., de Vries, S., Flanders, J., Folke, C., Frumkin, H., Gross, J.J., Hartig, T., Kahn, P.H., Kuo, M., Lawler, J.J., Levin, P.S., Lindahl, T., Meyer-Lindenberg, A., Mitchell, R., Ouyang, Z.Y., Roe, J., Scarlett, L., Smith, J.R., van den Bosch, M., Wheeler, B.W., White, M.P., Zheng, H., & Daily, G.C. (2019). Nature and mental health: An ecosystem service perspective. *Science Advances*, *5*, eaax0903

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5-32

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (2017). Classification And Regression Trees

Carpenter, S.R., Mooney, H.A., Agard, J., Capistrano, D., Defries, R.S., Diaz, S., Dietz, T., Duraiappah, A.K., Oteng-Yeboah, A., Pereira, H.M., Perrings, C., Reid, W.V., Sarukhan, J., Scholes, R.J., & Whyte, A. (2009). Science for managing ecosystem services: Beyond the Millennium Ecosystem Assessment. *Proceedings of the National Academy of Sciences*, *106*, 1305-1312

Chameides, W.L., Lindsay, R.W., Richardson, J., & Kiang, C.S. (1988). The role of biogenic hydrocarbons in urban photochemical smog: Atlanta as a case study. *Science*, *241*, 1473-1475

Chaplin-Kramer, R., Sharp, R.P., Weil, C., Bennett, E.M., Pascual, U., Arkema, K.K., Brauman, K.A., Bryant, B.P., Guerry, A.D., Haddad, N.M., Hamann, M., Hamel, P., Johnson, J.A., Mandle, L., Pereira, H.M., Polasky, S., Ruckelshaus, M., Shaw, M.R., Silver, J.M., Vogl, A.L., & Daily, G.C. (2019). Global modeling of nature’s contributions to people. *Science*, *366*, 255-258

Costanza, R., D'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., & Van Den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*, 253-260

Costanza, R., De Groot, R., Sutton, P., Van Der Ploeg, S., Anderson, S.J., Kubiszewski, I., Farber, S., & Turner, R.K. (2014). Changes in the global value of ecosystem services. *Global Environmental Change*, *26*, 152-158

Czajkowski, M., & Kretowski, M. (2016). The role of decision tree representation in regression problems – An evolutionary perspective. *Applied Soft Computing*, *48*, 458-475

Diaz, S., Pascual, U., Stenseke, M., Martin-Lopez, B., Watson, R.T., Molnar, Z., Hill, R., Chan, K.M.A., Baste, I.A., Brauman, K.A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P.W., van Oudenhoven, A.P.E., van der Plaat, F., Schroter, M., Lavorel, S., Aumeeruddy-Thomas, Y., Bukvareva, E., Davies, K., Demissew, S., Erpul, G., Failler, P., Guerra, C.A., Hewitt, C.L., Keune, H., Lindley, S., & Shirayama, Y. (2018). Assessing nature's contributions to people. *Science, 359*, 270-272

Diaz, S., Settele, J., Brondizio, E.S., Ngo, H.T., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Ichii, K., Liu, J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnar, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.J., Visseren-Hamakers, I., Willis, K.J., & Zayas, C.N. (2019). Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science, 366*, eaax3100

Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, *2*, 253-260

Douglas, O., Lennon, M., & Scott, M. (2017). Green space benefits for health and well-being: A life-course approach for urban planning, design and management. *Cities, 66*, 53-62

Dzhambov, A.M., Browning, M.H.E.M., Markevych, I., Hartig, T., & Lercher, P. (2020). Analytical approaches to testing pathways linking greenspace to health: A scoping review of the empirical literature. *Environmental Research, 186*, 109613

Dzhambov, A.M., Markevych, I., Hartig, T., Tilov, B., Arabadzhiev, Z., Stoyanov, D., Gatseva, P., & Dimitrova, D.D. (2018). Multiple pathways link urban green- and bluespace to mental health in young adults. *Environmental Research*, *166*, 223-233

Eitelberg, D.A., Van Vliet, J., Doelman, J.C., Stehfest, E., & Verburg, P.H. (2016). Demand for biodiversity protection and carbon storage as drivers of global land change scenarios*, 40*, 101-111

El-Metwally, A., Javed, S., Razzak, H.A., Aldossari, K.K., Aldiab, A., Al-Ghamdi, S.H., Househ, M., Shubair, M.M., & Al-Zahrani, J.M. (2018). The factor structure of the general health questionnaire (GHQ12) in Saudi Arabia. *BMC Health Services Research, 18*

Felipe-Lucia, M.R., Soliveres, S., Penone, C., Manning, P., van der Plas, F., Boch, S., Prati, D., Ammer, C., Schall, P., Gossner, M.M., Bauhus, J., Buscot, F., Blaser, S., Bluthgen, N., de Frutos, A., Ehbrecht, M., Frank, K., Goldmann, K., Hansel, F., Jung, K., Kahl, T., Nauss, T., Oelmann, Y., Pena, R., Polle, A., Renner, S., Schloter, M., Schoning, I., Schrumpf, M., Schulze, E.D., Solly, E., Sorkau, E., Stempfhuber, B., Tschapka, M., Weisser, W.W., Wubet, T., Fischer, M., & Allan, E. (2018). Multiple forest attributes underpin the supply of multiple ecosystem services. *Nature Communications, 9*, 4839

Frey, B.S., Luechinger, S., & Stutzer, A. (2010). The Life Satisfaction Approach to Environmental Valuation. In G.C. Rausser, V.K. Smith, & D. Zilberman (Eds.), *Annual Review of Resource Economics, Vol 2, 2010* (pp. 139-160). Palo Alto: Annual Reviews

Friedman, J.H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*, 1189-1232

Gong, P., Liu, H., Zhang, M.N., Li, C.C., Wang, J., Huang, H.B., Clinton, N., Ji, L.Y., Li, W.Y., Bai, Y.Q., Chen, B., Xu, B., Zhu, Z.L., Yuan, C., Suen, H.P., Guo, J., Xu, N., Li, W.J., Zhao, Y.Y., Yang, J., Yu, C.Q., Wang, X., Fu, H.H., Yu, L., Dronova, I., Hui, F.M., Cheng, X., Shi, X.L., Xiao, F.J., Liu, Q.F., & Song, L.C. (2019). Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Science Bulletin, 64*, 370-373

Greenwell, B.M. (2017). pdp: an R Package for constructing partial dependence plots. *R J., 9*, 421

Hankins, M. (2008). The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC Public Health, 8*, 355

Hartig, T., & Kahn, P.H. (2016). Living in cities, naturally. *Science, 352*, 938-940

Hartig, T., Mitchell, R., de Vries, S., & Frumkin, H. (2014). Nature and Health. In J.E. Fielding (Ed.), *Annual Review of Public Health, Vol 35* (pp. 207-+). Palo Alto: Annual Reviews

Jebb, A.T., Tay, L., Diener, E., & Oishi, S. (2018). Happiness, income satiation and turning points around the world. *Nature Human Behaviour, 2*, 33-38

Killingsworth, M.A. (2021). Experienced well-being rises with income, even above $75,000 per year. *Proceedings of the National Academy of Sciences, 118*, e2016976118

Kim, Y., & Kim, Y. (2022). Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society, 79*, 103677

Kopmann, A., & Rehdanz, K. (2013). A human well-being approach for assessing the value of natural land areas. *Ecological Economics, 93*, 20-33

Krekel, C., Kolbe, J., & Wuestemann, H. (2016). The greener, the happier? The effect of urban land use on residential well-being. *Ecological Economics, 121*, 117-127

Lee, A.C.K., & Maheswaran, R. (2011). The health benefits of urban green spaces: a review of the evidence. *Journal of Public Health, 33*, 212-222

Li, C., & Managi, S. (2021a). Contribution of on-road transportation to PM2.5. *Scientific Reports, 11*, 21320

Li, C., & Managi, S. (2021b). Land cover matters to human well-being. *Scientific Reports, 11*

Li, C., & Managi, S. (2021c). Spatial Variability of the Relationship between Air Pollution and Well-being. *Sustainable Cities and Society*, 103447

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news, 2*, 18-22

Lundberg, S.M., Erion, G., Chen, H., Degrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*, 56-67

Lundberg, S.M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems, 30*

Mackerron, G., & Mourato, S. (2009). Life satisfaction and air quality in London. *Ecological Economics, 68*, 1441-1453

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change, 23*, 992-1000

Malek, Ž., & Verburg, P.H. (2020). Mapping global patterns of land use decision-making. *Global Environmental Change, 65*, 102170

Markevych, I., Schoierer, J., Hartig, T., Chudnovsky, A., Hystad, P., Dzhambov, A.M., de Vries, S., Triguero-Mas, M., Brauer, M., Nieuwenhuijsen, M.J., Lupp, G., Richardson, E.A., Astell-Burt, T., Dimitrova, D., Feng, X.Q., Sadeh, M., Standl, M., Heinrich, J., &

Fuertes, E. (2017). Exploring pathways linking greenspace to health: Theoretical and methodological guidance. *Environmental Research, 158*, 301-317

Mendoza-Ponce, A., Corona-Núñez, R., Kraxner, F., Leduc, S., & Patrizio, P. (2018). Identifying effects of land use cover changes and climate change on terrestrial ecosystems and carbon stocks in Mexico. *Global Environmental Change, 53*, 12-23

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com

OECD (2017). *Green Growth Indicators 2017*.

Pouso, S., Borja, Á., Fleming, L.E., Gómez-Baggethun, E., White, M.P., & Uyarra, M.C. (2021). Contact with blue-green spaces during the COVID-19 pandemic lockdown beneficial for mental health. *Science of The Total Environment*, *756*, 143984

Quek, K.F., Low, W.Y., Razack, A.H., & Loh, C.S. (2001). Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study. *Psychiatry and Clinical Neurosciences, 55*, 509-513

Seresinhe, C.I., Preis, T., MacKerron, G., & Moat, H.S. (2019). Happiness is Greater in More Scenic Locations. *Scientific Reports, 9*, 11

Seresinhe, C.I., Preis, T., & Moat, H.S. (2015). Quantifying the Impact of Scenic Environments on Health. *Scientific Reports*, *5*, 16899

Tsurumi, T., Imauji, A., & Managi, S. (2018). Greenery and Subjective Well-being: Assessing the Monetary Value of Greenery by Type. *Ecological Economics, 148*, 152-169

Tsurumi, T., & Managi, S. (2015). Environmental value of green spaces in Japan: An application of the life satisfaction approach. *Ecological Economics*, *120*, 1-12

Wang, L., Zhou, Y., Wang, F., Ding, L., Love, P.E.D., & Li, S. (2021). The Influence of the Built Environment on People's Mental Health: An Empirical Classification of Causal Factors. *Sustainable Cities and Society, 74*, 103185

Wang, R., Helbich, M., Yao, Y., Zhang, J., Liu, P., Yuan, Y., & Liu, Y. (2019). Urban greenery and mental wellbeing in adults: Cross-sectional mediation analyses on multiple pathways across different greenery measures. *Environmental Research*, *176*, 108535

White, M.P., Alcock, I., Wheeler, B.W., & Depledge, M.H. (2013). Would You Be Happier Living in a Greener Urban Area? A Fixed-Effects Analysis of Panel Data. *Psychological Science, 24*, 920-928

White, M.P., Pahl, S., Wheeler, B.W., Depledge, M.H., & Fleming, L.E. (2017). Natural environments and subjective wellbeing: Different types of exposure are associated with different aspects of wellbeing. *Health & Place*, *45*, 77-84

Wu, L., & Kim, S.K. (2021). Health outcomes of urban green space in China: Evidence from Beijing. *Sustainable Cities and Society, 65*, 102604

Yuan, L., Shin, K., & Managi, S. (2018). Subjective Well-being and Environmental Quality: The Impact of Air Pollution and Green Coverage in China. *Ecological Economics, 153*, 124-138

Appendix:

Table A1: Numbers of Observations in Each Country and Survey Method

Country	Number of Observations	Survey Method	GDP in 2017 (current USD)	Population in 2017
Egypt	1,010	Face-to-Face-Based	2.3537E+11	9.6443E+07
South Africa	1,110	Web-Based	3.4955E+11	5.7000E+07
China	18,931	Web-Based	1.2143E+13	1.3864E+09
India	6,562	Both	2.6522E+12	1.3387E+09
Indonesia	2,363	Both	1.0154E+12	2.6465E+08
Japan	10,098	Web-Based	4.8600E+12	1.2679E+08
Kazakhstan	1,000	Face-to-Face-Based	1.6681E+11	1.8038E+07
Malaysia	1,077	Web-Based	3.1896E+11	3.1105E+07
Mongolia	500	Face-to-Face-Based	1.1426E+10	3.1138E+06
Myanmar	1,073	Face-to-Face-Based	6.6719E+10	5.3383E+07
Philippines	1,672	Web-Based	3.1362E+11	1.0517E+08
Singapore	550	Web-Based	3.3841E+11	5.6123E+06
Sri Lanka	284	Web-Based	8.8020E+10	2.1444E+07
Thailand	1,115	Web-Based	4.5528E+11	6.9210E+07
Turkey	1,954	Web-Based	8.5268E+11	8.1102E+07
Vietnam	1,497	Both	2.2378E+11	9.4597E+07
Czech	1,178	Web-Based	2.1591E+11	1.0594E+07
France	2,130	Web-Based	2.5863E+12	6.6865E+07
Germany	3,165	Web-Based	3.6567E+12	8.2657E+07
Greece	1,358	Web-Based	2.0309E+11	1.0755E+07
Hungary	1,354	Web-Based	1.4151E+11	9.7880E+06
Italy	2,106	Web-Based	1.9570E+12	6.0537E+07
Netherlands	1,371	Web-Based	8.3181E+11	1.7131E+07
Poland	2,218	Web-Based	5.2622E+11	3.7975E+07
Romania	472	Web-Based	2.1170E+11	1.9587E+07

Russia	2,118	Web-Based	1.5786E+12	1.4450E+08
Spain	2,032	Web-Based	1.3093E+12	4.6593E+07
Sweden	1,330	Web-Based	5.4055E+11	1.0058E+07
United Kingdom	2,993	Web-Based	2.6662E+12	6.6059E+07
Canada	1,332	Web-Based	1.6469E+12	3.6543E+07
Mexico	1,669	Web-Based	1.1577E+12	1.2478E+08
United States	10,620	Web-Based	1.9485E+13	3.2499E+08
Australia	2,004	Web-Based	1.3301E+12	2.4602E+07
Brazil	2,255	Web-Based	2.0536E+12	2.0783E+08
Chile	1,174	Web-Based	2.7775E+11	1.8470E+07
Colombia	1,089	Web-Based	3.1179E+11	4.8901E+07
Venezuela *	807	Face-to-Face-Based	1.4384E+11	2.9390E+07
Total	95,571		6.6924E+13	5.1513E+09

The proportion of the world population: 68.58%

The proportion of world GDP: 82.67%

Note: Data on population and GDP are provided by World Bank. World Bank did not provide the GDP of Venezuela in 2017.

Population: https://data.worldbank.org/indicator/sp.pop.totl

GDP: https://data.worldbank.org/indicator/ny.gdp.mktp.cd

GDP of Venezuela: https://countryeconomy.com/gdp/venezuela

Table A2: Descriptive Statistics of Features

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Mental Health Score	89,273	24.319	6.301	0	21	29	36
DIG	89,273	-0.317	0.719	-0.997	-0.773	-0.121	3.000
Social Class	89,273	2.960	0.822	1	3	3	5
Student Dummy	89,273	0.056	0.231	0	0	0	1
Worker Dummy	89,273	0.533	0.499	0	0	1	1
Company Owner Dummy	89,273	0.021	0.142	0	0	0	1
Government Officer Dummy	89,273	0.031	0.174	0	0	0	1
Self-employed Dummy	89,273	0.076	0.265	0	0	0	1
Professional Job Dummy	89,273	0.035	0.183	0	0	0	1
Housewife Dummy	89,273	0.084	0.277	0	0	0	1
Unemployed Dummy	89,273	0.086	0.280	0	0	0	1
Pleasure	89,273	3.167	0.809	1	3	4	4
Anger	89,273	2.351	0.935	1	2	3	4
Sadness	89,273	2.347	0.954	1	2	3	4
Enjoyment	89,273	3.052	0.830	1	3	4	4
Smile	89,273	3.318	0.777	1	3	4	4
Euthusiastic	89,273	0.423	0.494	0	0	1	1
Critical	89,273	0.168	0.374	0	0	0	1
Dependable	89,273	0.648	0.478	0	0	1	1
Anxious	89,273	0.245	0.430	0	0	0	1
Open to New Experience	89,273	0.507	0.500	0	0	1	1
Reserved	89,273	0.430	0.495	0	0	1	1

Sympathetic	89,273	0.623	0.485	0	0	1	1
Careless	89,273	0.119	0.324	0	0	0	1
Calm	89,273	0.499	0.500	0	0	1	1
Uncreative	89,273	0.177	0.382	0	0	0	1
Urban Center Dummy	89,273	0.688	0.463	0	0	1	1
Urban Area Dummy	89,273	0.144	0.352	0	0	0	1
Rural Area Dummy	89,273	0.167	0.373	0	0	0	1
Income Group	89,273	2.795	0.900	1	2	3	5
Female Dummy	89,273	0.489	0.500	0	0	1	1
Age	89,273	42.797	14.817	18	30	54	99
Self-reported Health	89,273	3.824	0.883	1	3	4	5
Bachelor Dummy	89,273	0.390	0.488	0	0	1	1
Master Dummy	89,273	0.090	0.287	0	0	0	1
PhD Dummy	89,273	0.018	0.134	0	0	0	1
Community Livable	89,273	4.023	0.852	1	4	5	5
Community Attachment	89,273	3.621	1.038	1	3	4	5
Community Safety	89,273	3.017	0.748	0	3	3	4
Children Number	89,273	1.212	1.216	0	0	2	10
Cropland (%)	89,273	13.668	19.065	0.000	0.996	18.257	99.804
Forest (%)	89,273	12.482	19.495	0.000	0.513	15.169	100.000
Grassland (%)	89,273	10.469	14.771	0.000	1.057	13.935	99.316
Shrubland (%)	89,273	1.081	3.876	0.000	0.000	0.363	40.000
Wetland (%)	89,273	0.083	0.268	0.000	0.000	0.047	3.000
Water (%)	89,273	3.391	8.050	0.000	0.017	2.292	50.000
Urban Land (%)	89,273	57.991	33.761	0.000	29.081	88.476	100.000
Bare Land (%)	89,273	0.525	2.159	0.000	0.000	0.168	20.000

X	89,273	47.756	79.521	-128.6	1.025	116.397	153.555
Y	89,273	29.558	20.730	-53.3	21.029	42.002	69.558

Table A3: Descriptions of Features

Aspect	Predictor Name	Descriptions or Questions of Predictors
	Mental Health Score	The output variable
Income	DIG	Because of the economic gap between countries, absolute annual individual income is less effective in classifying the level of human mental health among people from different countries. Therefore, we employ a new variable DIG.
Social Class	Social Class	Which **social class** do you feel that you belong within your country? (5: Upper - 1: Lower)
Job	Student Dummy	Are you a student, now? (1: yes, 0: otherwise)
Job	Worker Dummy	Are you a worker including both full-time and part-time, now? (1: yes, 0: otherwise)
Job	Company Owner Dummy	Are you a company owner, now? (1: yes, 0: otherwise)
Job	Government Officer Dummy	Are you a government officer, now? (1: yes, 0: otherwise)
Job	Self-employed Dummy	Are you self-employed, now? (1: yes, 0: otherwise)
Job	Professional Job Dummy	Are you doing a profesional job, such as professor, lawyer, doctor, now? (1: yes, 0: otherwise)
Job	Housewife Dummy	Are you a housewife or househusband, now? (1: yes, 0: otherwise)
Job	Unemployed Dummy	Are you unemployed, now? (1: yes, 0: otherwise)
Emotion Weekly	Pleasure	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Emotion Weekly	Anger	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)

Emotion Weekly	Sadness	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Emotion Weekly	Enjoyment	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Emotion Weekly	Smile	How often have you felt or experienced the following feelings or actions within a week? (4: often - 1:not at all)
Personality	Euthusiastic	Do you see yourself as someone who is euthusiatic? (1: yes - 0: otherwise)
Personality	Critical	Do you see yourself as someone who is critical? (1: yes - 0: otherwise)
Personality	Dependable	Do you see yourself as someone who is dependable? (1: yes - 0: otherwise)
Personality	Anxious	Do you see yourself as someone who is anxious? (1: yes - 0: otherwise)
Personality	Open to New Experience	Do you see yourself as someone who opens to new experience? (1: yes - 0: otherwise)
Personality	Reserved	Do you see yourself as someone who is reserved? (1: yes - 0: otherwise)
Personality	Sympathetic	Do you see yourself as someone who is sympathetic? (1: yes - 0: otherwise)
Personality	Careless	Do you see yourself as someone who is disorganized? (1: yes - 0: otherwise)
Personality	Calm	Do you see yourself as someone who is calm? (1: yes - 0: otherwise)
Personality	Uncreative	Do you see yourself as someone who is conventional? (1: yes - 0: otherwise)
Geographical variable	Urban Center Dummy	The data are extracted from ESA. Data source (European Commission): https://ghsl.jrc.ec.europa.eu/datasets.php
Geographical variable	Urban Area Dummy	The data are extracted from ESA. Data source (European Commission): https://ghsl.jrc.ec.europa.eu/datasets.php
Geographical variable	Rural Area Dummy	The data are extracted from ESA. Data source (European Commission): https://ghsl.jrc.ec.europa.eu/datasets.php

Income Group	Income Group	Which **income group** do you feel that you belong within your country? (5: Upper - 1: Lower)
Gender	Female Dummy	Are you female? (1: yes, 0: otherwise)
Age	Age	Please tell us your age.
Physical Health	Self-reported Health	All in all, how would you describe your state of health? (5: very good - 1: very poor)
Education	Bachelor Dummy	Are you with bachelor degree? (1: yes, 0: otherwise)
Education	Master Dummy	Are you with master degree? (1: yes, 0: otherwise)
Education	PhD Dummy	Are you with Ph.D. degree? (1: yes, 0: otherwise)
Living Environmental Aspect	Community Liviability	How livable is your neighborhood? (5: very livable - 1: not livable)
Living Environmental Aspect	Community Attachment	How attached are you to your local community? (4: extremly attached - 1: completely detached)
Living Environmental Aspect	Community Safety	Please tell us about safety of your neighborhood. (4: very safe - 1: very dangerous)
Family Aspect	Children Number	How many children do you have?
Land Cover	Cropland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Forest (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Grassland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Shrubland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Wetland (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Water (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/

Land Cover	Urban Land (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/
Land Cover	Bare Land (%)	Percentage of a certain land type. Data Source: http://data.ess.tsinghua.edu.cn/

Note: The features that do not mention the data source are from our surveys.

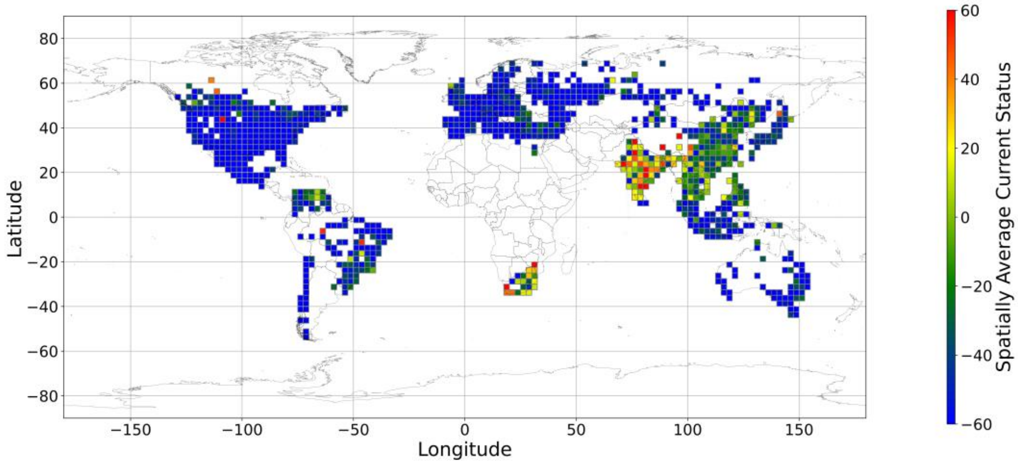

Figure A1.a: The Spatially Average Current Status of DIG

(Note: Cell size is 2.5° × 2.5°)

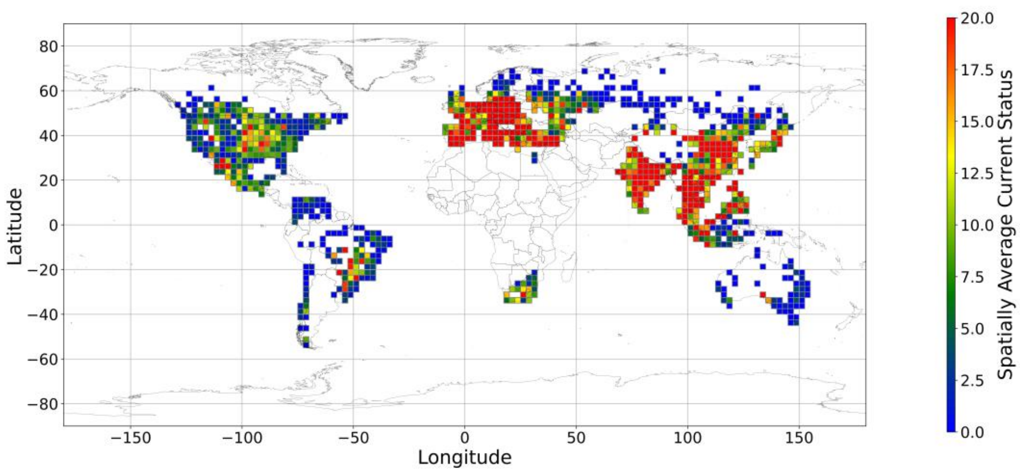

Figure A1.b: The Spatially Average Current Status of Cropland

(Note: Cell size is 2.5° × 2.5°)

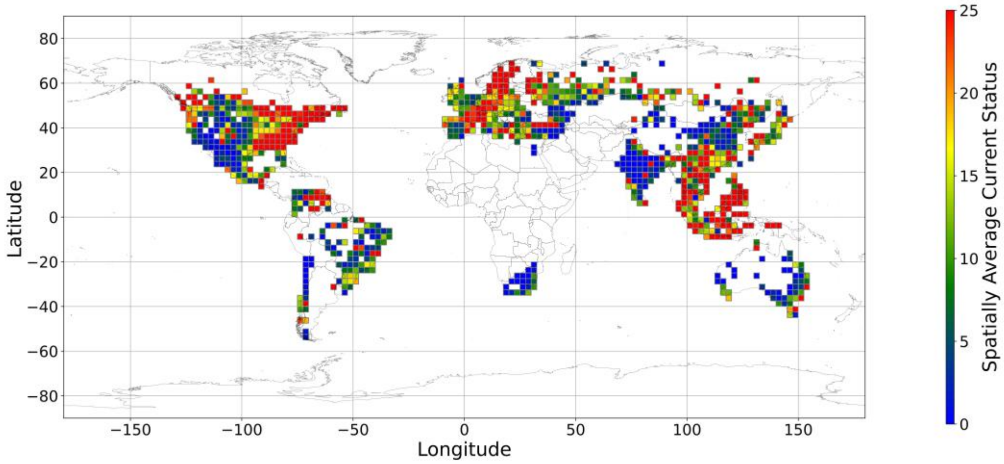

Figure A1.c: The Spatially Average Current Status of Forest

(Note: Cell size is 2.5° × 2.5°)

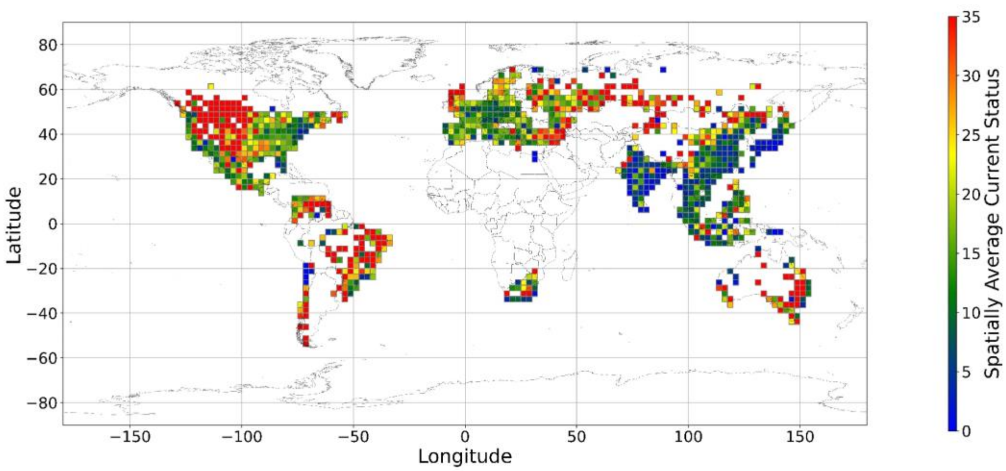

Figure A1.d: The Spatially Average Current Status of Grassland

(Note: Cell size is 2.5° × 2.5°)

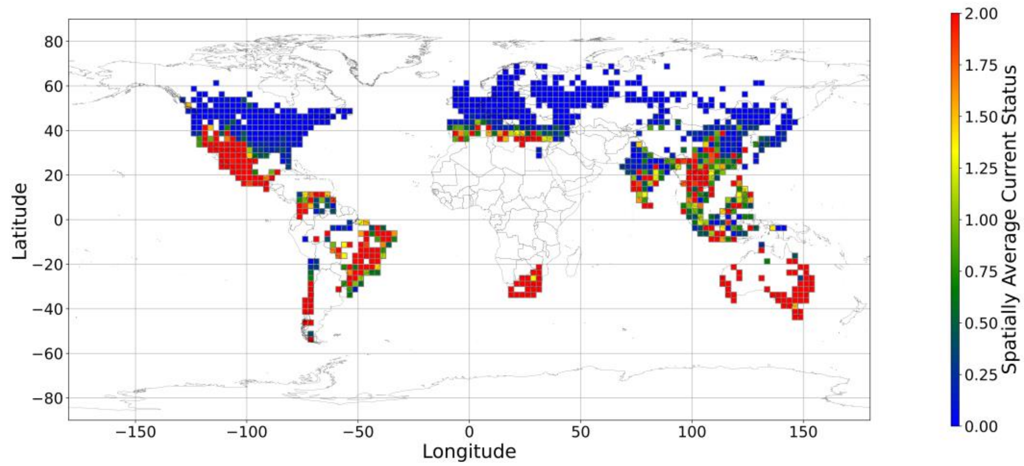

Figure A1.e: The Spatially Average Current Status of Shrubland

(Note: Cell size is 2.5° × 2.5°)

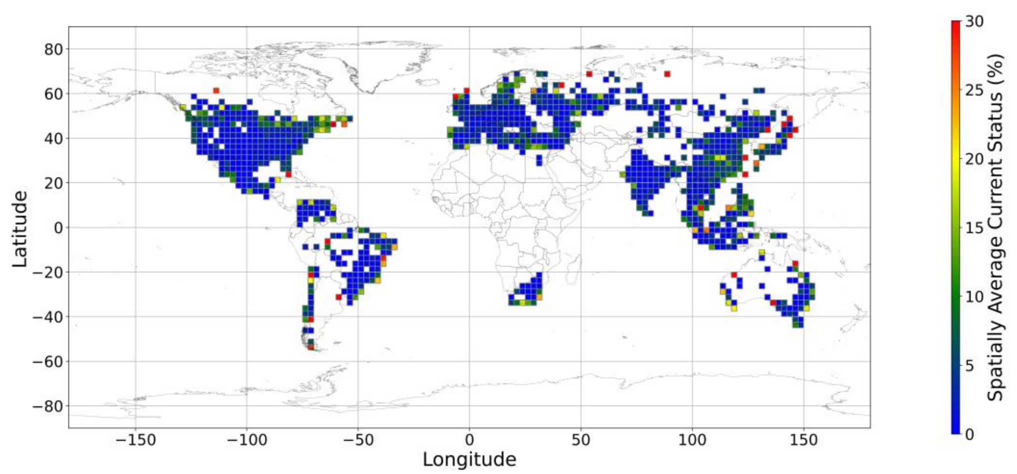

Figure A1.f: The Spatially Average Current Status of Water

(Note: Cell size is 2.5° × 2.5°)

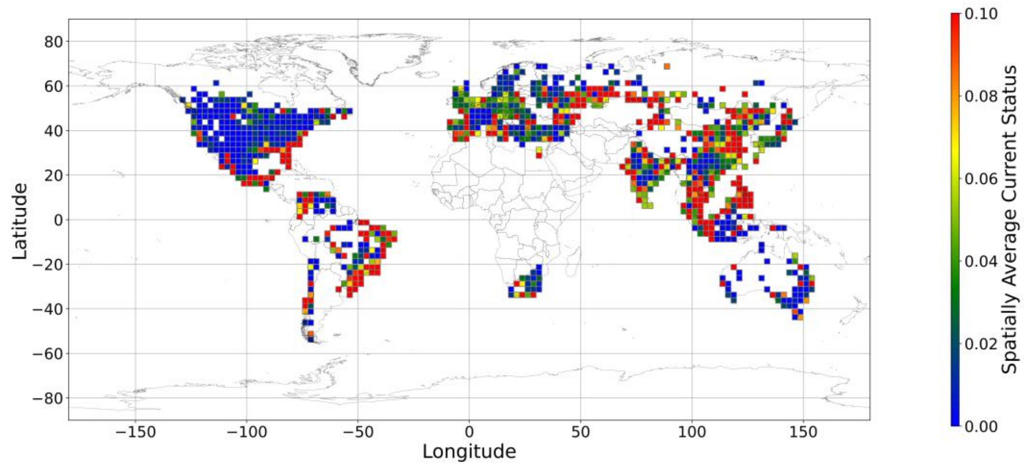

Figure A1.g: The Spatially Average Current Status of Wetland

(Note: Cell size is 2.5° × 2.5°)

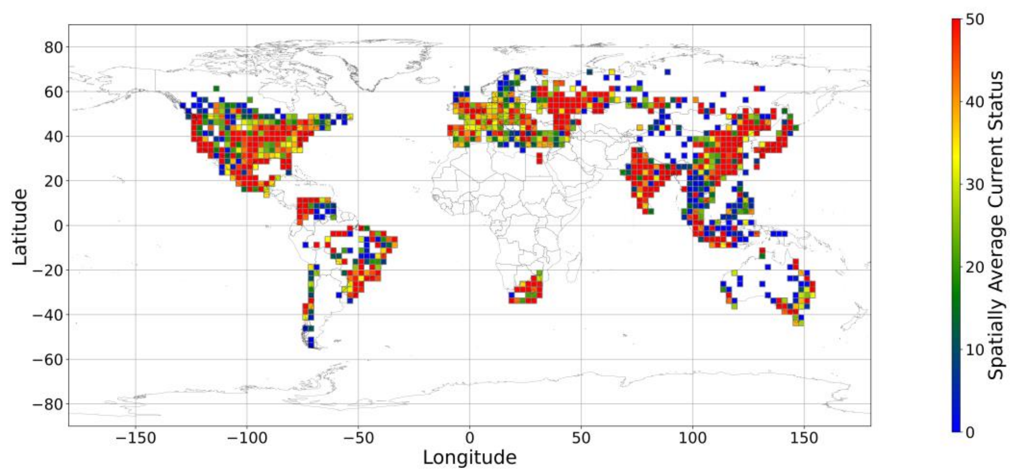

Figure A1.h: The Spatially Average Current Status of Urban Land

(Note: Cell size is 2.5° × 2.5°)

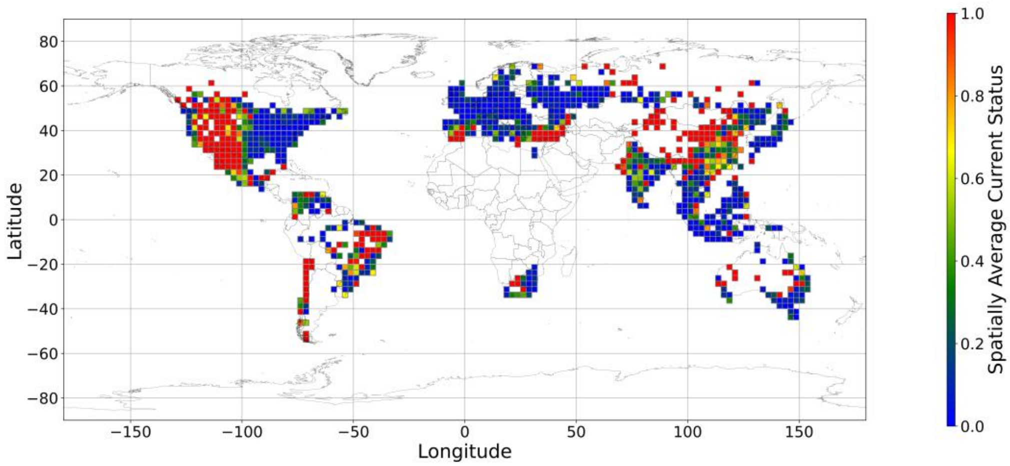

Figure A1.i: The Spatially Average Current Status of Bare Land

(Note: Cell size is 2.5° × 2.5°)