05

Methodology

"Composite [indices] involve a long sequence of steps that need to be followed meticulously" (Greco et al., 2018). The GGPM team applied a stepwise approach to enhance the transparency, replicability, and credibility of the Green Growth Index (Figure 11). This approach conforms to "good practices" in developing composite indices (Nardo, Saisana, Saltelli, & Tarantola, 2005; OECD & JRC, 2008). After concept building (chapter 4), the second step was an empirical application to systematically address methodological issues, such as data selection and statistical tests as well as normalization, weights, and aggregation of indicators. This chapter explains details of these methods. The third step, which aimed to check the robustness of the Green Growth Index, measured the explanatory power of the indicators and dimension subindices as well as the sensitivity and uncertainty levels of

the index. The fourth step, which focused on the presentation of the indicators, dimension subindices, and the Green Growth Index, required attention to enhance the comprehensibility and policy relevance of the results. This step considered not only the illustration of the results in maps, diagrams, and tables but also their assessments using benchmarks and ranks. This report, however, presents only selected results because most of the analyses will be discussed in GGGI's forthcoming Global Green Transformation Report (see chapter 9.1). It is a flagship report that will serve as a core part of GGGI's initiative to promote the model of green growth and showcase successful country experiences and approaches, supplemented by data, analysis, and stakeholder engagement.

Figure 11 Stepwise approach for developing the Green Growth Index

Steps	Activities	Rationale and description	Chapter
Concept building	• Objectives and purpose	• Guide development of the concept and inform purpose of the global index	1
	• Concept and definition	• Present underlying concepts for the Green Growth Index and its dimensions	4.1
	• Dimensions and indicators	• Asses policy relevance of indicators and data for each green growth dimension	4.2
Empirical application	• Data selection and statistical tests	• Set criteria for data selection, conduct correlation analysis, check for outliers	5.1–5.5
	• Normalization and benchmarking	• Identify lower/upper bounds for normalization and sustainability targets	5.6, 5.9
	• Aggregation and weights	• Assess weights and define aggregation methods (i.e. arithmetic, geometric)	5.7–5.8
Robustness check	• Sensitivity analysis	• Check the sensitivity of the global index to changes in indicators and targets	5.10.1
	• Uncertainty analysis	• Analyze impacts of changes in the framework and aggregation methods	5.10.2
	• Consistency check	• Correlate the Index to other green growth related indices	8.2
Presentation	• Global pattern of Index and dimensions	• Present green growth dimensions and global index on maps and tables	2, App. 1
	• Regional and dimension structure	• Compare Green Growth Index and dimension sub-indices across regions	6.1
	• Country and indicator specifics	• Analyze results for top performing countries and underlying indicators	6.2–6.3

5.1 Indicator selection

The conceptual framework should provide guidance on the choice of indicators (chapter 4.2), but the metrics or data to be selected to measure these indicators can be subjective, particularly when the "desired data" are not available (OECD & JRC, 2008). The selection criteria should thus be consistent with the objectives and purpose of developing the index. Because the Green Growth Index aims to measure green growth performance across countries and regions this year and the succeeding ones, GGPM used the following criteria in selecting indicators:

- Relevance of the indicator to the green growth dimensions based on conceptual and empirical evidence;

- Coverage of more than 140 countries, which include a large number of GGGI member and partner countries;

- Availability of time series data to allow updates of the index on a regular interval; and

- Accessibility of the data to allow replication of methods and check the credibility of their sources to enhance data acceptability.
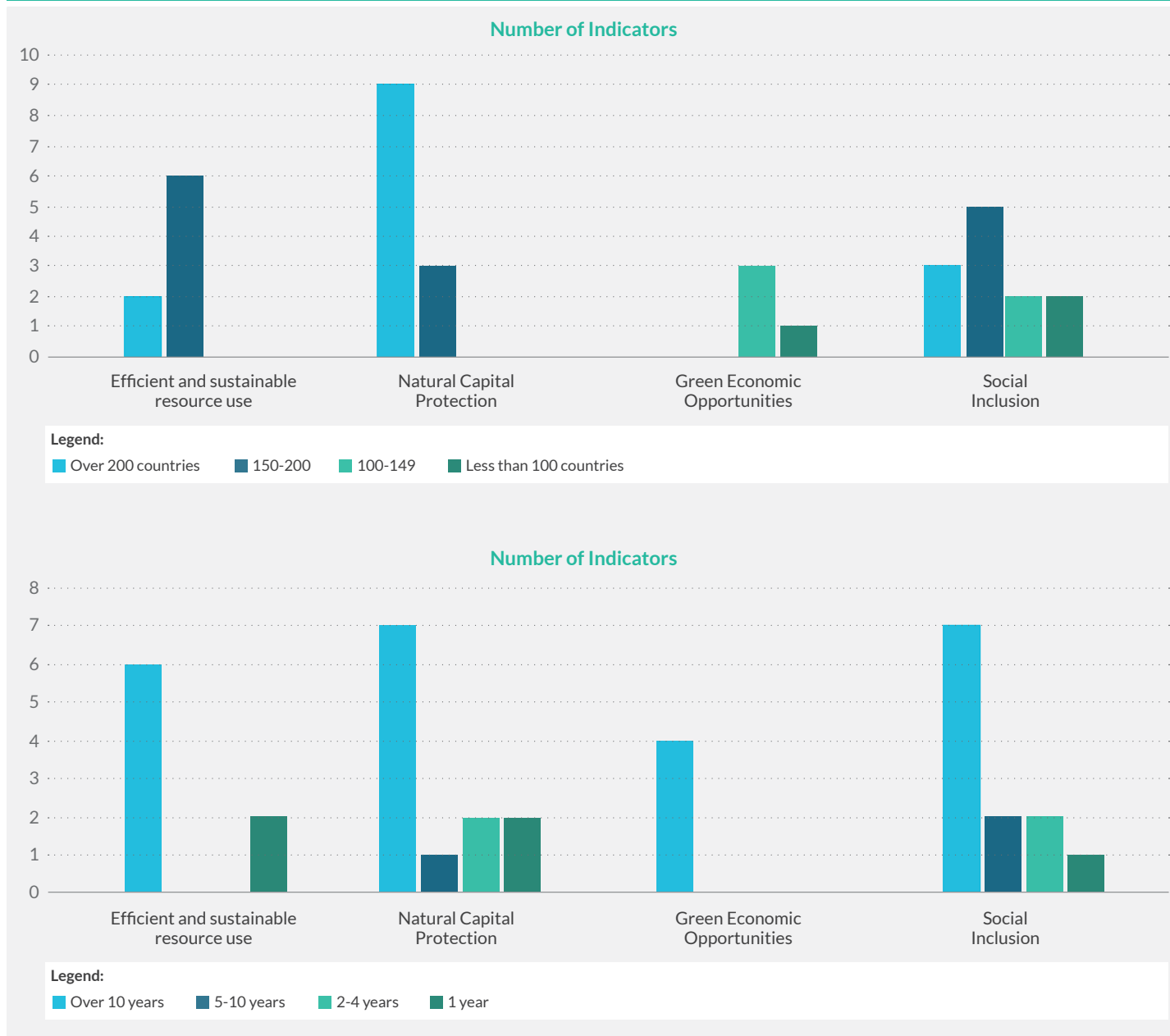
Literature review was conducted to provide evidence on the relevance of the indicators to the green growth dimensions and pillars (chapter 4.2; Acosta, 2019). Some of the indicators are, however, "proxy variables" because the desired indicators are either not available or there was a dearth of relevant data (see discussion in chapter 7.1 on indicators and proxy variables). Although the GGPM team aimed to have a wide data coverage in terms of the number of countries and years, some of the more relevant indicators did not meet these criteria. For example, there was data for less than 100 countries on one indicator for green economic opportunities, which is the share of patent publications in environmental technology to total patents, and two indicators for social inclusion, namely the share of youth (aged 15-24 years) not in education, employment, or training as well as the proportion of urban population living in slums (Figure 12). No alternative proxy variables are currently available for these indicators. Indicators for social inclusion, however, are expected to improve in the coming years because they are SDG indicators. Also, there was data for only one year for two indicators for efficient and sustainable resource use, specifically water use efficiency and average soil organic carbon content; for two indicators

for natural capital protection, specifically the municipal solid waste (MSW) generation per capita and soil biodiversity, or the potential level of diversity living in soils; and for one indicator for social inclusion, specifically the proportion of population above statutory pensionable age receiving a pension. Most of these indicators are proxy variables and expected to be replaced by more desired

data in the next few years. For example, FAO is currently finalizing its database for soil nutrients, which would be an alternative data source for soil organic content and soil biodiversity. Further improvements are also expected in data for water use efficiency and statutory pensions because they are SDG indicators.

Figure 12 Characteristics of the indicators based on country coverage and years of data availability

Number of Indicators

Legend:
- Over 200 countries
- 150-200
- 100-149
- Less than 100 countries

Number of Indicators

Legend:
- Over 10 years
- 5-10 years
- 2-4 years
- 1 year

Data for all indicators included in the Green Growth Index are publicly available online. The data were mainly collected from international organizations; this offers important advantages for measuring performance across countries. For example, collecting data from national agencies for more than 100 countries will be cumbersome, whereas data from international organizations are collected from national agencies and have undergone consistency checks. The United Nations coordinates statistical activities "to guarantee integrated systems of collection, processing and dissemination of data" (Eurostat, n.d.). Nonetheless, during the

regional consultation workshops, some regional experts expressed concerns over using data from international organizations (Acosta et al., 2019). To address these concerns, GGGI will encourage regional experts to undertake additional consistency check of the data once the data used in the development of the Green Growth Index become available online. Moreover, GGGI will help to communicate any concerns on the correctness and validity of the data to the international organizations that are responsible for producing and publishing the data.

5.2 Data scaling

"To have an objective comparison across small and large countries, scaling of variables by an appropriate size measure, e.g., population, income, trade volume, and populated land area, etc. is required" (OECD and JRC, 2008b: p.23). More than 70 percent of the 36 indicators are scaled data. They mainly use denominator data on gross domestic product (GDP) or gross national income (GNI), such as for primary energy supply, domestic material consumption, and adjusted net savings; area, such as for water use efficiency, PM2.5 air pollution, soil organic carbon content, organic agriculture, key biodiversity areas, terrestrial and marine biodiversity, and forest area; available resources or size of sector, such as for freshwater withdrawal, environmental export, green employment, and environment technology patent; and population, such as for material footprint, DALY rate as affected by unsafe water, municipal solid waste, GHG emissions, access to safe water and sanitation, access to electricity and clean fuels, and mobile and fixed broadband.

Three composite indices, which by default are scaled, were used as indicators, including the Red List Index, inequality in income based on the Atkinson Index, and the Healthcare Access and Quality Index. The Red List Index measures the variation in total extinction across species groups. The income inequality measure developed by Atkinson is based on the proportion of the total income that a given society would have to forego to allow equal income shares among the population (Afonso, LaFleur, & Alarcón, 2015). The Healthcare Access and Quality Index is based on the study of the Global Burden of Diseases (GBD), which used 32 causes from which death should not occur in the presence of effective care (Fullman et al., 2018). It is not uncommon to use indices in developing a composite index. Indices are particularly useful when one indicator is not sufficient to measure different issues that equally need attention or when one indicator only partially captures the problem or its solutions. Acosta (2019) provides detailed descriptions of the indicators to enhance comprehensibility of these indices.

5.3 Data imputation

A direct and most common approach to address missing data is to simply exclude or omit them (Gelman & Hill, 2007; He, 2010; Kang, 2013). The Green Growth Index partly adopts this approach. This is applied to indicators with time series data, where indicators are excluded when they have missing data for two consecutive years prior to the baseline year, which refers to the year that was used in computing the index. Examples of sustainability indices that do not apply data imputation include the Environmental Vulnerability Index of the South Pacific Applied Geoscience Commission, the UNEP Green Economy Progress Index, and ADB's Inclusive Green Growth Index. Kang (2013) emphasized the problems with missing data, including reduction in statistical power, bias in estimation of parameters, reduced representativeness of the samples, and increased complexity of analysis. While these are very relevant for complex modelling analysis, using simple and transparent aggregation methods to generate the Green Growth Index can reduce these problems (Chapter 5.8). Moreover, He (2010) explained that when data are missing completely at random (MCAR), analysis with missing data is unbiased. In most cases, there are no clear basis on whether data are missing at random, which is a prerequisite in most imputation methods (Nardo et al., 2005). Gelman & Hill (2007) also pointed out that excluding indicators with missing data will reduce the number of samples in the analysis.

Imputation methods, such as mean imputation, linear interpolation, regression analyses, maximum likelihood, multiple imputation, are widely used to fill in missing data (Horton & Kleinman, 2007; OECD & JRC, 2008; Kang, 2013; Wicklin, 2017). Examples of sustainability indices that apply data imputation include the Global Green Economy Index of DC, which uses the mean of the five closest countries; the African Green Growth Index of AfDB, which uses the mean of normalized indicators; the Ecological Footprint of the Global Footprint Network, which uses inter- or extrapolation; the Environmental Performance index of the Yale University and Columbia University,

which imputes the closest data points and uses extrapolation; the Sustainable Society Index of the Sustainable Society Foundation, which uses expert judgment; and the Happy Planet Index of the New Economics Foundation, which imputes data from the closest years. He (2010) categorized the methods of mean imputation and of treating missing data as a separate category as ad hoc because imputation is based on implausible assumptions, noting that "these methods impute the missing data only once and then proceed to the completed data analysis" (He, 2010: p.3). Single imputation methods are known to underestimate variance and standard errors because they assume to know the unobserved value with certainty (He, 2010; OECD & JRC, 2008). As far as the computation of composite indices is concerned, there are serious statistical problems associated with these imputation methods, which can affect the reliability of the analysis. For example, mean imputed data will not only reduce the variance but also change the correlation between the indicators (Wicklin, 2017). Both are problematic because a good variance is important to capture differences in scores across countries and, as discussed in Chapter 5.5, correlation is important to identify redundant indicators. In short, there are trade-offs when using data imputation, and decisions often depend on subjective judgement. The motivations for using, and not using, imputation methods should thus be justified because "[n]o imputation model is free of assumptions" (OECD & JRC, 2008:p.25). In order to minimize the statistical implications of various imputation methods, the GGPM team adopted the simplest approach of the Happy Planet Index, which imputed data only from the closest years; for instance, missing data for 2017 was imputed by data from 2016. In very few cases, the mean of the closest years was used when there was a lack of time series data to observe the trend, and only two data points were available.

Table 1 provides information on data availability for the indicators and which indicators that were subjected to imputation. Out of the 36 indicators, 12 required imputations. However, four out of 10 indicators only needed imputation for one country. The indicators with the largest number of countries subjected to imputation include GJ1 Share of green employment in total manufacturing employment (GT1) and share of youth (aged 15-24 years) not in education, employment or training (SE3). Data for GJ1 were estimated by the United Nations Industrial Development Organization (UNIDO) based on the methods developed by Moll de Alba & Todorov (2018, 2019 in press). SE3 is an SDG indicator. Data for both indicators are expected to improve in the next years.

Table 1 Characteristics of the indicators in terms of data availability and required imputation

Indicator	Available Data	Baseline data	Number countries	Required imputation	Number of countries imputed	Type of imputation
Efficient and sustainable resource use						
EE1	1990–2015	2015	191	Yes	3	Data from 2014
EE2	1990–2015	2015	212	No	-	-
EW1	2015	2015	165	No	-	-
EW2	1998–2007, 2014	2014	184	No	-	-
SL1	2019	2019	243	No	-	-
SL2	2004–2016	2016	162	Yes	1	Data from 2015
ME1	1970–2015	2015	186	No	-	-
ME2	1990–2015	2015	174	No	-	-
Natural Capital Protection						
EQ1	1990–2016	2016	194	No	-	-
EQ2	2000–2017	2017	195	No	-	-
EQ3	2018	2018	216	No	-	-
GE1	1960–2014	2014	201	Yes	1	Data from 2013
GE2	1990–2010	2010	203	No	-	-
GE3	1961–2016	2016	226	No	-	-
BE1	2000–2018	2018	225	No	-	-
BE2	1990–2016	2016	208	Yes	1	Data from 2015
BE3	2016	2016	218	No	-	-
CV1	1993–2016	2016	223	No	-	-
CV2	2014–2017	2017	184	No	-	-
CV3	2016, 2017	2017	210	No	1	Data from 2016
Green Economic Opportunities						
GV1	1990–2017	2016	126	Yes	7	Closest data from 2012 to 2015
GT1	2000–2017	2016	148	Yes	15	Data from 2014 or 2015**
GJ1	2000–2015	2015	119	No	-	-
GN1	1980–2017	2016	93	Yes	10	Data from 2015**
Social Inclusion						
AB1	2000–2015	2015	117	No	-	-
AB2	2000–2017	2015	214	No	-	-
AB3	2000–2017	2017	203	No	-	-
GB1	1990, 1997–2017	2018	193	No	-	-
GB2	2011, 2014, 2017	2017	144	Yes	7	Data from 2014
GB3	2009–2018	2018	187	No	-	-
SE1	2010–2017	2017	156	Yes	5	Data from 2016

Table 1 Characteristics of the indicators in terms of data availability and required imputation *(continued)*

Indicator	Available Data	Baseline data	Number countries	Required imputation	Number of countries imputed	Type of imputation
Social Inclusion						
SE2	2000–2016	2016	203	No	-	-
SE3	1990–2018	2016	88	Yes	23	Data from 2015 or 2017
SP1	2015	2015	175	No	-	-
SP2	1990–2015	2015	194	No	-	-
SP3	2000, 2005, 2010, 2014, 2016	2016	118	Yes	8	Data from 2014

*Refers to Figure 1 for the definition of the indicator codes
**Few datapoints were imputed using mean of data from 2015 and 2017

5.4 Distribution and outliers

An outlier is an observed value that has an "abnormal distance," whether extremely large or small value, from other values of a dataset (NIST-SEMATECH, 2013). Outliers can "distort mean, standard deviation and the covariance structure of the indicator" and alter correlation between indicators (Mishra, 2008). They also affect the normalized values of the indicators and thus need to be identified and accounted for (Nardo et al., 2005; OECD & JRC, 2008). Boxplots of the indicators were computed to show the distribution of numerical data and identify extreme values or outliers in the indicators. Figure 13 illustrates the boxplot for the ratio of the total primary energy supply to GDP, showing the presence of extreme outliers. It also shows the interpretation of the boxplots of the indicators.

Figure 13 Illustration and interpretation of the boxplots for the ratio of the total primary energy supply to GDP
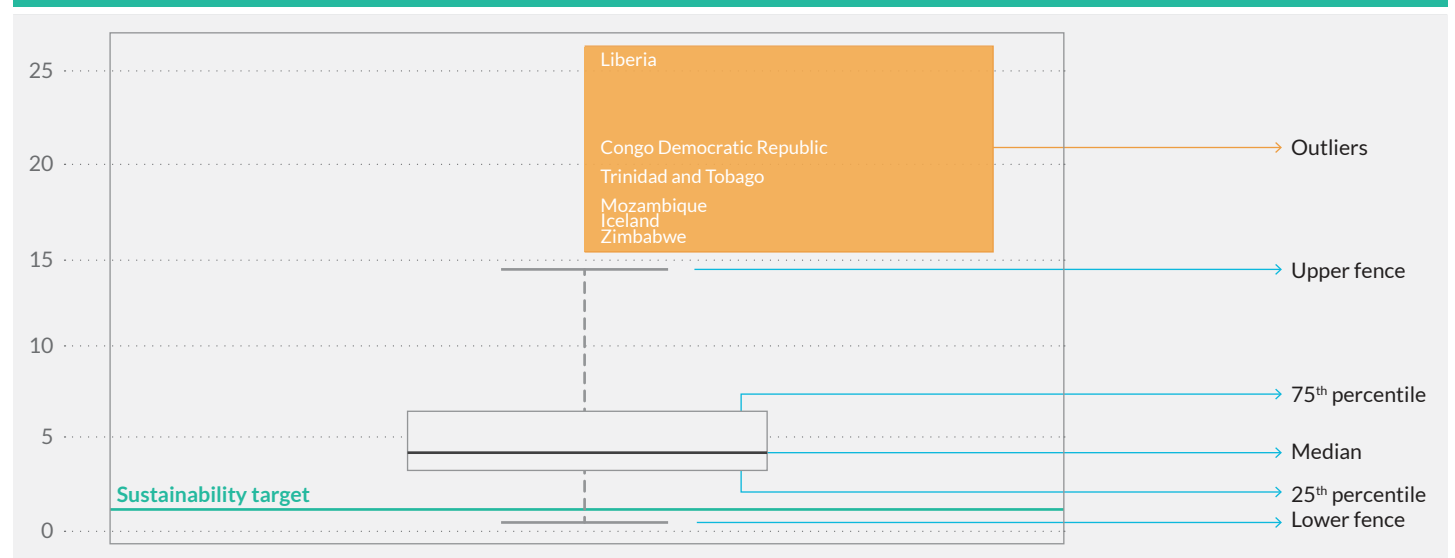
Table 2 summarizes the information from the boxplots, which were used to identify the outliers and the indicators that needed capping, where:

IQR = 75th percentile - 25th percentile
Lower fence = 25th percentile - μ x IQR
Upper fence = 75th percentile + μ x IQR

With μ = 3.0 the multiplier.

Although 2.2 is the recommended multiplier (Hoaglin & Iglewicz, 1987; Iglewicz & Banerjee, 2001), the GGPM team used a relatively higher multiplier to avoid generating too many extreme outliers and capping the data of many countries. Moreover, 3.0 is mostly applied in many standard statistical software to compute for extreme outliers. In some cases, the normalization approach that was used to compute the Green Growth Index allowed capping of the outliers through benchmarking. As explained in detail in Chapter 5.6.2, this will depend on the relationship of the indicator to green growth, whether negative or positive, and value of the indicators relative to the sustainability targets, whether above or below. When extreme outliers cannot be capped through benchmarking, they were capped prior to normalization. This is the case for the following indicators. Table 2 presents the number of capped values.

- **EE1:** Ratio of total primary energy supply to GDP (MJ per $2011 PPP GDP)
- **EW2:** Share of freshwater withdrawal to available freshwater resources (Percent)
- **ME2:** Total material footprint (MF) per capita (MF tons per capita)
- **EQ1:** PM2.5 air pollution, mean annual population weighted exposure (Micrograms per m³)

- **EQ2:** DALY rate as affected by unsafe water sources (DALY lost per 100,000 persons)
- **EQ3:** Municipal solid waste (MSW) generation per capita (Tons per year per capita)
- **GE1:** Ratio of CO_2 emissions to population, excluding AFOLU (Metric tons per capita)
- **GE2:** Ratio of non-CO_2 emissions to population, excluding AFOLU (Tons per capita)
- **GE3:** Ratio of non-CO_2 emissions in agriculture to population (Gigagrams per 1,000 persons)
- **GV1:** Adjusted net savings minus natural resources and pollution damages (Percent of GNI)
- **SE2:** Ratio of urban-rural access to basic services, such as water, sanitation, and electricity (Percent)

Capping outliers implies replacing extreme values with other values that more or less correspond to the structure of the rest of the dataset or the normal distribution. For the Green Growth Index, the GGPM team used the values of the lower and upper fences depending on whether the extreme outliers are beyond lower or upper fences as shown in Appendix 2. Except for the adjusted net savings minus natural resources and pollution damages (GV1), all other indicators with extreme outliers took the upper fence as their capped values.

Table 2 Summary of information for identifying and capping outliers *(continued)*

Indicator codes	25th Percentile	75th Percentile	IQR	Lower Fence	Upper Fence	Sustainability Targets*	Number Outliers**
Green Economic Opportunities							
GV1	1.02	14.73	13.71	-40.11	55.86	32.44	2
GT1	0.48	3.80	3.32	-9.48	13.76	13.52	0
GJ1	0.01	0.08	0.07	-0.20	0.29	0.14	0
GN1	0	0.02	0.02	-0.06	0.08	0.08	0
Social Inclusion							
AB1	51.93	92.88	40.95	-70.91	215.71	100.00	0
AB2	50.76	100.00	49.25	-96.98	247.74	100.00	0
AB3	43.82	78.44	34.62	-60.03	182.28	100.00	0
GB1	12.60	29.50	16.90	-38.10	80.20	50.00	0
GB2	1.03	1.26	0.23	0.36	1.93	1.00	0
GB3	50.00	100.00	50.00	-100.00	250.00	100.00	0
SE1	17.25	28.23	10.98	-15.68	61.15	7.96	0
SE2	1.00	1.36	0.36	-0.08	2.44	1.00	31
SE3	10.88	27.38	16.50	-38.62	76.88	0.00	0
SP1	16.87	98.55	81.68	-228.17	343.59	100.00	0
SP2	49.70	76.60	26.90	-31.00	157.30	100.00	0
SP3	12.33	53.08	40.75	-109.93	175.33	0.00	0

*Refers to Table 4 for details of the sustainability targets.
**Refers to outliers that were capped prior to normalization.
*** The share of freshwater withdrawal to available freshwater resources has a lower bound (25 percent) and an upper bound (75 percent). The extreme outliers refer to the upper bound, so the capped outliers assumed the values of the upper fence.

Table 2 Summary of information for identifying and capping outliers

Indicator codes	25th Percentile	75th Percentile	IQR	Lower Fence	Upper Fence	Sustainability Targets*	Number Outliers**
Efficient and Sustainable Resource Use							
EE1	3.49	6.36	2.88	-5.15	15.00	1.09	6
EE2	5.08	46.91	41.83	-120.40	172.39	51.40	0
EW1	4.00	32.30	28.30	-80.90	117.20	265.76	0
EW2***	1.96	30.10	28.14	-82.46	114.52	25.00	11
SL1	37.75	108.26	70.51	-173.79	319.79	289.34	0
SL2	0.14	3.20	3.06	-9.04	12.38	11.90	0
ME1	1.02	5.08	4.06	-11.15	17.25	0.17	0
ME2	3.55	20.89	17.34	-48.47	72.91	5.00	3
Natural Capital Protection							
EQ1	15.28	46.25	30.97	-77.62	139.14	10.00	4
EQ2	33.52	1085.98	1052.47	-3123.88	4243.38	0.00	7
EQ3	0.17	0.48	0.32	-0.78	1.43	0.00	1
GE1	0.81	6.19	5.38	-15.32	22.32	0.05	6
GE2	0.14	0.56	0.42	-1.13	1.82	0.00	17
GE3	0.22	0.93	0.71	-1.91	3.06	0.00	9
BE1	25.19	65.78	40.59	-96.58	187.55	100.00	0
BE2	10.93	48.06	37.13	-100.46	159.45	17.00	0
BE3	0.48	0.93	0.45	-0.86	2.27	1.16	0
CV1	0.78	0.94	0.16	0.30	1.42	1.00	0
CV2	25.00	80.00	55.00	-140.00	245.00	100.00	0
CV3	1.81	18.45	16.64	-48.11	68.37	13.50	0

5.5 Correlation of indicators

Bivariate correlation was used to analyze the strength of the association between the indicators in each dimension. Pearson correlation was the appropriate technique to use for the Green Growth Index because its indicators are continuous, and only a few of them have extreme outliers (chapter 5.4). Chok's (2008) study reveals that the correlation coefficient generated from this technique could improve statistical power even for distributions with moderate skewness. Its coefficient can take values from -1 to +1, where -1 shows perfectly linear but with negative relationship, +1 shows perfectly linear and with positive relationship, and 0 shows no linear relationship between the indicators (Bolboaca & Jäntschi, 2006). In the case of the Green Growth Index, the absolute values of the coefficients are more important than their signs. The aim of the correlation analysis is twofold: the first is to identify redundant indicators with very strong correlation, inducing double counting on the weights or the coefficient values; and the second is to verify whether indicators have acceptable levels of association in their respective dimensions or the p-value.

There are no clear rules on how to rate the values of the coefficients. According to Schober, Boer, & Schwarte (2018), many studies agree that "a coefficient of less than 0.1 indicates a negligible and more than 0.9 a very strong relationship, values in between are disputable" (Schober, Boer, & Schwarte, 2018: p.1765). In order to validate our indicators, it is necessary to have an acceptable significant correlation between 0.1 and 0.9. However, some experts consider these values very low and very high, respectively. The GGPM team thus interpreted the coefficient values according to a different range: 1 to 0.9 as very high; 0.89 to 0.7 and 0.1 to 0.29 as acceptable; 0.69 to 0.3 as ideal; and less than 0.1 as very low. The significance level of the correlation coefficient is represented by the p-value. When the p-value is below 0.01, then confidence in the correlation is 99 percent, or a 1 percent level of significance. When the p-value is between 0.01 and 0.05, then confidence is 95 percent, or a 5 percent level of significance, and when it is between 0.05 and 0.10, then confidence is 90 percent, or a 10 percent level of significance. Here the GGPM team investigated the absolute values of the correlation coefficients, only considering those with levels of significance that are equal or greater than 10 percent. Table 3 summarizes the results of the correlation analysis for each dimension, presenting those coefficients whose levels of significance are 10 percent or higher. Appendix 3 presents detailed results of the correlation analysis.

Table 3 Summary of the results of the correlation analysis between indicators in each dimension, in percent

Interpretation of the coefficient values	Efficient & sustainable resource use	Natural capital protection	Green economic opportunities	Social inclusion
1 – 0.9 (very high)	0	0	0	0
0.89 – 0.7 (acceptable)	0	0	0	10
0.69 – 0.3 (ideal)	29	41	0	43
0.1 – 0.29 (acceptable)	71	59	100	48
Less than 0.1 (very low)	0	0	0	0

Note: The numbers refer to the percentage of correlation coefficients with significance levels of 10 percent, 5 percent, and 1 percent (2-tailed). These values were drawn from the results presented in Appendix 3.

The correlation coefficients with significance levels of 10 percent or higher fall in the interval between 0.9 and 0.1 for all dimensions (Table 3), which means that no indicator has a very high level of correlation with another indicator. Many coefficients fall at an ideal level, between 0.3 and 0.7. However, a larger number of the coefficients are at an acceptable low level, between 0.1 and 0.3, particularly for indicators for green economic opportunities and efficient and sustainable resource use. About 10 percent of the correlation coefficients for social inclusion indicators are between 0.7 and 0.9, which is at an acceptable high level. The results of the correlation analysis reveal that there are no redundant indicators in our dataset, although many indicators have low, yet acceptable, levels of correlation. The only indicator with no statistically significant correlation with other indicators is the share of patent grants in environmental technology to total patent grants (GN1), one of the four indicators under the green economic opportunities dimension. This can be attributed to the small number of data points for this indicator,

having the lowest number even after imputation (Table 1). Overall, the correlation analysis confirms the validity and soundness of the model.

It is worth mentioning, however, that the indicators in the final framework are a result of an iterative process of statistical validation of the indicators. Other indicators were also considered in the framework but excluded and replaced with other indicators due to a very high correlation. These indicators include lower secondary completion rate, total (percentage of relevant age group); mean years of schooling (number of years); student-teacher ratio, primary school; gender inequality index; poverty headcount ratio at $1.90 per day; universal health coverage (UHC) service coverage index; wage and salaried workers, total (percentage of total employment); and share of GHG emissions and removals to population for AFOLU (Gigagrams per 1,000 persons).

5.6 Normalization of indicators

Normalization is a key method when developing a composite index, particularly when the index builds on multidimensional concepts and covers a large number of indicators. It helps to transform indicators with different units into uniform scales and unitless numbers that allow meaningful comparisons (Nardo et al., 2005; Pollesch & Dale, 2016); align indicators with positive and negative relationships to the phenomenon, which, in the case of this report, is green growth (Mazziotta & Pareto, 2013); and reduce uneven influence of indicators with extreme values on the index (Talukder, Hipel, & VanLoon, 2017). The most common methods for normalization include ranking; distance to target, or the best performer; standardization, or z-scores; re-scaling, or min-max transformation; and proportionate normalization (Nardo et al., 2005; Saisana & Saltelli, 2011; Mazziotta & Pareto, 2013; Talukder et al., 2017). There are no general rules for selecting appropriate normalization methods, so they are commonly based on subjective or expert judgement (Böhringer & Jochem, 2006; Hsu, Johnson, & Lloyd, 2013). But the choice of methods should consider properties of the indicators and objectives for constructing the composite index (Nardo et al., 2005; OECD, 2018).

Rescaling method, also known as min-max transformation, was chosen to normalize the indicators in the Green Growth Index for the following reasons:

- It is simple and the most widely used method, which will allow replication of the Green Growth Index by governments at the national and subnational levels.

- It can integrate upper and lower bounds in the method, which will reduce the problems of extreme values and partially correct for outliers.

- It allows application of targets in the method, which will represent benchmarking of sustainability targets.

5.6.1 Rescaling (min-max)

Generally, the method rescales a given indicator xi into different intervals with an identical range between 0 and 1 based on a minimum (X_{min}) and a maximum (X_{max}) *(Equation 1)*.

Equation 1

$$X^i_{norm} = \frac{x_i - X_{min}}{X_{max} - X_{min}}$$

where: X^{norm}_i = normalised i*th* indicator
$X = (x_1, x_2 ..., x_n)$
$n = 1, 2 ..., n$ number of countries

Many sustainability, environmental, and governance indices are using the rescaling method to normalize indicators. They include the Human Development Index of the United Nations Development Programme (UNDP), the Inclusive Green Growth Index of ADB, the Sustainable Society Index of the Sustainable Society Foundation (SSF), the Worldwide Governance Index of the World Bank (WB), the E-Government Development Index of the UN Public Administration Network, and the Democracy Index of the Economist Intelligence Unit (EIU). The range of the indices, however, is often not [0,1] because the rescaling method offers the advantage of setting boundaries (Talukder et al., 2017).

5.6.2 Benchmarking (lower/ upper bounds)

Equation 2 presents a more general mathematical function of the rescaling method in Equation 1 to include information on lower bound ***a*** and upper bound ***b***. The values of these boundaries are assigned arbitrarily and often depend on the objectives of the index. For example, ADB's Inclusive Green Growth Index has a range of 1 to 6 with the objective of aligning the scores with those of the World Bank's Worldwide Governance Index (Jha et al., 2018). The Green Growth Index used the range [1,100]. The lower bound of 1 is used instead of 0 because during the regional workshops (Chapter 3), some experts suggested avoiding using 0 in the index because it provides a negative notion and discourages performance improvement. Although the rescaling method generates unitless numbers with the objective of facilitating comparison across not only indicators but also years and countries, scores of zero could be misinterpreted to mean the lack of capacity to perform in a given indicator on green growth. The upper bound of 100 is used to imply achievement of the sustainability target for a given indicator (Chapter 5.6.3).

Equation 2

$$X^i_{norm} = a + \left(\frac{x_i - X_{min}}{X_{max} - X_{min}} \right) (b - a)$$

where: a = lower bound
b = upper bound

By integrating the targets into the rescaling method, the distance to sustainability targets can be directly measured from the scores of the indicators, or benchmarking (chapter 5.8). This approach is also referred to as the benchmarking normalization function, which "depends on indicator values each being mapped to some value based on a qualitative valuation of their level of sustainability" (Pollesch & Dale, 2016: p.198). OECD's Measuring Distance to the SDG Targets (OECD, 2019b, 2019a) and SDSN's SDG Index (Lafortune, Fuller, Moreno, Schmidt-traub, & Kroll, 2018; Sachs et al., 2019) applied this approach to measure country performance relative to the SDG targets. Pollesch & Dale (2016) compared how this approach was used in various studies to assess sustainability (e.g. Krajnc and Glavic, 2005; Castoldi and Bechini, 2010; Hayashi et al., 2014; Maxim, 2014; Pinar et al., 2014). In these studies, the boundaries were referred to as sustainability "thresholds," which were defined as either internal or external. Internal thresholds can refer to values that are specific to the system and the environmental or socio-economic sensitivities of the system being studied (Pollesch & Dale, 2016). The study of Pinar, et al. (2014) provided an example for using external thresholds, which were derived from outside sources, such as literature and international legislations.

The GGPM team used both internal and external thresholds, which, in the context of green growth, refer to the sustainability targets. In line with the study of Pinar et al. (2014), the external thresholds in the Green Growth Index are targets derived from literature. Specifically, these are targets that are explicitly agreed for the SDGs; implicit SDG targets based on the interpretations of OECD (2017b, 2019b, 2019a) and/ or SDSN (Sachs, Schmidt-Traub, Kroll, Lafortune, & Fuller, 2018; Sachs et al., 2019); or targets identified by experts for other international agreements, such as the air quality guidelines (WHO, 2005), Aichi targets (Leadley et al., 2014), and material resources (Bringezu, 2015). Meanwhile, the internal thresholds are targets derived from the mean values of the top county performers for specific indicators (Chapter 5.6.3).

The methods for integrating the boundaries or thresholds in the normalization function varied among different studies, mostly depending on the characteristics of the indicators used to measure these boundaries. In the case of the Green Growth Index, five different cases were identified for computing the upper bound b and integrating in the rescaling normalization method. Each case is elaborated below.

Case 1 was applied to indicators with a positive relationship to green growth and maximum values (X_{max}) that were less than the sustainability target (X^t). In this case, the upper bound ***b*** was based on the ratio between the difference of the maximum from the minimum value and the difference of the sustainability target from the minimum value (*Equation 3*). The reference point for both the maximum value and the sustainability target should be the minimum value of the indicator, which, in many cases, was not equivalent to zero. Case 1 assumed that none of the countries has reached the sustainability target of 100.

Case 1 assumptions

Equation 3

(i) positive relationship between the indicator and green growth
(ii) $X_{max} < X^t$
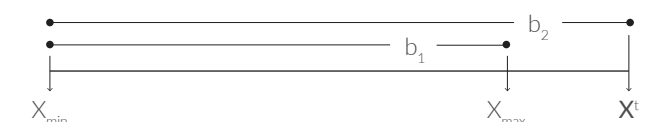
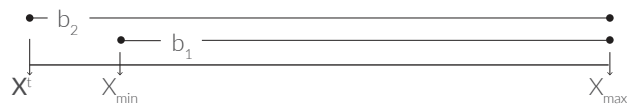$$X^i_{norm} = a + \left(\frac{x_i - X_{min}}{X_{max} - X_{min}} \right) (b - a)$$

where: $a = 1, b = \left(\frac{b_1}{b_2} \right) 100$

$b_1 = X_{max} - X_{min}, \ b_2 = X^t - X_{min}$

Case 2 was applied to indicators with a negative relationship to green growth and minimum values (X_{min}) that were greater than the sustainability target (X^t). Since the indicators have a negative relationship to green growth, the normalization function in Equation 4 was inverted. In this case, upper bound ***b*** was based on the ratio between the difference of the minimum from the maximum value and the difference of the sustainability target from the maximum value. The reference point for both the maximum value and the sustainability target should be the maximum value of the indicator. Similar to Case 1, Case 2 assumed that none of the countries has reached the sustainability target of 100.

Case 2 assumptions

Equation 4

(i) negative relationship between the indicator and green growth
(ii) $X^t < X_{min}$

$$X^i_{norm} = a + \left(\frac{x_i - X_{max}}{X_{min} - X_{max}} \right) (b - a)$$

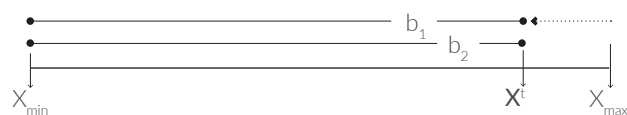where: $a = 1, b = \left(\frac{b_1}{b_2} \right) 100$

$b_1 = X_{min} - X_{max}, \quad b_2 = X^t - X_{max}$

Case 3 was applied to indicators with a positive relationship to green growth and some maximum values (X_{max}) that were greater than or equal to the sustainability target (X^t). The rescaling normalization function was modified, using the sustainability target as reference rather than the maximum value. For countries with values (x_i) that were greater than the sustainability target, their values for the indicator were modified by taking the value of the sustainability target. This assumed that they already met the target. This rescaling normalization method hence allowed the capping of any extreme values or outliers using the target value. Since upper bound b was based on the ratio between the difference of the maximum, which was capped using sustainability target, from the minimum value and the difference of sustainability target from the minimum value (*Equation 5*), b = 100. Case 3 assumed that some countries have reached the sustainability target of 100.

Case 3 assumptions

Equation 5

(i) positive relationship between the indicator and green growth
(ii) $X_{max} \geq X^t$

If $x_i > X^t$ then $x_i = X^t$

$$X^i_{norm} = a + \left(\frac{x_i - X_{min}}{X^t - X_{min}} \right) (b - a)$$

where: $a = 1, b = \left(\frac{b_1}{b_2} \right) 100$

$b_1 = X_{max} - X_{min}, \quad b_2 = X^t - X_{min}$

Case 4 was applied to indicators with a negative relationship to green growth and some minimum values (X_{min}) that were less than or equal to the sustainability target (X^t). Because the indicators have a negative relationship to green growth, the normalization function in Equation 6 was inverted. Moreover, the function was modified, using the sustainability target as reference rather than the minimum value. For countries with values (x_i) that were less than the sustainability target, their values for the indicator were modified by taking the value of the sustainability target. Similar to Case 3, the countries were assumed to

have already met the target and any extreme values or outliers were capped using the target value. Since upper bound b was based on the ratio between the difference of minimum, which was capped using sustainability target, from the maximum value, and the difference of the sustainability target from the maximum value (*Equation 6*), b = 100. Case 4 assumed that some countries have reached the sustainability target of 100.

Case 4 assumptions

Equation 6

(i) negative relationship between the indicator and green growth
(ii) $X^t \geq X_{min}$
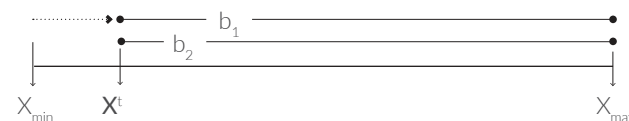
If $x_i > X^t$ then $x_i = X^t$

$$X^i_{norm} = a + \left(\frac{x_i - X_{max}}{X^t - X_{max}} \right) (b - a)$$

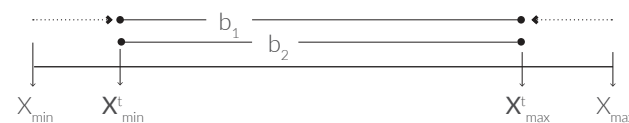where: $a = 1, b = \left(\frac{b_1}{b_2} \right) 100$

$b_1 = X_{min} - X_{max}, \quad b_2 = X^t - X_{max}$

Case 5 is a special case where there are both lower and upper bounds, which correspond to two sustainability targets: one at the minimum level and the other at the maximum level. This case was only applied to the share of freshwater withdrawal to total available freshwater, which has values lower than the minimum sustainability target and higher than the maximum sustainability target. For countries that met these conditions, their values for the indicator were modified by taking the values of the sustainability targets. Any extreme values or outliers were capped using these target values. Since upper bound b was based on the ratio of the same values, b = 100. This indicator has a negative relationship to green growth, so the normalization function in Equation 7 was inverted. Case 5 assumed that some countries have reached the sustainability target of 100.

Case 5 assumptions

Equation 7

(i) both lower and upper bounds for the indicator
(ii) negative relationship between the indicator and green growth
(iii) $X_{min} < X^t_{min}$ and $X^t_{max} < X_{max}$

If $x_i > X^t_{max}$ then $x_i = X^t_{max}$
If $x_i < X^t_{min}$ then $x_i = X^t_{min}$

$$X^i_{norm} = a + \left(\frac{x_i - X^t_{max}}{X^t_{min} - X^t_{max}} \right) (b - a)$$

where: $a = 1, b = \left(\frac{b_1}{b_2} \right) 100$

$b_1 = X_{min} - X^t_{max}, \quad b_2 = X^t_{min} - X^t_{max}$

5.6.3 Sustainability Targets

Figure 14 and Table 4 present the characteristics of the sustainability targets that were used to compute upper bound b in Chapter 5.6.2. Case 3 applied for more than half of the targets, and Case 4 applied to about a quarter of them. The former indicates that the indicators have a positive relationship to green growth and maximum values that were greater than the targets, while the latter suggests that indicators have a negative relationship to green growth and minimum values that were less than the targets. The number of indicators with a positive relationship to green growth is slightly higher than those with a negative relationship. The targets were grouped into three types: SDG targets; other targets, whose sources are not from the SDG indicators; and the mean of the top five performers. Where targets are not available from the SDG indicators

and other reliable literature, they were computed based the average values of the top five performing countries (bottom 5 performing countries for negative relationship to green growth). This approach was adopted from SDSN's Sustainable Development Report, which presents the SDG Index and Dashboards (Lafortune et al., 2018; Sachs et al., 2018, 2019). The targets in the Green Growth Index were aligned as much as possible with the SDG targets. Reference were thus made to those studies that identified targets for the SDGs, mainly OECD (2019a, 2019b) and SDSN (Sachs et al., 2018, 2019). For the SDG targets, the reference year was 2030, except for the share of marine biodiversity, which is 2020. Many countries have already achieved the 2030 targets for the SDG indicators (Table 4).

Figure 14 Characteristics of the sustainability targets used to benchmark the normalization functions
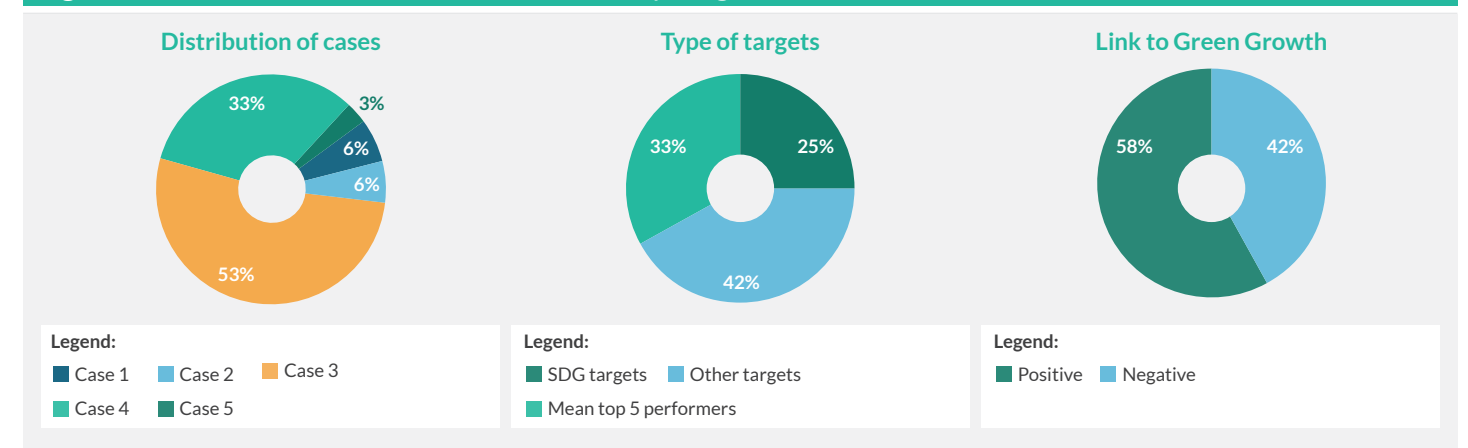
To sum up, the criteria for selecting the sustainability targets are based on the following:

1. For SDG indicators, SDG targets, both explicit and implicit, which were suggested in the OECD and SDSN reports were used. If the interpretation of implicit targets is different, the SDSN values, which are applied on a global context, were adopted.

2. For non-SDG indicators, targets suggested in scientific literature and reports from international organizations were used.

3. For SDG indicators not included in the OECD and SDSN reports, the mean of the top five performers was used.

4. For non-SDG indicators with no available information from the literature and reports, the mean of the top five performers was used.

Table 4 Details on the sustainability targets used to benchmark the normalization functions

Indicators	Link to green growth	Case	Min Max	Unstat sdg indicator	Targets	Countries reaching targets	Type of target	Source of data	Source of targets
Resource efficiency									
EE1: Ratio of total primary energy supply to GDP (MJ per $2011 PPP GDP)	negative	4	0.41 25.99	Yes	1.092 MJ per GDP	3	Mean top 5 performers	SE4ALL[i]	Method based on Sachs et al. (2019)
EE2: Share of renewables to total final energy consumption (Percent)	positive	3	0.00 95.82	Yes	51.4 percent	50	Other targets	SE4ALL	Sachs et al. (2019)[ii]
EW1: Water use efficiency (USD per m³)	positive	3	0.10 1157.90	Yes	265.7579346 USD per ᵐ3	5	Other targets	FAO	OECD (2019)
EW2: Share of freshwater withdrawal to available freshwater resources (Percent)	negative	5	0.00 2603.49	Yes	25 and 75 percent	130	Other targets	FAO	FAO 2017[iii]
SL1: Average soil organic carbon content (Ton per hectare)	positive	3	10.86 384.59	No	289.338 ton per hectare	2	Mean top 5 performers	FAO	Method based on Sachs et al. (2019)
SL2: Share of organic agriculture to total agricultural land area (Percent)	positive	3	0.00 81.14	No	11.90 percent	11	Other targets	FAO	OECD 2017b[iv]
ME1: Total domestic material consumption (DMC) per unit of GDP (DMC kg per GDP)	negative	4	0.02 15.76	Yes	0.169685364 kg per USD	5	Other targets	IRP	OECD (2019)
ME2: Total material footprint (MF) per capita (MF tons per capita)	negative	4	0.40 116.73	Yes	5.0 MF tons per capita	60	Mean top 5 performers	IRP[v]	Stefan Bringezu (2015)
Natural capital protection									
EQ1: PM2.5 air pollution. mean annual population-weighted exposure (Micrograms per m³)	negative	4	3.76 203.74	Yes	10 micrograms per m³	23	Other targets	Brauer et al. 2016	WHO 2005; OECD (2019)
EQ2: DALY rate due to unsafe water sources (DALY lost per 100,000 persons)	negative	2	0.81 10961.17	Yes	0 in every 100,000 population	0	SDG Target (explicit)	IHME[vi]	OECD (2019)

Table 4 Details on the sustainability targets used to benchmark the normalization functions *(continued)*

Indicators	Link to green growth	Case	Min Max	Unstat sdg indicator	Targets	Countries reaching targets	Type of target	Source of data	Source of targets
Natural capital protection									
EQ3: Municipal solid waste (MSW) generation per capita (Tons per year per capita)	negative	4	0.00 1.59	No	0.001752675 ton per year per capita	1	Other targets	WB	Sachs et al. (2019)
GE1: Ratio CO₂ emissions to population, excluding AFOLU (Metric tons per capita)	negative	4	0.04 45.42	No	0.054 Metric tons per capita	3	Mean top 5 performers	CDIAC[vii]	Method based on Sachs et al. (2019)
GE2: Ratio non-CO₂ emissions to population, excluding AFOLU (Ton per capita)	negative	4	0.00 22.42	No	0 ton per capita	8	Mean top 5 performers	FAOSTAT	Method based on Sachs et al. (2019)
GE3: Ratio non-CO₂ emissions in Agriculture to population (Gigagrams per 1000 persons)	negative	4	0.00 8.45	No	0 gigagrams per 1000 persons	13	Mean top 5 performers	FAOSTAT	Method based on Sachs et al. (2019)
BE1: Average proportion of Key Biodiversity Areas covered by protected areas (Percent)	positive	3	0.00 100.00	Yes	100 percent	1	SDG target (implicit)	IUCN, UNEP-WCMC	Sachs et al. (2019)[viii]
BE2: Share of forest area to total land area (Percent)	positive	3	0.00 98.26	Yes	17 percent	137	Other targets	FAOSTAT	OECD (2019)
BE3: Soil biodiversity. potential level of diversity living in soils (Index)	positive	3	0.22 1.22	No	1.156 index	2	Mean top 5 performers	JRC-ESDAC[ix]	Method based on Sachs et al. (2019)
CV1: Red list index (Index)	positive	1	0.41 0.99	Yes	1 index	0	Other targets	BirdLife International and IUCN	OECD (2019); Sachs et al. (2019)
CV2: Tourism and recreation in coastal and marine areas (Score)	positive	3	0.00 100.00	No	100 score	20	Other targets	Ocean Health Index	Sachs et al. (2019)[x]
CV3: Share of terrestrial and marine protected areas to total territorial areas (Percent)	positive	3	0.00 99.46	Yes	13.5 percent for both terrestrial and marine	77	SDG Target (explicit) for marine; Other targets for terrestrial	UNEP-WCMC	(Leadly et. al., 2014)[xiii]

Table 4 Details on the sustainability targets used to benchmark the normalization functions *(continued)*

Indicators	Link to green growth	Case	Min Max	Unstat sdg indicator	Targets	Countries reaching targets	Type of target	Source of data	Source of targets
Green economic opportunites									
GV1: Adjusted net savings. minus natural resources and pollution damages (Percent GNI)	positive	3	-84.71 37.48	No	32.438 percent GNI	3	Mean top 5 performers	WB	Method based on Sachs et al. (2019)
GT1: Share of export of environmental goods (OECD and APEC class.) to total export (Percent)	positive	3	0.00 49.78	No	13.52 Percent	4	Mean top 5 performers	UN-COMTRADE	Method based on Sachs et al. (2019)
GJ1: Share of green employment in total manufacturing employment (Percent)	positive	3	0.00 0.14	No	0.136 percent	3	Mean top 5 performers	Moll de Alba and Todorov (2018, 2019 in press)	Method based on Sachs et al. (2019)
GN1: Share of patent publications in environmental technology to total patents (Percent)	positive	3	0.00 0.20	No	0.076 percent	3	Mean top 5 performers	WIPO [xiii]	Method based on Sachs et al. (2019)
Social inclusion									
AB1: Population with access to safely managed water and sanitation (Percent)	positive	3	6.44 100.00	Yes	100 percent for both water and sanitation	5	SDG Target (explicit)	WHO/ UNICEF[xiv]	OECD (2019); Sachs et al. (2019)
AB2: Population with access to electricity and clean fuels/technology (Percent)	positive	3	4.04 100.00	Yes	100 percent for both	57	SDG Target (explicit)	SE4ALL	Sachs et al. (2019)
AB3: Fixed Internet broadband and mobile cellular subscriptions (Number per 100 people)	positive	3	6.87 179.34	Yes	100 subscriptions per 100 people	5	SDG Target (explicit for mobile, implicit for internet)	ITU[xvi]	Sachs et al. (2019)
GB1: Proportion of seats held by women in national parliaments (Percent)	positive	3	0.00 61.30	Yes	50 percent for parliament	3	SDG Target (explicit)	IPU [xviii]	OECD (2019); Sachs et al. (2019)
GB2: Share of female to male with account in financial institution, age 15+ (Percent)	negative	4	1.00 1.85	Yes[xix]	1 equality ratio	14	Other targets	WB	Normative

Table 4 Details on the sustainability targets used to benchmark the normalization functions *(continued)*

Indicators	Link to green growth	Case	Min Max	Unstat sdg indicator	Targets	Countries reaching targets	Type of target	Source of data	Source of targets
Social inclusion									
GB3: Getting paid, covering laws and regulations for equal gender pay (Score)	positive	3	0.00 100.00	No	100 percent	53	Other targets	WB	Normative
SE1: Inequality in income based on Atkinson (Index)	negative	4	5.80 56.40	No	7.96 Index	2	Mean top 5 performers	UNDP	Method based on Sachs et al. (2019)
SE2: Ratio urban-rural access to basic services (water, sanitation and electricity) (Percent)	negative	4	1.00 94.83	Yes	1 equality ratio	119	Other targets	WHO/ UNICEF, SE4ALL	Normative
SE3: Share of youth (aged 15-24 years) not in education, employment or training (Percent)	negative	2	1.29 46.89	Yes	0 percent	0	SDG Target (explicit)	ILO	OECD (2019) [xxi]
SP1: Proportion of population above statutory pensionable age receiving a pension (Percent)	positive	3	0.00 100.00	Yes	100 percent	41	SDG Target (explicit)	ILO	OECD (2019)
SP2: Healthcare access and quality index (Index)	positive	1	32.50 94.60	No	100 percent	0	Other targets	GBD[xxii]	GBD 2018
SP3: Proportion of urban population living in slums (Percent)	negative	4	0.00 97.50	Yes	0 percent	3	Other targets	UN-Habitat	Normative

Note: Details on data sources are availabile in Acosta (2019).

[i] *Sustainable Energy for All (SE4ALL) database from the SE4ALL Global Tracking Framework led jointly by the World Bank, International Energy Agency, and the Energy Sector Management Assistance Program*
[ii] *Alternative target is 58.62368011 percent based on OECD report (2019)*
[iii] *Alternative targets are 10 percent and 12.5 percent based on OECD (2019) and Sachs et al. (2019), respectively*
[iv] *OECD (2017) metadata, based on Share of agricultural land area under certified organic farm management*
[v] *UN Environment: Secretariat of the International Resource Panel (IRP), website: resourcepanel@unep.org*
[vi] *Institute for Health Metrics and Evaluation (IHME)*
[vi] *Carbon Dioxide Information Analysis Center (CDIAC), Environmental Sciences Division, Oak Ridge National Laboratory, Tennessee, United States.*
[vii] *Alternative targets are 92.69 and 37.73 percent for mountain and terrestrial/freshwater based on OECD (2019)*
[i] *Joint Research Centre, European Soil Data Centre (JRC-ESDAC)*
[x] *Based on scores for other OHI indicators*
[xi] *World Database on Protected Areas (WDPA) where the compilation and management is carried out by United Nations Environment World Conservation Monitoring Centre (UNEP-WCMC) in collaboration with governments, non-governmental organizations, academia and industry. The data is available online through the Protected Planet website (protectedplanet.net).*
[xi] *Average value for 17 percent terrestrial and 10 percent marine*
[xiii] *World Intellectual Property Organization (WIPO)*
[xiv] *WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply, Sanitation and Hygiene (washdata.org).*
[xv] *Alternative targets are 100 percent for electricity and 95 percent for clean fuels based on OECD (2019)*
[xvi] *International Telecommunication Union (ITU), World Telecommunication/ICT Development Report and database*
[xvii] *Alternative targets are 40.37400055 percent for total fixed broadband subscriptions per 100 inhabitants and 100 percent for proportion of population covered by a mobile network, by technology, based on OECD (2019)*
[xviii] *Inter-Parliamentary Union (IPU)*
[xix] *Refers to the actual indicator and not to the ratio between female and male*
[xx] *Refers to the actual indicator and not to the ratio between urban and rural*
[xxi] *Alternative target is 8.1 percent based on Sachs et al. (2019)*
[xxii] *GBD (2015) Global Burden of Disease Study 2015*

5.7 Weights of indicators and dimensions

Weights determine the relative importance of the indicators to each other. It entails the use of expert or subjective judgement that can become complicated in case of a multidimensional concept (OECD & JRC, 2008; Michaela Saisana & Saltelli, 2011). Gan et al. (2017) broadly categorized methods for weighting indicators into three: statistic-based weighting, public/expert opinion-based weighting, and equal weighting.

Statistic-based weighting uses quantitative methods to identify explicit weights, such as the principal component analysis, the data envelopment analysis, and the conjoint analysis (Nardo et al., 2005; OECD & JRC, 2008; Greco et al., 2018). The principal component analysis (PCA) is widely used to transform data into fewer dimensions and provides summaries of characteristics of high-dimensional data (Lever, Krzywinski, & Altman, 2017; Lever et al. 2017), but it can also be used to generate weights for the indicators based on the factor loadings (Chao & Wu, 2017; Hong-jun & Jin-feng, 2013). The GGPM team used PCA to compute the weights for the indicators (Appendix 4). The PCA weights, however, were not used in computing the Green Growth Index for two reasons: first, properties of the data influence the weights, which are expected to change when a new dataset with different structures are added to the composite index (Chapter 7.1); second, according to OECD & JRC (2008), this weight construction method is not valid and can be misleading for policy-guiding indicators. The weights from the PCA were used for the robustness check (see chapter 5.10).

The analytic hierarchy process (AHP) and the budget allocation process are examples of public or expert opinion-based weighting (Hudrlíková, 2013). AHP is a participatory and multicriteria decision-making approach that informs about the relative importance of indicators based on their pairwise comparisons (Dedeke, 2013; Pakkar, 2014). In AHP, the subjective judgment of the experts influences the weights. To facilitate the participation of the experts in identifying weights for the indicators, a survey questionnaire on AHP was developed for the Green Growth Index and distributed during the regional consultation workshops. The results of AHP revealed that there is a large divergence in consensus not only across regions but also across dimensions of green growth (Appendix 4). For this reason, it makes it difficult to use the AHP results to assign weights to the indicators. A higher level of consensus would be needed to identify the appropriate weights for the indicators.

The GGPM team used equal weighting for the Green Growth Index. Equal weighting is the most commonly used method in composite indices (Gan et al., 2017; Greco et al., 2018). Equal weights, which are often based on normative assumptions or based on understanding of the underlying concepts, are applied in composite indices, such as the Human Development Index, the Ecological Footprint, the Genuine Saving Index, the Environmental Vulnerability Index, the Sustainable Society Index, and the Corruption Perception Index. By not using weights from either AHP or PCA, the GGPM team assumed implicitly that the indicators have equal weights. Explicitly, however, the indicators do not have equal weights because the dimensions have a different number of indicators. This is clearly revealed by the PCA results in Figure A4.1 (see Appendix 4), where more weights are estimated for dimensions with the least number of indicators.

in each dimension. Similar to level 1, the 25 percent rule on missing values was applied to dimensions with more than four indicator categories, such as in the case of resource efficiency and green economic opportunities. This rule was not applied for the indicator categories under natural capital protection and social inclusion, which have only three categories each.

At level 3, geometric aggregation was applied to the dimensions, and the 25 percent rule on missing values was not applied. At this level of aggregation, no dimension was allowed to easily substitute for the other dimensions to improve the Green Growth Index. Thus, as the level of aggregation increases, the level of substitutability decreases.

Table 5 Comparison of linear and geometric aggregations

Characteristics	Types of aggregation methods	
	Linear/Additive	Geometric/Multiplicative
Data properties	A useful method when all individual indicators have the same measurement units, and further ambiguities due to the scale effects have been neutralized.	An appropriate method when noncomparable and strictly positive individual indicators are expressed in different ratio scales.
	It is useful when the underlying indicators are correlated.	It is useful in the presence of minor outliers.
Compensability	Full and constant compensability is allowed, such that deficits in one dimension can be traded off or substituted with surplus in another. Weights are substitution rates and depend on the trade-off value.	Partial compensability, limiting the ability of indicators with very low scores to be fully compensated for by indicators with high scores. No indicator's range dominates the mean values.
Policy implications	Priority will be to continue specializing in sectors where country has a comparative advantage.	Priority will be to increase in performance in sectors with the lowest score to improve overall ranking.

Sources: (OECD & JRC, 2008; Nardo & Saisana, 2008; Munda & Nardo, 2005; Saisana & Saltelli, 2011; Hudrlíková, 2013; Gan et al., 2017; Greco et al., 2018).

5.8 Aggregation of indicators and dimensions

Aggregation reduces dimensionality and provides a single holistic value (Pollesch & Dale, 2016) to measure performance. The two most common and simple methods include linear aggregation using arithmetic mean and geometric aggregation using geometric mean (Santeramo, 2016), with the former being more widely applied than the latter (Greco et al., 2018). For example, the Environmental Vulnerability Index and the Corruption Perception Index use linear aggregation, while the Human Development Index and the Sustainable Society Index use the geometric aggregation. The choice of aggregation methods should consider the properties of data, level of compensability, and implications on policy (Table 5). Both methods were used at the different levels of aggregation of the Green Growth Index (Figure 15).

At level 1, the indicators were linearly aggregated into indicator categories using the arithmetic mean. An important consideration here is the compensability of the individual indicators in each

indicator category. This allows countries with poor performance in one indicator, for instance, due to lack of resources, to be compensated by another indicator in the same indicator category. In most cases, the level of correlation between indicators in the same category is not negligible (Chapter 5.5), which can be assumed that they have some degree of substitution. Moreover, at level 1 of aggregation, a rule on missing value for a category with more than four indicators was applied: Countries with more than 25 percent of missing values were dropped. This method was adopted from Jha et al. (2018) in developing ADB's Inclusive Green Growth Index, which allowed indicators with missing values to be "substituted" by other indicators. This rule was not applied for the indicators in resource efficiency and green economic opportunities, which have less than three indicators in each category.

At level 2, geometric aggregation was applied to the indicator categories to allow only partial compensability between indicators

Figure 15 Methods of aggregation at the indicator, indicator category, and dimension levels
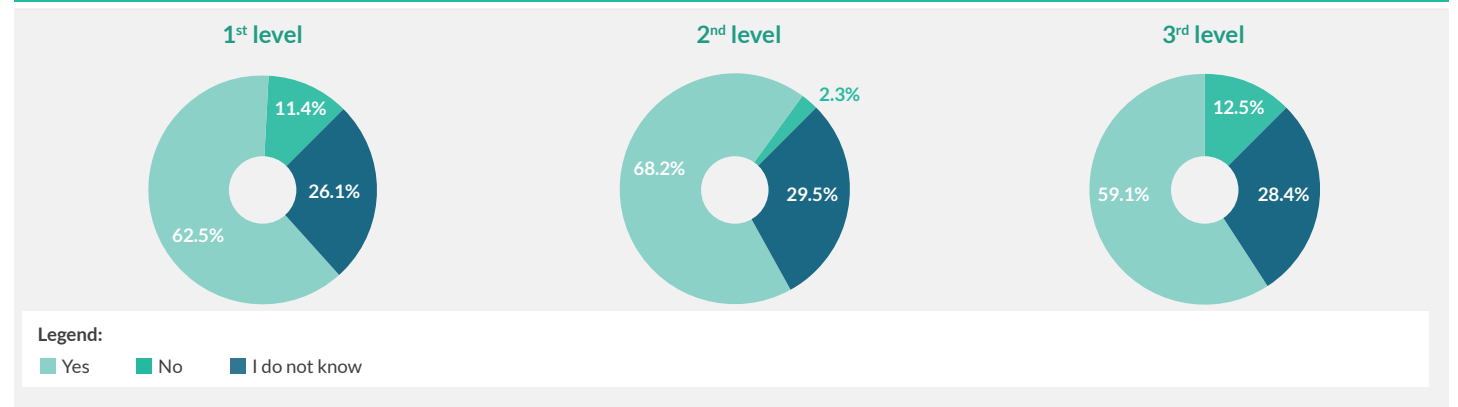

Normalized indicators

EE1	Ratio of total primary energy supply to GDP
EE2	Share of renewables to total final energy consumption
EW1	Water use efficiency
EW2	Share freshwater withdrawal to available freshwater
SL1	Average soil organic carbon content
SL2	Share of organic agriculture to agricultural area
ME1	Total domestic material consumption per GDP
ME2	Total material footprint per capita
EQ1	PM2.5, mean annual population-weighted exposure
EQ2	DALY rate as affected by unsafe water sources
EQ3	Municipal solid waste generation per capita
GE1	CO_2 emissions per capita, excluding AFOLU
GE2	Non-CO_2 emissions per capita, excluding AFOLU
GE3	Non-CO_2 emissions in agriculture per capita
BE1	Proportion of KBAs covered by protected areas
BE2	Share of forest area to total area
BE3	Soil biodiversity, potential level of diversity
CV1	Red list index
CV2	Tourism and recreation in coastal and marine areas
CV3	Share of terrestrial and marine PA's to territorial areas
GV1	Adjusted net savings
GT1	Share of environmental goods to total export
GJ1	Share of green employment in manufacturing
GN1	Share of environmental technology to total patents
AB1	Access to safely managed water and sanitation
AB2	Access to electricity and clean fuels/technology
AB3	Internet broadband and mobile cellular subscriptions
GB1	Seats held by women in national parliaments
GB2	Ratio of female to male with financial account
GB3	Laws and regulations for equal gender pay
SE1	Inequality in income based on Atkinson
SE2	Ratio urban-rural, safe water/sanitation and electricity
SE3	Youth not in education, employment or training
SP1	Proportion of population receiving pension
SP2	Healthcare access and quality index
SP3	Proportion of urban population living in slums

LEVEL 1 Linear aggregation of normalized indicators*

Efficient and sustainable energy
Efficient and sustainable water use
Sustainable land use
Material use efficiency
Environmental quality
Greenhouse gas emissions reductions
Biodiversity and ecosystem protection
Cultural and social value
Green investment
Green trade
Green employment
Green innovation
Access to basic services and resources
Gender balance
Social equity
Social protection

LEVEL 2 Geometric aggregation of indicator categories

Efficient and sustainable resource use
Natural capital protection
Green economic opportunities
Social inclusion

LEVEL 3 Geometric aggregation of dimensions

Green Growth Index

*No aggregation for indicators of green economic opportunities

Figure 16 Percentage of experts who agree on the methods used for the different levels of aggregation

1st level: 11.4% · 26.1% · 62.5%
2nd level: 2.3% · 29.5% · 68.2%
3rd level: 12.5% · 28.4% · 59.1%

Legend: Yes · No · I do not know

5.9 Ranks and benchmarks

Ranks and benchmarks are useful methods to measure green growth performance. During the regional consultation workshops, which constitute the second phase of consultations, the experts' opinions on how to rank the countries and which targets to use to benchmark the indicators were collected (Chapter 3). These topics need careful attention because they can influence the acceptability of any composite indices by policymakers, the public, and other stakeholders. Recognizing the continuous debates on the utility and credibility of composite indices, Saisana & Saltelli (2011: p.268) emphasize that indices "should never be seen as a goal, per se, regardless of their quality, [but] … as a starting point for initiating discussion and attracting public interest and concern".

"Rankings can be powerful tools of both branding and influence" (The Economist, 2014), but they also create controversies (Michaela Saisana & Saltelli, 2011; Chowdhury & Sundaram, 2016; Seth & McGillivray, 2018). Many popular indices, such as the Human Development Index, Environmental Performance Index, Corruption Perception Index, and Doing Business, use ranks to compare performance across countries. While the experts agreed on the usefulness of ranks to measure performance, they suggested avoiding the use of global ranks. They preferred using ranks only for groups of countries, by region or level of development, for instance, through which performance is more or less comparable.

"[A] set of indicators may have effect only when seen through a relevant benchmarking system that will give meaning to the produced measurements" (Benetatos, 2008: p.3). The methods and parameters require careful consideration when making decisions on benchmarking. The benchmarking method in the Green Growth Index was integrated in the normalization of indicators (Chapter 5.6). Benchmarking normalization is commonly applied in global sustainability indices, for instance, those developed by UNDP and OECD. The benchmarking parameters, specifically sustainability targets (chapter 5.6.3), were based on SDG targets as well as targets defined by other international organizations. Many experts suggested using SDGs and other internationally agreed targets, which the countries have to fulfil and achieve based on their international commitments, to benchmark the Index.

During the third phase of consultations, the expert reviewers were asked as to whether or not they agree on the aggregation methods used at different levels. This was important because when measuring performance relative to the SDGs, the choice of not only the indicators but also the methods influence countries' ranks (Miola & Schiltz, 2019). More than half of them agreed on the methods used to aggregate the Green Growth Index (Figure 16). However, the level of agreement slightly declined for the third level of aggregation. More than a quarter of the expert reviewers could not provide an answer to the question. The number of those who did not agree was small compared to those who agreed and who were not knowledgeable of the methods.

5.10 Robustness check

Composite indicators have faced criticism because such can be misleading if constructed poorly, and thus may be prone to misinterpretation (Saisana & Tarantola, 2002). Misinterpretation is prevented by conducting the final and essential step in the development of a composite index - evaluation of the confidence in the model as well as coming up with assumptions to support it.

Sensitivity and uncertainty analyses are standard procedures to evaluate the robustness of an index. These analyses increase the index transparency and evaluate which countries are advantaged or disadvantaged, thus permitting a necessary debate around the index (OECD & JRC, 2008). Sensitivity analysis measures the relative contribution to the output variance of individual sources of uncertainty on the input. Uncertainty analysis measures the impact on the output variance of uncertainty on the assumptions, method, and overall structure of the model. Both are closely linked and their combined approach shows

a more robust evaluation of the index confidence (Saisana et al., 2005). Results of the Monte Carlo models for the sensitivity and uncertainty analyses are presented below. More detailed results are available in other literatures (Flores, Acosta, Maharjan, & Peyriere, 2019).

5.10.1 Sensitivity analysis

The sensitivity analysis evaluates the impact on the Index of variations on the input. Two sources of uncertainty on the input exist: indicators and sustainability targets. The GGPM team manually and individually modifies the values of these inputs in a specific range and evaluates the impact on the Index. The six models used for the sensitivity analysis are described in Table 6.
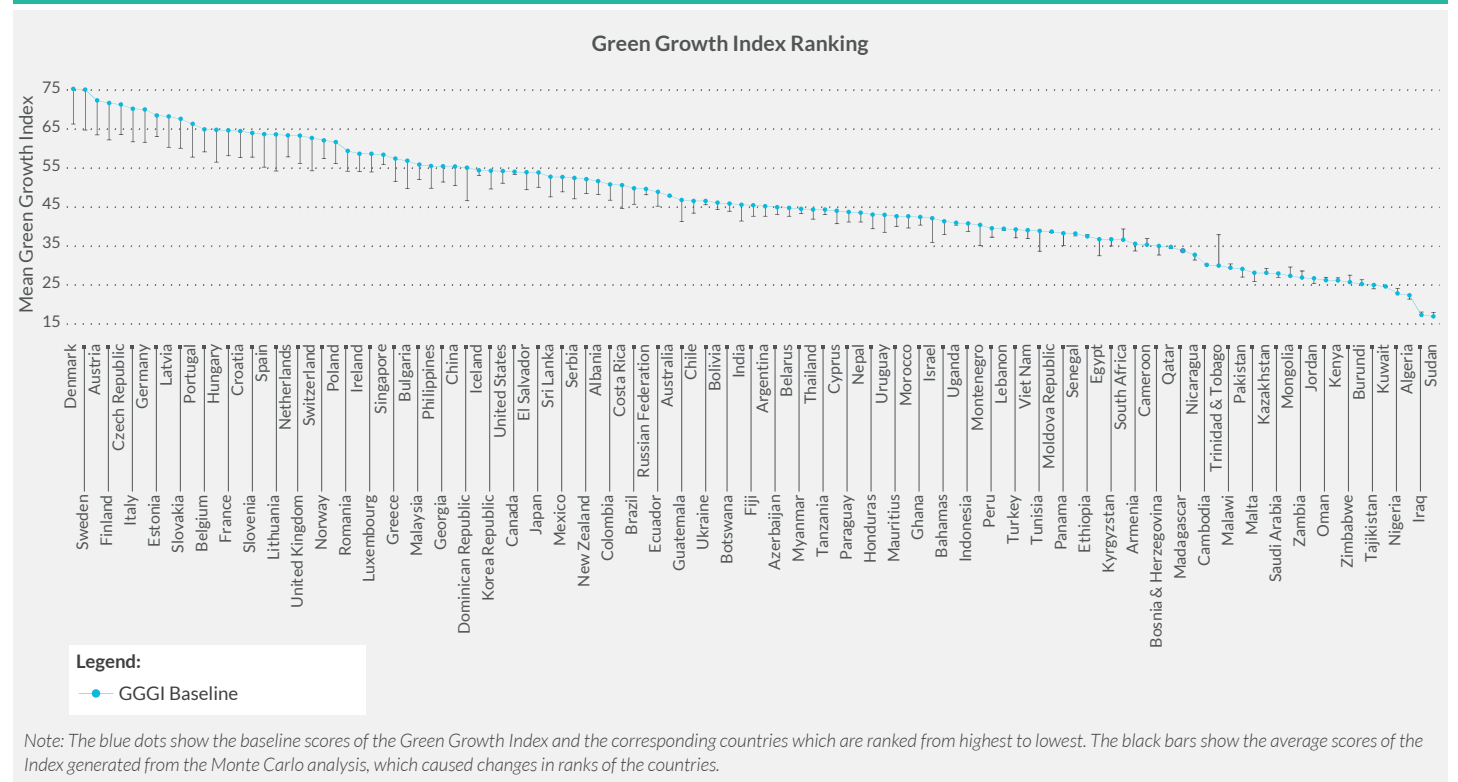
Table 6 Assumptions used in the six models for sensitivity analysis of the Green Growth Index

Types of sensitivity	Model assumptions	Baseline model
Change in values of indicators	Model A1.1 Increased values of indicators by intervals of 20 percent to 100 percent	Values of the indicators are based on baseline year (Table 1)
	Model A1.2 Decreased values of indicators by intervals of 20 percent to 100 percent	
Change in values of sustainability targets	Model A2.1 Instead mean of top 5 performers, used 90 percentiles for indicators with positive and 10 percentiles for negative relationship to green growth	Sustainability targets are mainly based on SDGs and other global targets (Table 4)
	Model A2.2 Increased values of sustainability targets by 50%, which is assumed to be target for 2050, except for targets which values are already 100	
Change in set of indicators	Model A3.1 Used indicators based on the 3rd draft framework (as described in Peyriere & Acosta)	Indicators are based on the framework with acceptable level of correlations (Figure 1)
	Model A3.2 Used indicators that were excluded from the final framework due to correlation analysis (see chapter 5.5)	

All in all, the sensitivity analysis based on the six models passed the robustness tests with good results (Flores et al., 2019). In this report, results of Monte Carlo analysis for Models A1.1 and A1.2 are presented. The input values were selected randomly within the specified range of change in indicator values (-100 and +100, at 20 percent interval). The values for the indicators were changed simultaneously at each iteration so that the interaction effects between the indicators were taken into consideration. This analysis enabled identification on how the scores and ranks of countries have changed within the specified range. The iteration was carried out over 1,000 times and a sequence of scores for the Green Growth Index was generated. Figure 17 summarizes the results of the

sensitivity analysis after randomizing the input values within ±20 percent. The results show that input variations caused only minimal changes on the scores of the Green Growth Index. Moreover, many countries maintained their rankings. There is an average change of 3.7 units in the index values among all 115 countries which resulted to 90 percent of the countries having a change in rank fewer than 8 places. The countries show an average change in ranks of 3.5 and with the top 30 countries shifting only 2.4 places. In addition, results of the sensitivity analysis show that changes in the input values to a certain amount have minimal impacts on countries with higher ranks than those with lower ranks.

Figure 17 Results of Monte Carlo model for the sensitivity analysis (1,000 iterations), Mean Green Growth Index scores

Note: The blue dots show the baseline scores of the Green Growth Index and the corresponding countries which are ranked from highest to lowest. The black bars show the average scores of the Index generated from the Monte Carlo analysis, which caused changes in ranks of the countries.

5.10.2 Uncertainty analysis

The uncertainty analysis evaluates the impact of changing the assumptions made and methods used to build the model of the Green Growth Index. There are four assumptions selected: aggregation, normalization, outliers and weights. These were easily measurable and had a high potential impact on the results and rankings of the index. The eight models for uncertainty analysis are described in Table 7.

Table 7 Assumptions used in the eight models for uncertainty analysis of the Green Growth Index

Sources of uncertainty	Model assumptions	Baseline model
Aggregation	Model B1.1 Used linear in all levels of aggregation	Aggregation methods were combined linear and geometric methods (chapter 5.8)
	Model B1.2 Used geometric in all levels of aggregation	
Normalization	Model B2.1 Applied standardization method using the sustainability targets	Normalization method was rescaling using sustainability targets (chapter 5.6)
	Model B2.2 Applied rescaling (min-max) method using b=100 as upper bound instead of the sustainability targets	
Outliers	Model B3.1 For all indicators, capped bottom values by 2.5 percentiles and top values by 98 percentiles	Outliers were capped based on lower and upper fences (chapter 5.4)
	Model B3.2 Used average of extreme outliers as value for capping (i.e., instead of upper fence)	
Weights	Model B4.1 Used weights based on estimates from Principal Component Analysis (see Appendix 4)	Weights were based on implicit values (i.e., no weights were assigned) (chapter 5.7)
	Model B4.2 Used weights based on mathematical distribution of the number of indicators	

Overall, the uncertainty analysis reveals that the impacts of changing model assumptions are acceptable and the model for the Green Growth Index is robust (Flores et al., 2019). To evaluate the overall impact of simultaneously applying the eight uncertainty models above, the GGPM team also applied a Monte Carlo analysis. Like in the sensitivity analysis, the aim was to analyze the changes in countries' scores and ranks for the Green Growth Index relative to the baseline model. Here, the assumptions were also randomized 1,000 times, building new scores and ranks for the Green Growth Index for each country each time. Figure 18 summarizes the results of the Monte Carlo analysis which reveal that the uncertainty is overall quite low and rankings are significantly maintained. About 48 percent of the countries show confidence intervals of three places or less, while 87 percent of the countries have a change in ranking of less than 10 places. On average, the countries show a change in ranks of 4.7, which is acceptable when ranking 115 countries. Results of the Monte Carlo analysis also show that changes in the assumptions on aggregation, normalization, outliers, and weights have lesser impacts on countries with higher ranks than those with lower ranks (Figure 18). This can be attributed to the larger divergence in the scores across indicators and indicator categories and dimensions in low ranking countries.

Figure 18 Results of Monte Carlo analysis for the uncertainty analysis (1,000 iterations), Mean Green Growth Index scores with 95 percent confidence intervals