Title: Developing a framework to improve global estimates of conservation area coverage

Rachel E. Sykes[1], Diego Juffe-Bignoli[1], Kristian Metcalfe[2], Helen M.K. O'Neill[1], P.J. Stephenson[3], Matthew J. Struebig[1], Piero Visconti[4], Neil D. Burgess[5, 6], Naomi Kingston[5], Zoe G. Davies[1] and Robert J. Smith[1]

[1] Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Canterbury, UK

[2] Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Penryn, Cornwall, UK

[3] IUCN SSC Species Monitoring Specialist Group, Laboratory for Conservation Biology, Department of Ecology & Evolution, University of Lausanne, Lausanne, Switzerland

[4] Biodiversity, Ecology and Conservation Group, Biodiversity and Natural Resources Management Programme, International Institute for Applied Systems Analysis, Laxenburg, Austria

[5] UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), Cambridge, UK

[6] Centre for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, Denmark

Word number: 6796

Number of figures and tables: 3 figures and 1 table

Keywords: conservation areas, other effective area-based conservation measures, protected areas, systematic conservation planning, targets

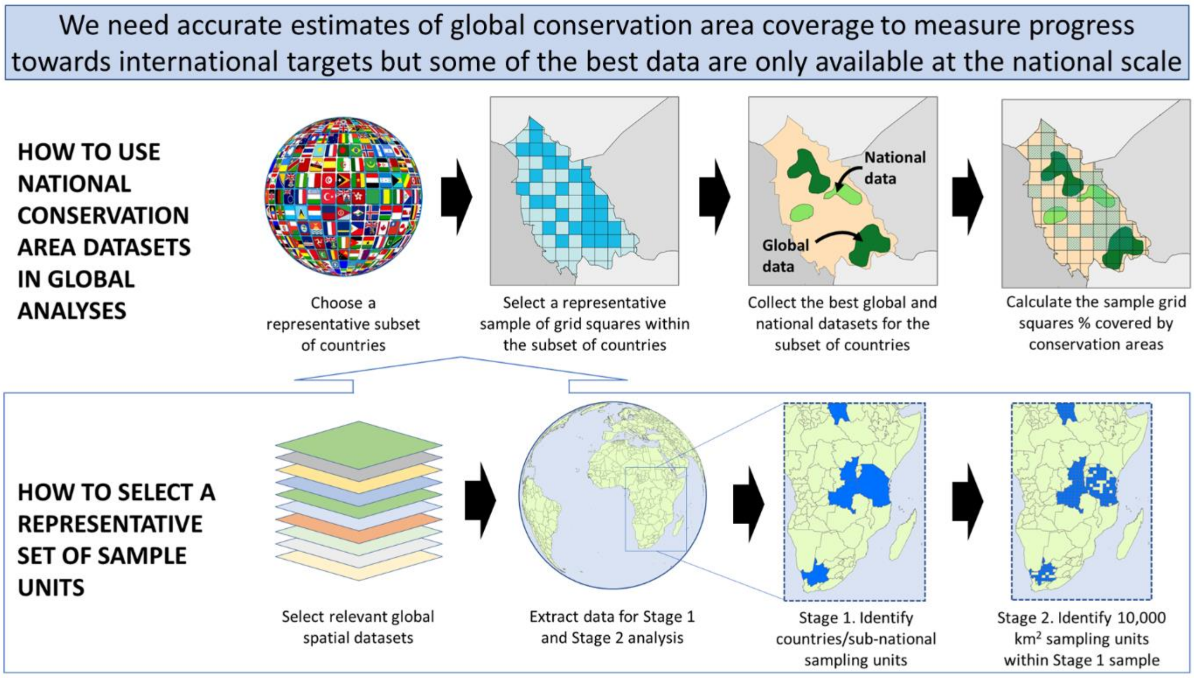
Author email addresses: Rachel E. Sykes (Rachel.Sykes@naturalengland.org.uk), Diego Juffe-Bignoli (diegojuffe@biodecisions.org), Kristian Metcalfe (Kristian.Metcalfe@exeter.ac.uk), Helen M.K. O'Neill (H.O'Neill@kent.ac.uk), P.J. Stephenson (stephensonpj@gmail.com), Matthew J. Struebig (M.J.Struebig@kent.ac.uk), Piero Visconti (visconti@iiasa.ac.at), Neil D. Burgess (Neil.Burgess@unep-wcmc.org), Naomi Kingston (Naomi.Kingston@unep-wcmc.org), Zoe G. Davies (Z.G.Davies@kent.ac.uk) and Robert J. Smith (R.J.Smith@kent.ac.uk)

Correspondence details: Professor Bob Smith, Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, Marlowe Building, University of Kent, Canterbury, Kent CT2 7NR; UK. Tel: +44 1227 823667. E-mail: R.J.Smith@kent.ac.uk.

Abstract

Habitat loss and species over-exploitation are among the leading threats to global biodiversity and international efforts to tackle these pressures often focus on area-based conservation management. There are ongoing discussions over what is an appropriate global conservation area coverage target, with some calls for 30% of the globe to be managed for conservation by 2030. To inform such debates it is important to know the extent and ecological representativeness of the current conservation area network, but this is hampered by gaps in existing global datasets. In particular, while data on privately and communally managed protected areas (PAs) and OECMs (other effective area-based conservation measures) are often available at the national or sub-national level, reporting delays and capacity issues means this information is not always incorporated into official datasets. To address this, here we propose a sample-based approach for producing more accurate metrics for describing the existing global conservation area network, based on selecting sampling areas that are representative of 10 factors that are known to influence conservation area establishment and biodiversity patterns. We then illustrate this approach by using a global search algorithm to identify a representative set of sampling units that cover 10% of the terrestrial realm across 25 countries. These sample units could be used as the focus of future data collation on different types of conservation area. Collecting and analysing data for this sample could produce quicker, more accurate estimates of global conservation area coverage and representativeness, complementing the existing international reporting system.

Graphical abstract



Developing a framework to improve global estimates of conservation area coverage

Introduction

Conservation areas are an essential component of global efforts to prevent biodiversity loss (Watson et al. 2014). To this end, the 196 signatories to the Convention on Biological Diversity committed through Aichi Target 11 to conserve 17% of the global terrestrial area within protected areas (PAs) and land under other effective area-based conservation measures (OECMs) by 2020 (CBD 2010). Progress towards Aichi Target 11 and other international commitments, such as the environment-related Sustainable Development Goals, is assessed using data from the World Database of Protected Areas (WDPA) and World Database of OECMs (WD-OECM). These databases are compiled and maintained by the UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), based on conservation area data approved by each national government or following an expert review and validation process (Lewis et al. 2019; Bingham et al. 2019; UNEP-WCMC and IUCN 2021). This makes the WDPA and WD-OECM extremely important sources of information, which need long-term, sustained resourcing to maintain their accuracy (Juffe-Bignoli et al. 2016). However, there are data limitations (Visconti et al. 2013), as some countries lack the capacity to provide up-to-date and accurate information, so it can take time for newer PAs to be included (UNEP-WCMC 2019). More generally, non-state PAs and OECMs are under-represented in the database (Bingham et al. 2017; Corrigan et al. 2018), partly because governments only recently started collecting data on conservation areas not owned or managed by the state. Additionally, some custodians of non-state conservation areas lack the capacity or are wary of providing information to governments about their land (Clements et al. 2018).

These limitations make it difficult to accurately measure progress towards international area-based conservation targets. It also makes it difficult to measure how well the global network represents biodiversity, especially as recent work suggests that non-state PAs can play a vital role in representing ecosystems that are missing from state PAs (Palfrey et al. 2022). Both these issues hamper the process of setting new targets, which is particularly important given that the international conservation community is pushing for more ambitious goals for conservation area extent and representativeness post-2020 (Dudley et al. 2018; Visconti et al. 2019; CBD 2021). Investing in improving the quality of global conservation area datasets will address this; there is an ongoing process for increasing the accuracy for state PAs and collecting information on non-state

PAs and OECMs (UNEP-WCMC 2019). Such work is vital but resource intensive and a long-term process (Juffe-Bignoli et al. 2016), so complementary approaches are needed.

One complementary approach could be to develop an analogous methodology to that of the Sampled Red List Index, which was designed to produce a better measure of species status and trends by accounting for taxonomic biases in the groups that have been fully assessed in the IUCN Red List (Baillie et al. 2008). Developing the Sampled Red List Index involved selecting a taxonomically representative set of species, conducting Red List assessments for those species that had not been assessed, and combining results for all of the selected species to produce a single measure (Lewis & Senior 2011). The equivalent for developing an estimate of global conservation area coverage would involve selecting a representative set of nations and then collecting the best available PA and OECM data to produce more accurate global indices. This would have similar benefits in terms of time and resources because data collection efforts would focus on a subset of countries. Just as importantly, producing this estimate based on a sample of data would not involve reporting results per country, so the analysis could use the latest and most accurate conservation area datasets without contradicting official data reported by governments. Here we present the first step in producing such a sampled approach for future estimates of global conservation area coverage and representativeness, developing a framework to identify a representative set of countries, and a set of sampling units within them.

Identifying a representative sample of countries, so that conservation area data from this subset can be used to estimate the extent to which the existing global PA and OECM network meets area and biodiversity targets, involves considering two sets of factors: drivers of conservation area establishment and drivers of biodiversity patterns. Establishment of conservation areas is influenced by a range of economic, political and social factors. For example, it is well known that conservation area coverage is higher on land of lower commercial value for agriculture or resource extraction (Loucks et al. 2008; Joppa & Pfaff 2009). Drivers of biodiversity patterns include latitude and elevation, as species and ecosystems show strong variation across these gradients (Gaston & Spicer 2013). Selecting a set of countries that best mirror these patterns is mathematically defined by the minimum set problem, so our framework is based on algorithms typically designed to solve these problems. This involves: selecting and mapping the features that influence conservation area extent and/or biodiversity pattern features; setting how much of each feature should be included in the sample, and; using complementarity-based algorithms to choose the best sets of countries that contain the specified amount of these features (Kukkala & Moilanen 2013).

Using this approach also involves choosing a cost metric, so that the prioritisation process minimises the cost while achieving the feature representation goals (Naidoo et al. 2006). In our case this metric needs to reflect the time and effort involved in collecting the conservation area data. PA and OECM datasets are generally collected and collated at the national level (Bingham et al. 2019), so each new country added to our sample would add an extra cost in terms of effort required. Thus, we define our cost metric as the number of countries in which our sample areas are found. Such a metric is a simplification, as the effort required will vary between countries based on their capacity and the number of conservation agencies that are responsible for national or sub-national data collection. We partially account for this in our study by dividing larger countries into their highest administrative units below the level of national government, such as states or provinces, to better match the devolved nature of conservation management and data collection in these countries.

Collecting data at the national level has one obvious disadvantage though, as these large sampling units are likely to contain some land that is not needed to meet the objectives, producing a less balanced sample because larger countries will be over-represented (Nhancale & Smith 2011). However, overcoming this simply involves repeating the spatial prioritisation using smaller sampling units within the subset of selected countries. Here we describe a sampling approach using this two-stage process to identify a representative set of countries and grid squares, designed to inform future efforts to collect, collate and supplement existing national PA and OECM datasets and produce more accurate measures of global patterns in conservation area coverage.

Methods

Our approach consisted of three steps, beginning with: (i) choosing socio-economic and biogeographic factors that represent drivers of conservation area extent and global biodiversity patterns, and (ii) defining and mapping the features that make up the categories within each factor. This was followed by (iii) a two-stage analysis: Stage 1 to identify the minimum set of countries needed to meet targets for each feature, and Stage 2 to identify sets of 10,000 km^2 grid squares that meet these targets within this subset of countries.

Choosing factors affecting biodiversity patterns and area-based conservation efforts

We conducted a literature review to identify factors that influence total conservation area network extent and patterns of global biodiversity. We then ran a workshop with 12 conservation area network experts to discuss these and other possible factors (Supplementary Materials Table S1) before generating a final list. This identified ten available global datasets that mapped these important factors: biomes, elevation, government effectiveness, islands and continents, landcover, latitude, income, population density of humans, realms and subregions (Table 1, Figure 1). Three of these factors were selected to represent only drivers of conservation area network extent, five to represent both drivers of conservation area network extent and global biodiversity patterns and two to represent only global biodiversity patterns (Supplementary Materials Table S2).

Table 1

Figure 1

Defining and mapping the features

To produce a representative sample, we needed to divide each factor into a number of categories (referred to as 'features' hereafter), either by using the existing classification system for categorical data or choosing appropriate thresholds for continuous data. We used the different datasets to produce a 1 km x 1 km resolution raster layer for each factor based on the Mollweide projection.

We used three datasets for the factors that only represent drivers of conservation area network extent (Figure 1). For government effectiveness features we used the World Bank's Worldwide Governance Indicators dataset, grouping countries into four categories based on government effectiveness scores of 0 - 24.9, 25 - 49.9, 50 - 74.9 and 75 - 100 (World Bank 2019a). For income features we used the World Bank low, lower-middle, upper-middle and high income country categories, which are based on per capita gross national income (World Bank 2019b). For human population density features we used UN data (UNPD 2013) classified into five categories using a logarithmic scale (0 - 0.9, 1 - 9.9, 10 - 99.9, 100 - 999.9 and > 1000 people per km^2) to ensure that the data adequately represented areas with very low and very high population densities.

We used five datasets for the factors that represent drivers of conservation area network extent and global patterns of biodiversity (Figure 1). For biome features we used WWF's global ecoregion GIS layer, where each of the 16 biomes is a broad ecosystem type (Olson et al. 2001). For elevation features, we used the Shuttle Radar Topography Mission's 1 km elevation data and divided these into five elevation categories of 0 – 299 m, 300 – 799 m, 800 – 1399 m, 1400 – 1999 m, ≥ 2000 m, based on existing studies of biodiversity and elevation gradients (Bruijnzeel & Veneklaas 1998; Linkie et al. 2010). For landcover features, we used the European Space Agency's GlobCover landcover map which divides the terrestrial realm into 12 broad landcover types (ESA GlobCover Project 2009). For the latitude features we created a latitudinal zone layer by dividing the globe into seven bands. Each band has a width of 20°, apart from at the poles where we used bands of 40° to avoid over-representing differences in these relatively small regions. For the subregion features we used the United Nations subregions classification to group countries into 22 categories (UNSD 2019).

We used two datasets for the two factors that only represent global biodiversity patterns (Figure 1). For the realm features we used WWF's global ecoregion dataset, where each of the eight realms is a large biogeographic unit (Olson et al. 2001). For the island and continent features, we used the Global Administrative Areas dataset (GADM 2018) and grouped them into five categories of < 1,000 km^2, ≥ 1,000 - 10,000 km^2, ≥ 10,000 - 100,000 km^2, ≥100,000 - 1,000,000 km^2 and "Continent" (>= 1,000,000 km^2). As part of this, we removed islands with an area < 1 km^2 because these are less likely to contain important terrestrial biodiversity (Whittaker & Fernandez-Palacios 2007). In addition, we classified islands as belonging to the continent feature if they were both <10 km^2 and within 100 km of a continent or Greenland, as these are likely to have similar species composition to their associated continents (Whittaker & Fernandez-Palacios 2007).

Spatial analysis

We used the Marxan software package (Ball et al. 2009) for the Stage 1 and Stage 2 analyses to identify the best set of sampling units, based on identifying a representative sample of the terrestrial realm by meeting targets for each of the 89 features across the ten factors, whilst minimising the number of countries selected. This is a novel use of Marxan, which is generally used to identify priority areas for conservation, whereas our analyses identify priority areas for data collection. We used Marxan in Stage 1 to identify a representative set of countries and territories. In Stage 2 we then identified 10,000 km^2 grid squares within these countries (Table 2), thus refining the sample from Stage 1 to avoid over-representing larger nations.

Marxan uses a simulated annealing algorithm, where each analysis involves running the software multiple times and producing a near-optimal portfolio each time. Marxan then produces two key outputs: the "best" output, which is the portfolio from the run with the lowest cost, and the "selection frequency" output, which counts the number of times each sampling unit appears in each of the portfolios. Sampling units with high selection scores are always needed to meet the targets; lower scoring sampling units can be swapped with similar sampling units without affecting target attainment (Ball et al. 2009).

For Stage 1, the sampling units were derived from the Database of Global Administrative Areas (GADM 2018) and consisted of countries for nations with an area < 1,000,000 km^2 or the highest sub-national administrative level polygons for larger countries (e.g. states, provinces, etc that are classified as L1 in the database and referred to as 'sub-national sampling units' hereafter). We took this approach because larger nations tend to have sub-national conservation agencies and legislation, so we wanted to minimise the number of these sub-national administrative units selected to avoid having to collate data from a large number of expert groups. We followed established practice for reporting terrestrial coverage statistics by excluding Antarctica from our analyses (Butchart et al. 2015). The Stage 2 sampling units were based on a global set of 100 km x 100 km grid squares, which was created in QGIS 3 using the Create Grid tool (QGIS 2019). We then used QGIS to clip this global grid layer with the national and sub-national sampling units used in Stage 1 to produce the final sampling unit layer.

We used the CLUZ plugin (Smith 2019) for QGIS to import the feature raster layers, calculate the area of each feature in each sampling unit and run Marxan. To ensure the sampling units selected in Stage 1 and 2 were representative of the terrestrial realm, we used Marxan to identify sampling units that when combined met the same percentage of total extent target for every feature. We carried out a sensitivity analysis to select this target, based on identifying a good compromise between sampling a sufficient proportion of the planet to produce a robust estimate of conservation area coverage, whilst minimising the number of national and sub-national sampling units. Based on this sensitivity analysis we chose a target value of 10%, as the number of sampling units required to meet higher targets increased more than two-fold (Supplementary Material Table S3). Thus, the set of sampling units identified by Marxan contained 10% of the total area of each of the 89 features.

The Stage 1 and Stage 2 analyses both involved 1,000 Marxan runs (see Supplementary Material for more details). The Stage 1 analysis was based on 900 national and sub-national sampling units. Each run consisted of 10 million iterations, and we set the costs so that Marxan ensured each portfolio met all the targets and also minimised the number of countries selected (Supplementary Material). The Stage 2 analysis was based on the 3,377 grid squares found within the national and sub-national sampling units selected in Stage 1. Each run consisted of 100 million iterations, and we set the costs so that Marxan ensured each portfolio met all the targets.

Comparative analyses

To assess whether the samples reflect global patterns, and whether the sample of grid squares is an improvement on the sample of national and sub-national sampling units in terms of representativeness, we undertook two analyses. The first analysis compared the percentage of the terrestrial realm covered by PAs with the percentage of each of the 1,000 Marxan outputs produced in the Stage 1 and Stage 2 analyses. We also compared them with a 1,000 randomly selected sets of national and sub-national sampling units (analogous to the Stage 1 Marxan analysis) and 1,000 randomly selected sets of the 100 km x 100 km sampling units (analogous to the Stage 2 Marxan analysis but based on all the sampling units across the global terrestrial realm, rather than those only found within the areas selected in the Stage 1 Marxan analysis). To do this, we developed a Python script (Van Rossum & Drake 2009) that added randomly selected sampling units until the set met or exceeded the combined area of the best respective Stage 1 or Stage 2 Marxan output. The PA data came from the publicly available WDPA dataset downloaded in May 2021 (UNEP-WCMC and IUCN 2021). It should be noted that this dataset does not include most PAs in China and India. We followed the standard protocol (UNEP-WCMC & IUCN 2016) by excluding PAs that are 'Proposed' or 'Not Reported' and UNESCO-MAB Biosphere Reserves. We also only used point data if the PA extent was recorded, converting it into a polygon of the required size by producing a buffer with the required radius around the point (UNEP-WCMC & IUCN 2016). We combined the PAs for each country and then used QGIS to calculate the total area of each grid square using the Clip and Dissolve functions, before using these data to calculate the area per Stage 1 administrative unit.

For the second analysis, we developed a Python script to calculate how many of the targets were met by each of the 1,000 Stage 1 and Stage 2 Marxan outputs and the 1,000 Stage 1 and Stage 2 randomly selected sets of sampling units.

Results

Stage 1 analysis

The best portfolio identified using Marxan consists of nine whole countries and territories, and 33 of the sub-national sampling units within another 16 countries (Figure 2a). These 25 countries and territories are: Argentina, Australia, Brazil, China, Democratic Republic of the Congo, Dominican Republic, France, French Polynesia, Greenland, Indonesia, India, Italy, Kazakhstan, Kiribati, Mexico, Mali, Papua New Guinea, Russia, Saudi Arabia, Sudan, South Georgia and the South Sandwich Islands, South Africa, Sweden, Tanzania and the United States of America. Only 17 of these 42 sampling units were selected in every one of the 1,000 portfolios identified by Marxan (Figure 2b), meaning that each of the other 25 sampling units could be swapped for sampling units containing similar amounts of the different features to produce similarly efficient portfolios.

Figure 2

Stage 2 analysis

The best portfolio identified by Marxan met all the targets and contained 2,231 of the 3,377 sampling units found within the Stage 1 sample, covering 10.89% of the terrestrial area (Figure 3a). The combined area of each of selected Stage 1 sampling units also selected in Stage 2 ranged between 31.5% for Australia and 100% for the Dominican Republic, with a median of 76.1% (Figure 3a); only 7 countries had less than half their Stage 1 area selected in Stage 2. The selection frequency results for Stage 2 mirrors this pattern, with low scores for sampling units where Marxan only needed to select a smaller proportion of the national and sub-national sampling units (Figure 3b).

Figure 3

Sampling comparison

The area of the terrestrial realm, excluding Antarctica, in our analysis is 135,008,972 km^2. The mean selected area of the 1,000 Stage 1 outputs was 16.82% (S.E. < 0.031) of the terrestrial realm and the mean selected area of the 1,000 Stage 2 outputs was 10.89%. (S.E. < 0.001). The publicly available WDPA data showed that 15.25% of the terrestrial realm is under protection, compared to a mean of 15.34% (S.E. = 0.069) for the 1000 Stage 1 outputs, and a mean of 15.97% (S.E. = 0.008) for the 1000 Stage 2 outputs. This compares to a mean area under protection for the Stage 1 random sets of sampling units of 15.26% (S.E. = 0.080) and for the Stage 2 random sets of sampling units of 15.22% (S.E. = 0.018).

The global area of the different features varied between < 0.001% for the Micronesia subregion and 94% for continents. All of the Stage 1 and Stage 2 Marxan outputs met all the 89 feature coverage targets, whereas the random sets for Stage 1 failed to meet a mean of 20.9 targets (S.E. = 0.152) and the random sets for Stage 2 failed to meet 15.9 targets (S.E. = 0.109). The mean percentage target met for the different factors for Stage 1 ranged between 12.94% of the elevation features and

22.18% of the subregion features and for Stage 2 ranged between 10.6% of the population density features and 14.8 % of the realms features (Table 1, Supplementary Material Table S4)

Discussion

Well-defined, measurable conservation targets, and accurate on-the-ground data to compare against them, are vital for driving forward progress towards our goal of a sustainable and ecologically healthy future for the planet (Mace et al. 2018). However, obtaining such data for the entire globe is a slow process (Juffe-Bignoli et al. 2016), a problem compounded by a lack of monitoring capacity in some high biodiversity countries (Stephenson et al. 2017). To augment the existing reporting systems, we propose using an approach pioneered by the Sampled Red List Index and similar projects (Butchart et al. 2007; Baillie et al. 2008). In this study we outline a framework for producing more accurate estimates of progress towards global conservation area targets by identifying a sample of countries and grid squares that are representative of the factors that shape total conservation area network extent and patterns of global biodiversity.

Choosing the factors and features

When choosing factors to include in our study, we sought to represent those that influence conservation area establishment across the world and broad patterns of biogeographic diversity. There is an established literature on the factors that shape global biodiversity patterns, so we can be confident that our final sample is representative at this global scale (Gaston & Spicer 2013). The literature on conservation area establishment factors is less well established, although we know that demographic, economic and governance factors are important (Mascia et al. 2014; Kroner et al. 2019), so differing social and socio-economic conditions will result in conservation area networks with differing extents (Bohn & Deacon 2000). More specifically, a number of previous studies have shown the importance of human population density and proxies of agricultural opportunity cost, such as elevation and landcover (Loucks et al. 2008; Joppa & Pfaff 2009) and the link between government effectiveness and wealth in determining conservation outcomes (Waldron et al. 2017).

Some factors that our expert group identified as potentially important could not be included because they have not been mapped at the global scale (Supplementary Material Table S1). Political and public support for conservation in each country, for example, may have an effect on conservation area establishment but global datasets were not available. This could be resolved in future through polling data and citizen science initiatives (McKinley et al. 2017). Collecting data on national land tenure systems might also be important, as this is likely to have a large impact on the extent of privately- and communally-managed PAs and OECMs in each country (Bingham et al. 2017). However, we did broadly account for this, as well as other potential factors, by using the geographic subregions dataset, ensuring representation of countries with shared legal, cultural and historical backgrounds. Another issue is that while some of our datasets are a snapshot of the current situation, conservation area coverage reflects both past and current circumstances, although governments often add or remove conservation areas in response to immediate conditions (Mascia & Pailler 2011; Radeloff et al. 2013).

Defining the sampling units and selecting the sample

The second key aim of our study was to ensure that the sampling approach was also a feasible basis for future data collection and study. Such data collection is resource intensive (Juffe-Bignoli et al. 2016), so we needed to balance between selecting a sample that was large enough to be sufficiently representative, but not so large as to make collecting data for every area in the sample unrealistic. We based Stage 1 of our framework on identifying countries and large within-country sub-regions to be included in our sample. This is because the nation state is the functional unit in conservation area data collection and reporting (Dallimer & Strange 2015) but large countries often have sub-national conservation agencies. Thus, by minimising the number of countries in our sample we also minimised the number of agencies and organisations involved in data collection. For the largest countries we also assumed that their conservation authorities would have a devolved structure, with national and sub-national agencies, hence our use of sub-national boundaries as sampling units. Research is needed to test these assumptions and better assess this trade-off between sample size and sampling effort.

The best portfolio identified in Stage 1 comprised nine whole countries and 33 administrative units in a further 16 countries. The selection frequency scores, which are based on how many times each sampling unit was selected in each of the runs, showed that only 17 of these sampling units were chosen every time (Figure 2b). The other sampling units are potentially interchangeable, which is important because if obtaining data from a particular country was impossible for logistical or political reasons, these units could be excluded and the analysis run again to find suitable replacements (Ball et al. 2009). The selection frequency results for the Stage 2 analysis also showed potentially interchangeable sampling units, mostly within the largest sub-national sampling units selected in Stage 1 containing additional land not needed to meet the targets (Figure 3). This Stage 2 result also shows the efficiency benefits of using a complementarity-based algorithm to select sample areas (Ball et al. 2009), as Marxan was able to meet the 10% targets for each feature in close to 10% of the sampling region, even though features belonging to different factors have different spatial distributions and extents. This involved selecting more than 10% for some features that are found in lots of sampling units and so are over-represented through meeting targets for other features (Table 1, Supplementary Material Tables S2 & S4). However, this is not expected to impact estimates of conservation area coverage based on the Stage 2 sample because the over-represented features include those with both high and low opportunity costs.

Comparing the analysis outputs in terms of their percentage in PAs is less straightforward for two reasons. First, the publicly available WDPA information used in this analysis does not include every PA provided by each country, as China and India have chosen to restrict some of their data (Bingham et al. 2019). This explains why the global PA coverage of the terrestrial realm calculated from our analysis of 15.25% is less than the 15.67% calculated by UNEP-WCMC based on all the data (UNEP-WCMC 2021). Second, our sampled approach was developed in part because of known limitations with the WDPA, so while comparing levels of protection based on the WDPA provides helpful insights, the results should not be seen as definitive. One clear trend from our results is that the Stage 1 and Stage 2 random sets of sampling units had near identical levels of PA coverage as the global figure. However, none of these random outputs also met all of the feature targets, so they would be less suitable for assessing the extent to which a sample of conservation areas represented biodiversity. The Stage 1 and Stage 2 outputs met all the targets, indicating that they could be used to measure conservation area representativeness, but the mean PA coverage for the Stage 2 outputs is 0.65% more than the global figure. More research is needed to understand this disparity, but in the short-term it would probably be prudent to modify conservation area estimates from this sampled approach based on the difference between the global and sample WDPA PA coverage.

Policy implications and wider relevance

Ongoing monitoring of progress towards conservation targets is essential but the required data are often lacking (Brooks et al. 2015). Resolving this will need more resources and capacity building (Stephenson et al. 2017), especially at the level of the nation state where most action is carried out and thus where guidance is most needed (Smith et al. 2009). At the same time, we need timely global estimates of progress to inform international policy. Our proposed solution for conservation area coverage is to identify a representative sample of countries and collect better data just from these, taking advantage of the availability of accurate information that has not yet been officially approved. Importantly, such a study would not need to report the estimated conservation area coverage for each country, avoiding problems associated with reporting unofficial national datasets.

In this study, we have shown that it is possible to identify such a representative sample of areas across the globe within a small enough number of countries to make data collection realistic. More research is needed on the trade-off between the percentage of the terrestrial realm included in the sample and the number of countries and sub-national administrative units required to provide the data. Nonetheless, we have demonstrated proof of concept and identified a sample of reasonable size that is also a realistic basis for data collection. Our sampling approach is also likely to be suitable for marine conservation areas, as the existing literature suggests that their distributions are similarly impacted by comparable social and socio-economic factors (Devillers et al. 2020).

The next step is to collect data on conservation areas within the sample we have identified, working with local experts and non-government sources. This should then be used to develop improved global conservation area metrics, measuring coverage, connectivity levels (Saura et al. 2018) and how well these conservation area networks represent biodiversity (Butchart et al. 2015). This will be particularly important for OECMs, as national- and regional-scale data suggest they enhance PA network connectivity and cover different biodiversity elements (Dudley et al. 2018). More broadly, we hope that this sampling approach could be used to produce global estimates of a range of other conservation metrics, related to costs and effectiveness (Coad et al. 2015; Iacona et al. 2018) and social impacts, governance and equity (Dawson et al. 2018; Naidoo et al. 2019). In doing so, our sampling approach could help monitor progress towards meeting a number of international conservation targets and policies.

Acknowledgements

We would like to thank Lincoln Fishpool and Mike Hoffmann for their role in identifying the factors used in the analysis. We acknowledge funding from the University of Kent’s PhD Scholarship and Faculty Grant programmes, and UK Research and Innovation's Global Challenges Research Fund (UKRI GCRF) through the Trade, Development and the Environment Hub project (project number ES/S008160/1).

References

Baillie JEM, Collen B, Amin R, Akcakaya HR, Butchart SHM, Brummitt N, Meagher TR, Ram M, Hilton-Taylor C, Mace GM. 2008. Toward monitoring global biodiversity. Conservation Letters **1**:18–26.

Ball I, Possingham H, Watts M. 2009. Marxan and relatives: Software for spatial conservation prioritization. Pages 185–195 in Moilanen A, Wilson K, Possingham H, editors. Spatial conservation prioritisation: quantitative methods and computational tools. Oxford University Press, Oxford, UK.

Bingham H, Fitzsimons J A, Redford K, Mitchell BA, Bezaury-Creel J, Cumming TL. 2017. Privately protected areas: advances and challenges in guidance, policy and documentation. PARKS **23**:13–28.

Bingham HC et al. 2019. Sixty years of tracking conservation progress using the World Database on Protected Areas. Nature Ecology & Evolution **3**:737–743.

Brooks TM, Butchart SHM, Cox NA, Heath M, Hilton-Taylor C, Hoffmann M, Kingston N, Rodríguez JP, Stuart SN, Smart J. 2015. Harnessing biodiversity and conservation knowledge products to track the Aichi Targets and Sustainable Development Goals. Biodiversity **16**:157–174.

Bruijnzeel LA, Veneklaas EJ. 1998. Climatic Conditions and Tropical Montane Forest Productivity: The Fog Has Not Lifted Yet. Ecology **79**:3–9.

Butchart SHM et al. 2015. Shortfalls and Solutions for Meeting National and Global Conservation Area Targets. Conservation Letters **8**:329–337.

Butchart SHM, Akçakaya HR, Chanson J, Baillie JEM, Collen B, Quader S, Turner WR, Amin R, Stuart SN, Hilton-Taylor C. 2007. Improvements to the Red List Index. PLOS ONE **2**:e140.

CBD. 2010. X/2.Strategic Plan for Biodiversity 2011-2020. Available from https://www.cbd.int/decision/cop/?id=12268.

CBD. 2021. First draft of the Post-2020 Global Biodiversity Framework. CBD/WG2020/3/3. Convention on Biological Diversity.

Clements HS, Selinske MJ, Archibald CL, Cooke B, Fitzsimons JA, Groce JE, Torabi N, Hardy MJ. 2018. Fairness and Transparency Are Required for the Inclusion of Privately Protected Areas in Publicly Accessible Conservation Databases. Land **7**:96.

Coad L et al. 2015. Measuring impact of protected area management interventions: current and future use of the Global Database of Protected Area Management Effectiveness. Philosophical Transactions of the Royal Society B: Biological Sciences **370**:20140281.

Corrigan C, Bingham H, Shi Y, Lewis E, Chauvenet A, Kingston N. 2018. Quantifying the contribution to biodiversity conservation of protected areas governed by indigenous peoples and local communities. Biological Conservation **227**:403–412.

Dallimer M, Strange N. 2015. Why socio-political borders and boundaries matter in conservation. Trends in Ecology & Evolution **30**:132–139.

Dawson N, Martin A, Danielsen F. 2018. Assessing Equity in Protected Area Governance: Approaches to Promote Just and Effective Conservation. Conservation Letters **11**:e12388.

Devillers R, Pressey RL, Ward TJ, Grech A, Kittinger JN, Edgar GJ, Watson RA. 2020. Residual marine protected areas five years on: Are we still favouring ease of establishment over need for protection? Aquatic Conservation: Marine and Freshwater Ecosystems **30**:1758–1764.

Dudley N, Jonas H, Nelson F, Parrish J, Pyhälä A, Stolton S, Watson JEM. 2018. The essential role of other effective area-based conservation measures in achieving big bold conservation targets. Global Ecology and Conservation **15**:e00424.

ESA GlobCover Project. 2009. GlobCover 2009 (Global Land Cover Map). Available from http://due.esrin.esa.int/page_globcover.php.

GADM. 2018. Global Administrative Areas. Available from https://gadm.org/data.html.

Gaston KJ, Spicer JI. 2013. Biodiversity: An Introduction. John Wiley & Sons.

Iacona GD et al. 2018. Standardized reporting of the costs of management interventions for biodiversity conservation. Conservation Biology **32**:979–988.

Joppa LN, Pfaff A. 2009. High and Far: Biases in the Location of Protected Areas. PLOS ONE **4**:e8273.

Juffe-Bignoli D et al. 2016. Assessing the Cost of Global Biodiversity and Conservation Knowledge. PLOS ONE **11**:e0160640.

Kukkala AS, Moilanen A. 2013. Core concepts of spatial prioritisation in systematic conservation planning. Biological Reviews **88**:443–464.

Lewis E, MacSharry B, Juffe-Bignoli D, Harris N, Burrows G, Kingston N, Burgess ND. 2019. Dynamics in the global protected-area estate since 2004. Conservation Biology **33**:570–579.

Lewis OT, Senior MJM. 2011. Assessing conservation status and trends for the world's butterflies: the Sampled Red List Index approach. Journal of Insect Conservation **15**:121–128.

Linkie M, Rood E, Smith RJ. 2010. Modelling the effectiveness of enforcement strategies for avoiding tropical deforestation in Kerinci Seblat National Park, Sumatra. Biodiversity and Conservation **19**:973–984.

Loucks C, Ricketts TH, Naidoo R, Lamoreux J, Hoekstra J. 2008. Explaining the global pattern of protected area coverage: relative importance of vertebrate biodiversity, human activities and agricultural suitability. Journal of Biogeography **35**:1337–1348.

Mace GM, Barrett M, Burgess ND, Cornell SE, Freeman R, Grooten M, Purvis A. 2018. Aiming higher to bend the curve of biodiversity loss. Nature Sustainability **1**:448–451.

Mascia MB, Pailler S. 2011. Protected area downgrading, downsizing, and degazettement (PADDD) and its conservation implications. Conservation Letters **4**:9–20.

McKinley DC et al. 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. Biological Conservation **208**:15–28.

Naidoo R et al. 2019. Evaluating the impacts of protected areas on human well-being across the developing world. Science Advances **5**:eaav3006.

Naidoo R, Balmford A, Ferraro PJ, Polasky S, Ricketts TH, Rouget M. 2006. Integrating economic costs into conservation planning. Trends in Ecology & Evolution **21**:681–687.

Nhancale B, Smith R. 2011. The influence of planning unit characteristics on the efficiency and spatial pattern of systematic conservation planning assessments. Biodiversity and Conservation **20**:1821–1835.

Olson DM et al. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience **51**:933–938.

Palfrey R, Oldekop JA, Holmes G. 2022. Privately protected areas increase global protected area coverage and connectivity. Nature Ecology & Evolution **6**:730–737.

QGIS. 2019. QGIS Geographic Information System. QGIS Association. Available from http://www.qgis.org.

Radeloff VC, Beaudry F, Brooks TM, Butsic V, Dubinin M, Kuemmerle T, Pidgeon AM. 2013. Hot moments for biodiversity conservation. Conservation Letters **6**:58–65.

Saura S, Bertzky B, Bastin L, Battistella L, Mandrici A, Dubois G. 2018. Protected area connectivity: Shortfalls in global targets and country-level priorities. Biological Conservation **219**:53–67.

Smith R. 2019. The CLUZ plugin for QGIS: designing conservation area systems and other ecological networks. Research Ideas and Outcomes **5**:e33510.

Smith RJ, Veríssimo D, Leader-Williams N, Cowling RM, Knight AT. 2009. Let the locals lead. Nature **462**:280–281.

Stephenson PJ et al. 2017. Unblocking the flow of biodiversity data for decision-making in Africa. Biological Conservation **213**:335–340.

UNEP-WCMC. 2019. User Manual for the World Database on Protected Areas and world database on other effective area based conservation measures: 1.6. UNEP-WCMC, Cambridge, UK. Available from http://wcmc.io/WDPA_Manual.

UNEP-WCMC. 2021. Protected areas map of the world, July 2021 Available at: www.protectedplanet.net.

UNEP-WCMC and IUCN. 2021. Protected Planet: The World Database on Protected Areas, [May 2021]. Available at: www.protectedplanet.net. UNEP-WCMC and IUCN, Cambridge, UK.

UNEP-WCMC, IUCN. 2016. Protected Planet Report 2016. UNEP-WCMC and IUCN, Cambridge UK and Gland, Switzerland.

UNPD. 2013. World Population Prospects: The 2012 Revision. Department of Economic and Social Affairs, United Nations Population Division. Available from https://population.un.org/wpp/Download/Standard/Population/ (accessed November 28, 2019).

UNSD. 2019. Methodology: Standard country or area codes for statistical use (M49). United Nations Statistics Division. Available from https://unstats.un.org/unsd/methodology/m49/ (accessed November 28, 2019).

Van Rossum G, Drake F. 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Visconti P, Butchart SHM, Brooks TM, Langhammer PF, Marnewick D, Vergara S, Yanosky A, Watson JEM. 2019. Protected area targets post-2020. Science **364**:239–241.

Visconti P, Marco MD, Álvarez-Romero JG, Januchowski-Hartley SR, Pressey RL, Weeks R, Rondinini C. 2013. Effects of Errors and Gaps in Spatial Data Sets on Assessment of Conservation Progress. Conservation Biology **27**:1000–1010.

Waldron A, Miller DC, Redding D, Mooers A, Kuhn TS, Nibbelink N, Roberts JT, Tobias JA, Gittleman JL. 2017. Reductions in global biodiversity loss predicted from conservation spending. Nature **551**:364–367.

Watson JEM, Dudley N, Segan DB, Hockings M. 2014. The performance and potential of protected areas. Nature **515**:67–73. Nature Publishing Group.

Whittaker RJ, Fernandez-Palacios JM. 2007. Island Biogeography: Ecology, Evolution, and Conservation. Oxford University Press, Oxford, UK.

World Bank. 2019a. Worldwide Governance Indicators DataBank. Available from https://databank.worldbank.org/source/worldwide-governance-indicators (accessed November 28, 2019).

World Bank. 2019b. GDP (current US$). Available from https://data.worldbank.org/indicator/NY.GDP.MKTP.CD (accessed November 28, 2019).

Table 1. Details of the factors used in the analysis that are likely to shape total conservation area network extent and patterns of global biodiversity, the extent of the feature with the smallest and largest area for each factor in the terrestrial realm, and the per factor mean percentage coverage of each feature identified in the Stage 1 and Stage 2 best portfolios.

Factor	**Number of features**	**Global area of feature with smallest extent (%)**	**Global area of feature with largest extent (%)**	**Stage 1 mean of % of each feature in the selected sample**	**Stage 2 mean of % of each feature in the selected sample**
Biomes	16	0.24	20.67	15.23	10.92
Elevation	5	5.38	41.24	12.94	10.73
Government effectiveness	4	17.34	35.99	15.28	11.01
Income	4	10.66	44.91	15.45	10.60
Islands and continents	5	0.36	94.23	14.95	11.61
Landcover	12	0.01	19.41	17.26	11.04
Latitude	7	0.16	23.74	17.79	13.87
Population density	5	0.79	39.82	13.76	10.59
Realms	9	<0.01	38.95	21.79	14.82
Subregions	22	<0.01	15.95	22.18	13.87

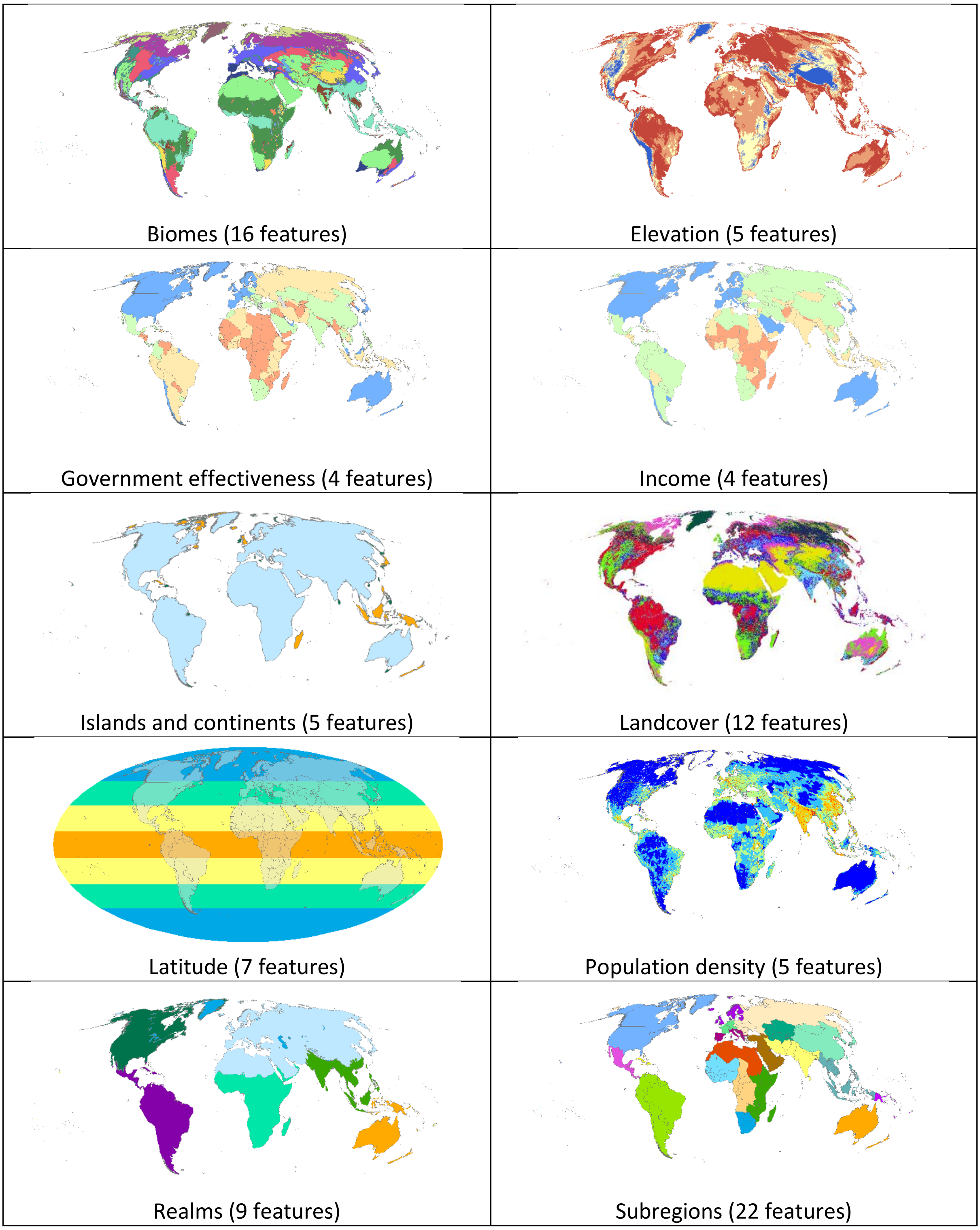
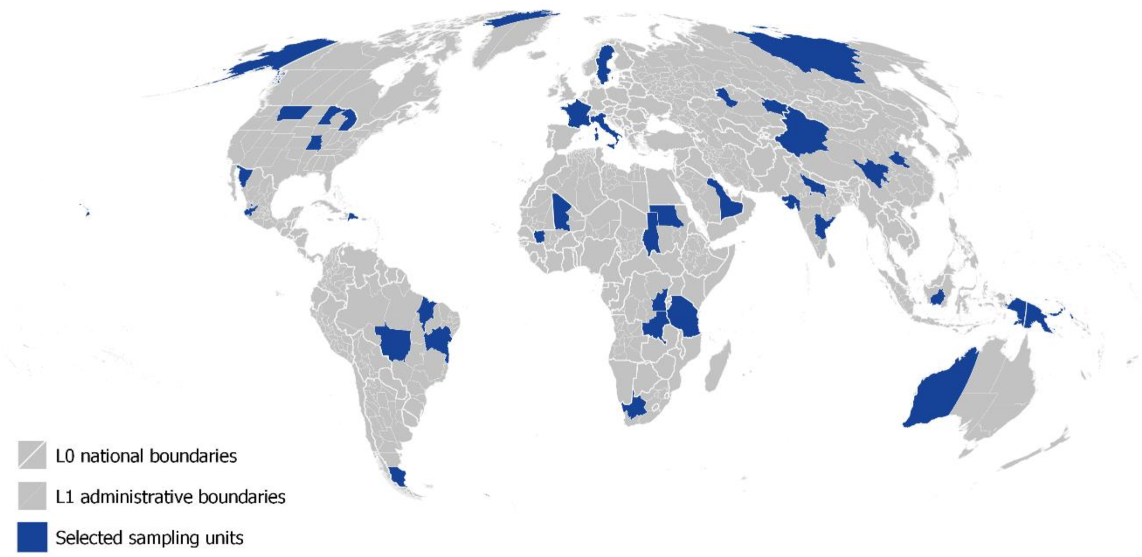


Figure 1: Maps of the ten factors used in the analysis to identify a representative sample of countries demonstrating drivers of conservation area extent and drivers of global biodiversity patterns. Details of the features that make up each factor are given in Supplementary Material Table S4.

(a)



(b)

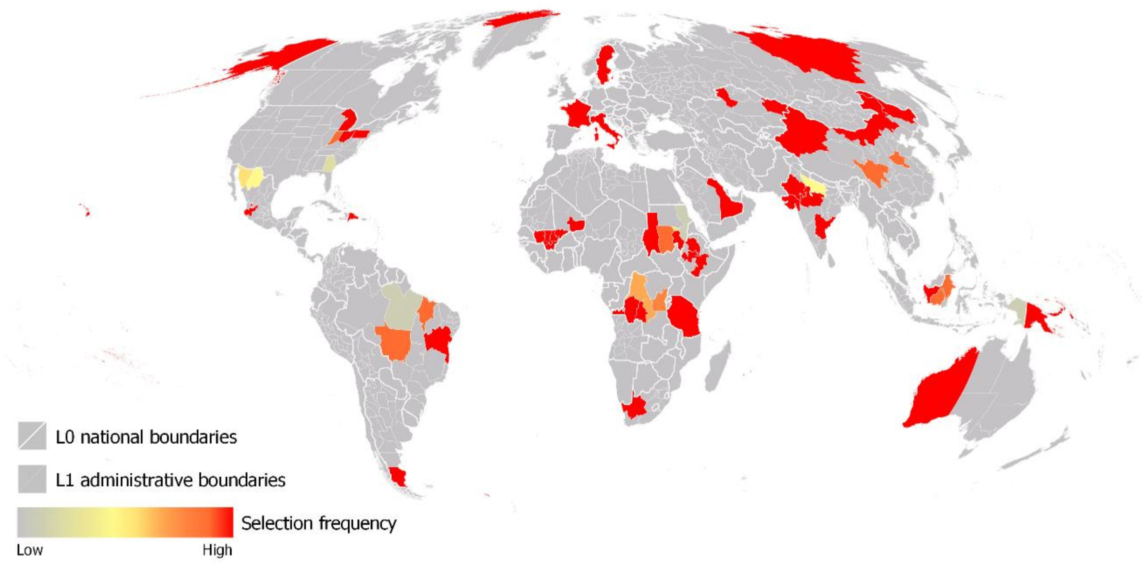
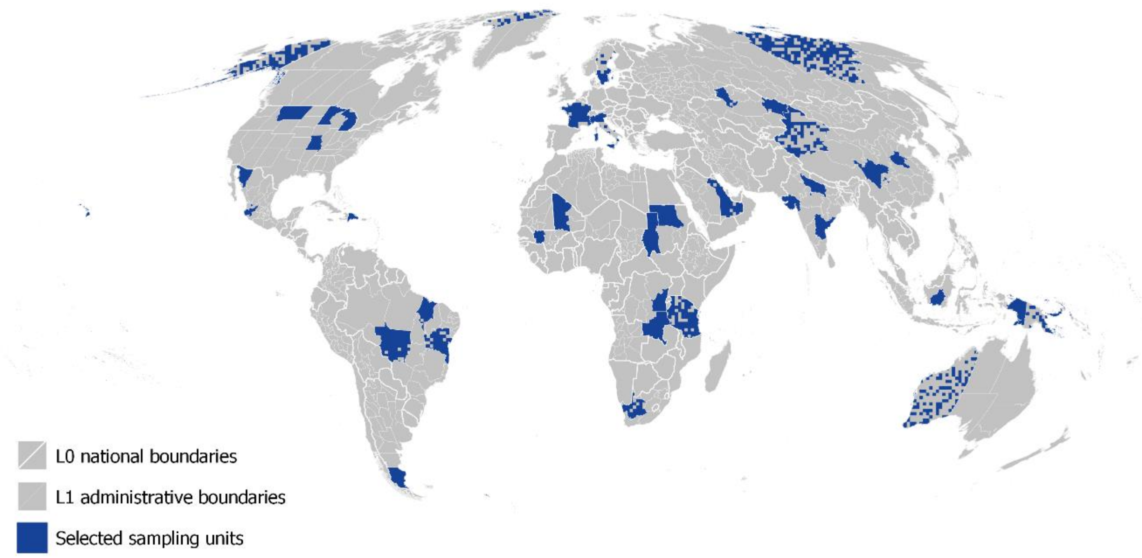


Figure 2: (a) Sample of countries (national sampling units) and administrative units (sub-national sampling units) that meet 10% targets, selected based on 1,000 Marxan runs and selecting the result with the smallest number of sampling units, most even spread across the continents and with sampling units with the highest mean selection frequency. (b) Selection frequency scores from Marxan showing the number of times each sampling unit was selected across the 1,000 runs used to identify the sample.

(a)



(b)

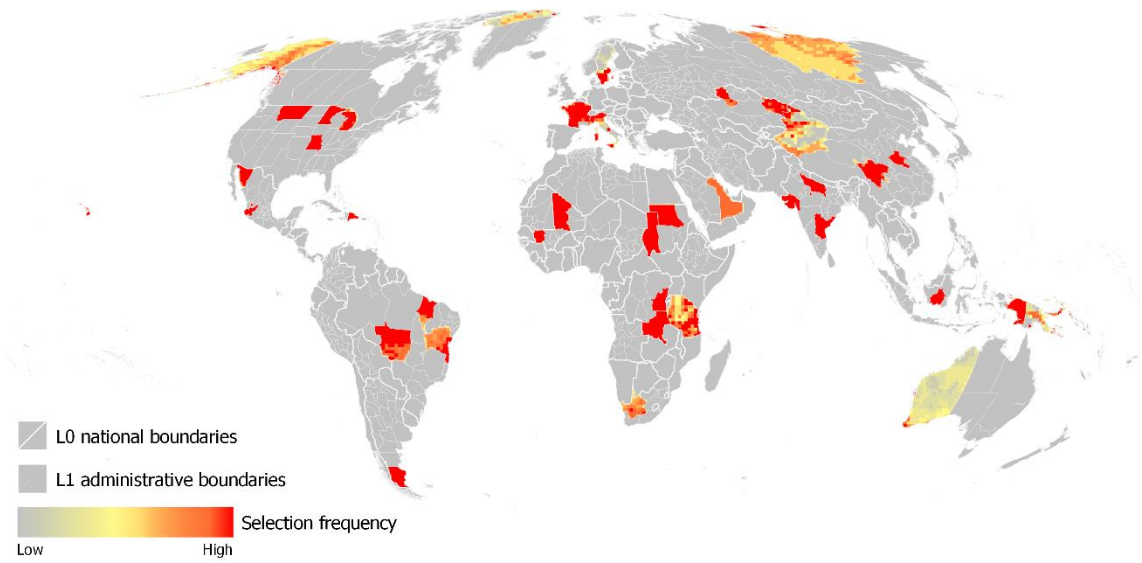


Figure 3. (a) Sample of 100 x 100 km grid squares found in the focal countries (national sampling units) and administrative units (sub-national sampling units) selected by Marxan that best meets 10% targets for biogeographic and conservation area extent factors while minimising sample area. (b) Selection frequency scores from Marxan showing the number of times each sampling unit was selected across the 1,000 runs used to identify the best sample.

Supplementary Materials: developing a framework to improve global estimates of conservation area coverage

Initial list of factors to be included in spatial prioritisation

Table S1: Factors discussed in our expert workshop as potential drivers of conservation area establishment and drivers of biodiversity patterns

African, Caribbean and Pacific Group of States	Latitude
Age of PA network	Legal system type
Carbon payments	Major habitat types
Climate vulnerability indices	PA investment
Completeness of WDPA country records	PA management effectiveness
Continents	PA management record
Corruption	PA visitor numbers
Degraded and pristine areas	Political groupings (e.g. ex-Soviet, ex-colonial)
Ecoregions	Political stability
Endemism	Rates of forest loss
Freshwater	Rates of habitat conversion
Islands	Religious groupings
Land tenure	Sacred sites
Landcover trends	Size of country
Language groups	Within country variability

Description of the factors included in spatial prioritisation

Table S2. Factors that drive global biodiversity patterns and conservation area extent used in our analysis.

Factor	**Drivers of conservation area extent**	**search**
Biomes	Conservation area extent is higher in biomes with less land suitable for agriculture, such as deserts, and rock and ice (Hoekstra et al. 2005).	Biodiversity differs greatly between biomes, with ecosystem types sharing similar species compositions (Gaston & Spicer 2013).
Elevation	Conservation area extent tends to increase at higher elevations (Joppa & Pfaff 2009).	Species composition varies across elevation gradients (Gaston & Spicer 2013).
Government effectiveness	Stable countries with higher bureaucratic quality have greater capacity to expand conservation area networks (Laurance 2004).	
Income (per capita)	Wealthier countries have more resources to fund the expansion of conservation area networks (Waldron et al. 2013).	
Islands and continents		Islands are often geographically and biologically distinct, with unique and highly threatened biodiversity (Gaston & Spicer 2013).
Landcover	Conservation area extent differs between landcover types (Joppa & Pfaff 2009).	Species composition varies between vegetation types and land-uses (Gaston & Spicer 2013).
Latitude	Conservation area extent is higher at latitudes with less land suitable for agriculture, such as closer to the poles (Hoekstra et al. 2005).	Species composition shows strong latitudinal gradients (Gaston & Spicer 2013).
Population density	Conservation area extent is lower in regions with high human population density (Joppa & Pfaff 2009).	
Realms		Biodiversity shows strong biogeographic patterns at the continental scale (Gaston & Spicer 2013).
Subregions	Sub-sections of continents have relatively similar histories, economies and legislative frameworks (Siegfried et al. 1998).	Biodiversity shows strong biogeographic patterns at the sub-continental scale (Gaston & Spicer 2013).

Sensitivity analysis to set targets

The sensitivity analysis explored the trade-off between the area of the terrestrial realm selected to be a potential sample for future studies on conservation area extent and representativeness, and the number of sampling units selected (where each sampling unit was a country or, for countries with an area ≥1 million km^2, sub-national units such as provinces and states). This was based on the premise that selecting a larger percentage of the planet would produce a more robust sample but selecting more countries and provinces would increase the time and resources needed to collect the conservation area data. So, we used the conservation planning system developed for Stage 1 to run eight Marxan analysis using the same percentage target for each feature in each analysis. These different targets were 1%, 2%, 5%, 10%, 20%, 30%, 40% and 50% of the total extent of each feature, and each analysis consisted of 100 runs of 10,000,000 iterations. We used a Boundary Length Modifier of 1.5 (Ball et al. 2009), a value that we determined through testing to best ensure that Marxan chose enough sub-national units from the same countries to meet the targets. We then counted the number of whole countries and the number of sampling units in the 'best' solution for each of the eight analyses.

The number of sampling units selected by Marxan to meet the targets for the 89 conservation features ranged from 23 for the 1% targets to 206 for the 40% targets (Table S3). The number of sampling units more than doubled when comparing results from using 10% and 20% targets, with a levelling off when the targets were ≥ 30%. Based on these results, we decided to use 10% targets for the main analyses.

Table S3. The number of sampling units and countries selected to meet specific percentage targets for each of the 89 features. Sampling units consisted of whole countries for nations with an area <1 million km^2 and highest level sub-national units (provinces, states, etc) for countries with an area ≥1 million km^2.

Conservation feature targets (%)	Number of sampling units selected	Number of countries selected
1	23	22
2	24	22
5	30	24
10	50	25
20	117	27
30	204	32
40	206	32
50	205	32

Marxan analysis

We used Marxan for a two-stage analysis: Stage 1 identified the minimum set of countries needed to meet targets for each feature and Stage 2 identified sets of 10,000 km^2 grid squares that meet these targets within this subset of countries. This involved setting Species Penalty Factors to ensure that all the targets were met and we used a value of 10 for every feature for both analyses, as our initial testing showed this value produced efficient results.

For the Stage 1 analysis, we sought to select a set of national and sub-national sampling units that represented all the features, while also minimising the number of countries selected. To do this we set the combined sampling unit cost of each country as 1, so that selecting more countries was more costly. To account for the larger nations being split into several sampling units, based on the sub-national administrative units, we set the sampling unit costs as the inverse of the number of sub-national units in the country. For example, each of South Africa's nine provinces had a sampling unit cost of 0.111. In addition, we needed to ensure that Marxan met targets by selecting the sub-national sampling units from the same countries whenever possible. To do this we manipulated the Marxan boundary cost file so it appeared that every sub-national sampling unit in the same country shared a boundary. This meant that if Marxan selected one sub-national sampling unit in a particular country then it would be less costly to select subsequent sub-national sampling units from the same country. To make sure this cost would be the same per country, we set the boundary length equal to the inverse of the number of different sub-national boundary pairs in each country so, for instance, the nine provinces in South Africa produced 45 combinations of sub-national pairs and so the boundary length was 0.0222. This manipulation of the boundary cost data has been used in previous studies to ensure that certain sampling units are more likely to be selected together, even when they are not physically adjacent (Possingham et al. 2005; Hermoso et al. 2011).

We ran the Stage 1 analysis using Marxan, which consisted of 1,000 runs of 10,000,000 iterations and used a Boundary Length Modifier value of 1.5. These parameters were selected to produce results that minimised the number of countries and sampling units selected. Of the resulting 1,000 portfolios, 284 had equally low costs, i.e., contained exactly the same number of countries and sampling units. To produce our final list of sample countries we therefore needed to develop our own scoring system to choose between these low-cost portfolios. We did this by first selecting portfolios with the most even spread of countries selected across the continents to further improve the representativeness of the sample. We then identified the portfolios containing sampling units that were selected most often in the 1,000 runs, based on calculating their mean selection frequency score. This provided us with our final set of national and sub-national sampling units that were then used in the Stage 2 analysis.

Thus, our first step in Stage 2 was to update the planning system to specify in CLUZ that all of the 100 km x 100 km sampling units found outside the national and sub-national regions selected in Stage 1 should be excluded from subsequent Marxan analyses. The Stage 2 Marxan analysis also consisted of 1,000 runs but we used 100,000,000 iterations after carrying out a sensitivity analysis that showed more iterations were needed to produce efficient results because Stage 2 involved a larger number of planning units than Stage 1. For Stage 2 we used a Boundary Length Modifier value of 0 because we were not interested in selecting adjacent sampling units. For this finer-scale analysis we used the sampling unit area as the cost metric and did not account for boundary cost. This was because in Stage 2 we were simply seeking to identify the smallest area of land needed to meet the targets, as the logistical cost of collecting PA coverage data is not affected by whether the Stage 2 sampling units neighbour each other.

References

Ball, I., Possingham, H. & Watts, M. (2009). Marxan and relatives: Software for spatial conservation prioritization. In: *Spatial conservation prioritisation: quantitative methods and computational tools* (eds. Moilanen, A., Wilson, K. & Possingham, H.). Oxford University Press, Oxford, UK, pp. 185–195.

Gaston, K.J. & Spicer, J.I. (2013). *Biodiversity: An Introduction*. John Wiley & Sons.

Hoekstra, J.M., Boucher, T.M., Ricketts, T.H. & Roberts, C. (2005). Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters*, 8, 23–29.

Joppa, L.N. & Pfaff, A. (2009). High and Far: Biases in the Location of Protected Areas. *PLOS ONE*, 4, e8273.

Laurance, W.F. (2004). The perils of payoff: corruption as a threat to global biodiversity. *Trends in Ecology & Evolution*, 19, 399–401.

Siegfried, W.R., Benn, G.A. & Gelderblom, C.M. (1998). Regional assessment and conservation implications of landscape characteristics of African national parks. *Biological Conservation*, 84, 131–140.

Waldron, A., Mooers, A.O., Miller, D.C., Nibbelink, N., Redding, D., Kuhn, T.S., Roberts, J.T. & Gittleman, J.L. (2013). Targeting global conservation funding to limit immediate biodiversity declines. *PNAS*, 110, 12144–12148.

Details of the features used in the analysis and their levels of representation in the selected samples

Table S4: Details of all the features used in the analysis, their total extent, the proportion of the Stage 1 and Stage 2 best outputs covered by each, and the proportion of the terrestrial realm covered by the Stage 1 and Stage 2 best outputs.

Category	Feature	Global extent (km^2)	Terrestrial realm covered by feature (%)	Stage 1 sample covered by feature (%)	Stage 2 sample covered by feature (%)	Global extent found in Stage1 sample (%)	Global extent found in Stage 2 sample (%)
Biomes	Tropical and subtropical moist broadleaf forest	19,847,759	14.67	10.78	13.49	11.40	10.00
Biomes	Tropical and subtropical dry broadleaf forest	3,017,092	2.23	1.92	2.54	13.36	12.32
Biomes	Tropical and subtropical coniferous forest	711,296	0.53	0.34	0.48	9.98	9.98
Biomes	Temperate broadleaf and mixed forest	12,772,448	9.44	6.64	8.68	10.91	10.01
Biomes	Temperate conifer forest	4,075,868	3.01	1.97	2.77	10.16	10.01
Biomes	Boreal Forest/taiga	15,046,636	11.12	14.11	10.99	19.68	10.73
Biomes	Tropical and Subtropical grasslands, savannas and shrubland	20,285,917	14.99	14.04	15.73	14.53	11.29
Biomes	Temperate grasslands, savannas and shrublands	10,098,291	7.46	4.92	6.86	10.22	10.00
Biomes	Flooded grasslands and savannas	1,094,839	0.81	0.63	0.83	12.03	11.58
Biomes	Montane grasslands and shrublands	5,203,199	3.85	3.28	3.58	13.23	10.19
Biomes	Tundra	8,206,496	6.07	9.22	7.90	23.59	14.10
Biomes	Mediterranean forests, woodlands and scrub	3,210,402	2.37	3.59	2.34	23.50	10.00
Biomes	Deserts and xeric shrublands	27,969,796	20.67	25.21	21.03	18.92	11.13
Biomes	Mangroves	320,823	0.24	0.21	0.23	13.99	10.95
Biomes	Inland water	1,039,692	0.77	0.84	0.96	16.96	12.47
Biomes	Rock and ice	1,973,619	1.46	1.99	1.34	21.16	10.05
Realms	Australasia	9,232,561	6.82	15.77	8.78	35.85	13.82
Realms	Antarctic	11,159	0.01	0.02	0.01	33.00	15.72

Realms	Afrotropics	21,769,183	16.09	13.08	16.22	12.62	10.99
Realms	Indomalay	8,513,981	6.29	4.08	5.79	10.06	10.00
Realms	Nearctic	20,398,341	15.08	12.92	13.86	13.30	10.00
Realms	Neotropics	19,368,174	14.31	10.32	13.16	11.19	10.00
Realms	Oceania	43,247	0.03	0.09	0.10	45.13	41.67
Realms	Palearctic	52,705,510	38.95	40.89	39.84	16.28	11.07
Realms	Snow and ice	2,832,017	2.09	2.52	2.01	18.66	10.15
Elevation	0 – 299 m	55,813,693	41.25	37.82	39.26	14.23	11.61
Elevation	300 – 799 m	43,299,328	32.00	35.59	33.97	17.26	11.88
Elevation	800 – 1399 m	19,827,397	14.65	16.57	15.77	17.54	10.07
Elevation	1400 – 1999 m	7,279,095	5.38	4.47	4.95	12.88	10.08
Elevation	>= 2000 m	8,627,189	6.38	5.32	5.86	12.94	10.00
Islands and continents	< 1,000 km^2	487,462	0.36	0.29	0.33	12.28	12.22
Islands and continents	1,000 to 10,000 km^2	660,808	0.49	0.60	0.50	18.95	10.02
Islands and continents	10,000 to 100,000 km^2	1,621,613	1.20	0.86	1.11	11.11	11.27
Islands and continents	100,000 to 1,000,000 km^2	5,009,245	3.70	4.04	3.95	16.92	10.84
Islands and continents	> 1,000,000 km^2	127,492,420	94.23	94.20	94.10	15.51	13.69
Landcover	Croplands	10,044,523	7.42	7.92	9.62	16.54	10.47
Landcover	Croplands mosaic	17,948,478	13.27	10.76	12.81	12.59	10.00
Landcover	Closed forest	25,436,142	18.80	13.85	17.29	11.43	13.02
Landcover	Open forest	12,323,377	9.11	12.82	11.32	21.85	10.50
Landcover	Mosaic grassland/shrubland	26,265,135	19.41	20.05	18.32	16.02	10.16
Landcover	Sparse vegetation	13,551,920	10.02	12.60	9.22	19.52	10.66
Landcover	Flooded forest/grassland	1,902,386	1.41	1.53	1.45	16.92	10.21
Landcover	Artificial surfaces	317,365	0.23	0.20	0.22	12.94	10.64
Landcover	Bare areas	21,608,578	15.97	15.35	15.53	14.91	10.18
Landcover	Water bodies	2,980,599	2.20	2.08	2.18	14.68	10.21
Landcover	Snow and ice	2,913,595	2.15	2.82	2.02	20.29	15.85

Landcover	No data	14,186	0.01	0.02	0.02	29.48	10.53
Latitude	50N to 90N	31,826,862	23.52	27.39	22.89	18.06	10.00
Latitude	30N to 50N	32,126,360	23.74	20.68	21.83	13.51	11.75
Latitude	10N to 30N	26,501,375	19.59	15.65	21.21	12.40	11.31
Latitude	-10S to 10N	20,617,051	15.24	13.90	16.00	14.15	11.09
Latitude	-30S to -10S	18,842,279	13.93	18.04	14.11	20.10	10.12
Latitude	-50S to -30S	5,146,310	3.80	4.04	3.54	16.48	28.55
Latitude	-90S to -50S	214,058	0.16	0.30	0.41	29.83	14.29
Income classification	Low income	14,417,961	10.66	11.06	14.00	16.10	10.56
Income classification	Lower middle income	22,038,475	16.29	12.63	15.92	12.03	10.00
Income classification	Upper middle income	60,767,325	44.91	40.77	41.32	14.08	11.07
Income classification	High income	38,082,584	28.15	35.54	28.76	19.59	10.78
Population density	0 to 0.9	53,883,215	39.82	46.44	39.53	18.09	11.62
Population density	1 to 9.9	39,359,881	29.09	28.03	31.21	14.95	10.02
Population density	10 to 99.9	27,781,643	20.53	16.15	18.89	12.20	11.12
Population density	100 to 999.9	9,292,487	6.87	5.54	7.02	12.52	10.13
Population density	1000+	1,070,380	0.79	0.56	0.73	11.05	10.00
Govt. Effectiveness	0 - 24.9	23,463,373	17.34	11.34	15.95	10.15	10.42
Govt. Effectiveness	25 - 49.9	48,702,809	35.99	33.62	34.59	14.49	13.09
Govt. Effectiveness	50 - 74.9	28,037,538	20.72	23.55	24.89	17.64	10.23
Govt. Effectiveness	75 – 100	35,102,625	25.94	31.49	24.58	18.83	10.30
Subregions	Australia and New Zealand	7,985,635	5.90	12.08	5.43	31.75	10.00
Subregions	Caribbean	233,427	0.17	0.23	0.33	20.70	20.70
Subregions	Central America	2,481,651	1.83	1.25	1.70	10.57	10.11
Subregions	Central Asia	4,380,003	3.24	2.29	2.98	10.99	10.02
Subregions	Eastern Africa	7,049,679	5.21	4.52	4.79	13.45	10.00
Subregions	Eastern Asia	11,598,707	8.57	10.87	9.46	19.68	11.95
Subregions	Eastern Europe	18,604,967	13.75	14.62	12.64	16.49	10.00

Subregions	Melanesia	544,908	0.40	2.22	1.38	85.63	34.56
Subregions	Micronesia	3,576	0.00	0.00	0.00	27.68	10.26
Subregions	Middle Africa	6,608,246	4.88	3.55	5.01	11.27	11.25
Subregions	Northern Africa	7,647,985	5.65	4.80	6.74	13.18	12.96
Subregions	Northern America	21,581,549	15.95	14.61	14.67	14.21	10.00
Subregions	Northern Europe	1,803,994	1.33	2.13	1.23	24.83	10.00
Subregions	Polynesia	8,613	0.01	0.02	0.01	46.37	18.77
Subregions	South America	17,845,353	13.19	9.80	12.40	11.53	10.23
Subregions	South-eastern Asia	4,483,416	3.31	2.24	3.05	10.50	10.01
Subregions	Southern Africa	2,681,065	1.98	1.73	1.82	13.54	10.00
Subregions	Southern Asia	6,710,677	4.96	3.35	4.75	10.49	10.41
Subregions	Southern Europe	1,316,461	0.97	1.43	1.11	22.85	13.23
Subregions	Western Africa	6,082,789	4.50	2.99	4.20	10.33	10.06
Subregions	Western Asia	4,528,985	3.35	2.62	3.08	12.16	10.00
Subregions	Western Europe	1,102,673	0.81	2.62	3.24	49.81	40.67