# Nature4SDGs Integrated Dataset

---

## Table of contents

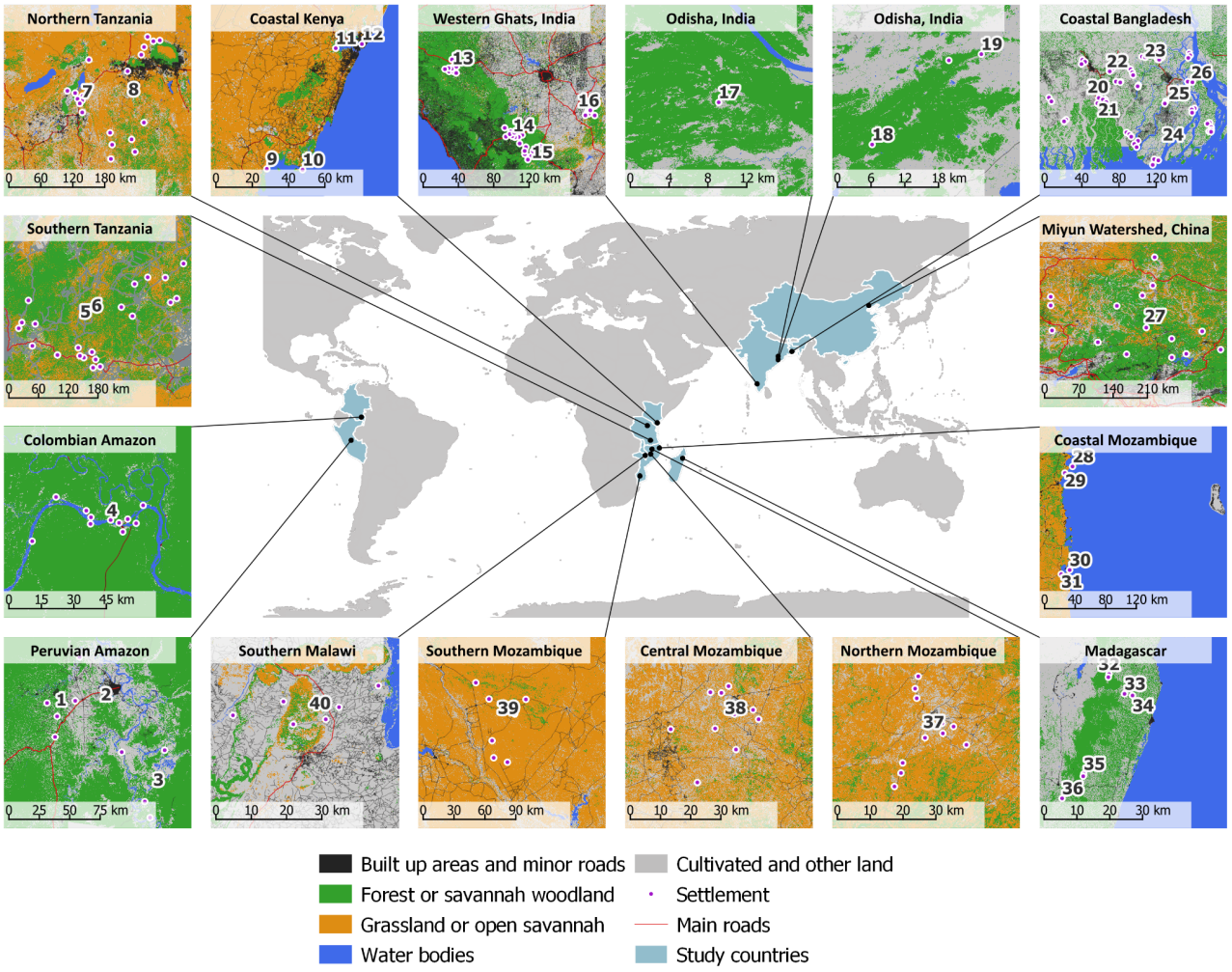
# 1. Introduction

This dataset was compiled for analyses in the research project ‘Nature's contribution to poverty alleviation, human wellbeing and the SDGs’ (Nature4SDGs) (NERC Grant NE/S012850/1). The dataset integrates secondary data on rural livelihoods, multi-dimensional human wellbeing, household demographics, resource tenure and social-ecological context across 10,971 households in 232 settlements in ten low- and middle-income countries (Table 1 and Figure 1). It primarily draws upon nine existing household surveys, and their associated site descriptions and qualitative interviews. It also draws upon existing global geospatial datasets to provide further village-level information on the social-ecological context.

*Table 1. Overview of the original data included in this dataset.*

| **Project acronym** | **Countries** | **No. of settlements** | **No. of households** | **Landscape types** |
|---|---|---|---|---|
| ACES | Mozambique | 27 | 1614 | Woodland, agriculture |
| ASSETS | Colombia<br>Peru<br>Malawi | 11<br>9<br>6 | 195<br>250<br>675 | PER & COL: Forest, agriculture, riverine<br>MWI: Forest, agriculture |
| DELTAS | Bangladesh | 63 | 1586 | Coastal, marine, mangroves, agriculture |
| P4GES | Madagascar | 7 | 603 | Forest, agriculture |
| PEFESPA | India (Odisha) | 4 | 127 | Forest, agriculture |
| PIMA | Tanzania | 42 | 1922 | North: Grassland pasture, savannah (arid/semi-arid)<br>South: Miombo woodland (subhumid, significant woodland canopy throughout) |
| SENTINEL | India | 40 | 1112 | Forest, agriculture |
| SPACES | Kenya<br>Mozambique | 4<br>4 | 786<br>351 | Coastal, marine, mangroves, agriculture |
| Miyun | China | 15 | 1750 | Forest, agriculture |

*Figure 1. Map of the study countries and sites of the original projects.*




Below follows: a description of the data sources; an overview of how these feed into each component of the integrated dataset; and a description of the methods used to compile the data in each component of the datasets.

---

# 2. Data sources

## 2.1. Household surveys, site descriptions and interviews

The household surveys were conducted during nine different previous research projects. Each project had a different research focus, but all collected similar data on ecosystem services, income and human wellbeing. All surveys collected information on annual livelihoods for a period between 2011 to 2015. The surveys also generated a variety of qualitative information from site descriptions and interviews. The coverage of the nine original datasets is summarised in Table 1, and below follows the citations containing details of the original methodologies, and links to the project websites, where available:

- DELTAS (Adams et al., 2016), website: Assessing Health, Livelihoods, Ecosystem Services and Poverty Alleviation in Populous Deltas;
- P4GES (Poudyal et al., 2018), website: Can Paying 4 Global Ecosystem Services Reduce Poverty?;

- ACES (Smith et al., 2019), website: Abrupt Changes in Ecosystem Services and Wellbeing in Mozambican Woodlands;
- PIMA (Bluwstein et al., 2018) website: Poverty and Ecosystem Impacts of Tanzania's Wildlife Management Areas;
- SPACES (Chaigneau et al., 2019), website: Sustainable Poverty Alleviation from Coastal Ecosystem Services;
- ASSETS (Angarita-Baéz et al., 2017; Ramirez-Gomez et al., 2015) website: Attaining Sustainable Services from Ecosystems through Trade-off Scenarios;
- PEFESPA (Lakerveld et al., 2015), website: The Political Ecology of Forest Ecosystem Services and Poverty Alleviation;
- SENTINEL (Devagiri et al., 2015), website: Sentinel Landscapes Framework Assessment in Western Ghats;
- Miyun (Robinson et al., 2019): Disaggregating livelihood dependence on ecosystem services to inform land management.

### 2.2. Global datasets

To supplement the household surveys, we also used several global datasets to generate information on the social-ecological context of each settlement. We generated this information using data on the spatial location of the settlements (which is confidential and not included in this integrated dataset; see next section). See the section on Social-Ecological Context for full details of these datasets.

---

# 3. Overview of the integrated dataset

### 3.1 Overview

The integrated dataset is comprised of four components, compiled as four separate .csv files:

1. Household demographics and wellbeing (**n4s_hh.csv**), where each row is a household
2. Household livelihood sources (**n4s_lvl.csv**), where each row is a livelihood source associated with a particular household
3. Settlement-level resource tenure systems (**n4s_rts.csv**), where each row is a natural resource associated with a settlement.
4. Settlement-level social-ecological context (**n4s_setts.csv**), where each row is a settlement

The four components can be linked together through corresponding ID variables (in a relational database structure; Figure 2). Some variables are not present across all sites. The implications of this missingness are discussed below.

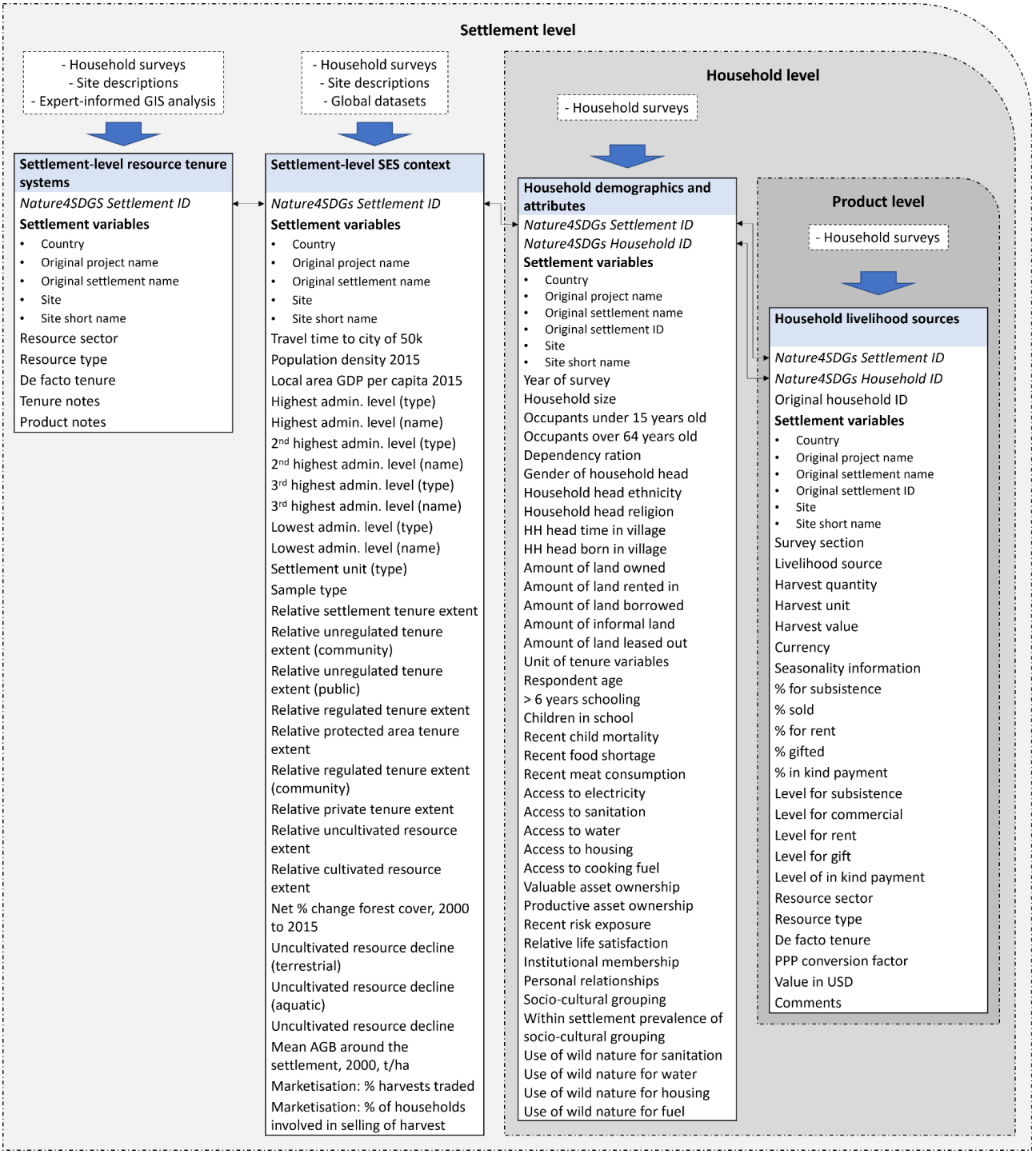
We also provide further three csv files to help navigate and use the dataset:

1. **n4s_ids.csv**: to help link between csv files, a file with all corresponding IDs, including the IDs from the original source datasets (should you want to link back to the original source databases)
2. **n4s_variable_names.csv**: a csv file with all variable names and descriptions

3. **n4s_ls.csv**: which contains more detailed information on the extent, proportion and area of the different land covers and associated resource tenure systems in each settlement. See section 4.4.

Spatial locations of the villages in any one area are subject to the confidentiality rules of the original research project and so are not published here. You can request access to these data by contacting the relevant people at the links above.

*Figure 2. Summary of dataset structure and sources.*



## 3.2 Data structure and IDs

Observations are organised in the following hierarchical structure: *project→country→site→ settlement→household*. Project refers to the original research project in which the data were collected, while site designates groupings of villages within a particular context. Some projects used these site-level distinctions to purposively select subgroups of villages from different contexts. *Table 2. Summary of coverage of household datasets*. Coverage of particular variables varies between datasets (see subsequent sections).

| **ID** | **ID Description** | **ID construction description** |
|---|---|---|
| pjnm | The project name of the original survey | Abbreviation of project name of original survey: "ac" = ACES, "asc" = ASSETS Colombia, "asm" = ASSETS Malawi, "asp" = ASSETS Peru, "de" = DELTAS , "mi" = Miyun, "od" = Orissa (PEFESPA), "pa" = P4GES, "pi" = PIMA, "se" = SENTINEL, "sp" = SPACES. |
| cn | The country | ISO3 country code |
| site | The social-ecological grouping of settlements used in the research design of the original survey | Qualitative name used in the original survey documentation |
| sett_id | The settlement | Comprised of: pjnm / ‘s’ + unique numeric settlement id / first six letters of village name |
| hhid | The respondent household | Comprised of: sett_id / ‘h’ + unique numeric household id |

## 3.3 Sampling

The observations at household and village level in this published dataset can be treated as follows:

- Households are random samples (or a census) of households within a village. See the ‘sample’ variable in *n4s_setts.csv* for the full details for each settlement
- Settlements were purposively selected based on the focus of the research question of the original study.

For all of the original datasets (linked above), household data were already randomly sampled or censused, except for ACES and SPACES. So that all households within this new published dataset could be treated as random within the village, we thus resampled the households for these two projects. ACES had a stratified random sample of households in each village, sampling equally from each strata in a four-level wealth rank (i.e. likely oversampled the highest and lowest parts of the distribution). Using supplementary data from the original study (wealth rankings of all households within each village), we thus resampled the ACES households to reflect the overall wealth rank distributions in each village. For SPACES, all households were randomly sampled within each village except for the village of Kongowea (*sp/s226/Kongow*), where households engaged in fishing were oversampled. We thus removed these oversampled households from the dataset. The original weighted datasets for ACES and SPACES are available in their archives linked above.

## 3.4 Missingness, imputation and data equivalence

While the original surveys shared a focus on ecosystem services and wellbeing, some of the variables differed slightly between the surveys. Additionally, each original survey has its own strengths and weaknesses in data quality (e.g. some have very robust livelihoods data, while others do not). In using this combined dataset, special consideration is therefore needed of

missingness, as well as data quality and equivalence, between original datasets. In Section 4 we highlight particular issues with missingness and data in each of the components of this integrated dataset. Immediately below we provide a more general overview of these issues.

Data missingness

For data missingness, there are two types:

- 'Real' missingness where, while a variable may have been collected in the original survey, there are occasionally a limited number of household values missing. This is due to more traditional issues around non-responses or enumerator errors in the original survey.
- 'Question absence' missingness, where the variable was not collected for an entire village or region in the original surveys.

Real missingness is coded as -9 in the dataset, while question absence is coded as -8.

Imputation

In the published dataset, we have avoided imputing missing data for all variables except for harvest value (*h_val*) in the livelihoods data frame (*n4s_lvl*). For all other variables, missingness can be imputed from the raw data provided if users so wish. For harvest value, robust imputation requires site-specific knowledge on similarities between villages and harvest unit equivalence. We thus undertook to impute missing *h_val* values for the combined dataset.

Missingness of the harvest value data primarily occurred because respondents could not provide values for every harvest in every household (e.g. because some households do not trade every type of harvest). We thus used a hierarchical strategy to impute best estimates of harvest value where it was missing. In order of preference:

1. Where other households in the village had reported a harvest value for the same livelihood source (*lvl_source*) and unit (*h_unit*), we took the median of the harvest in the village;
2. Where the above was not possible, we used field notes on *h_unit* equivalences to impute the within-village median value of the same *lvl_source* with different units;
3. Where the above was not possible, we extended imputation to include median values from nearby villages that according to site experts were qualitatively similar in their socio-ecological context;
4. Where the above was not possible, we used field notes and expert opinion from local site experts to estimate harvest values.

Data equivalence

While we have sought to generate common variables across the original surveys, the varying origin of these variables means that data quality and equivalence need to be carefully considered when designing an analysis. In Section 4 we highlight particular issues with data quality and equivalence for each part of the integrated dataset. Generally, in any one analysis particular sites can be assumed to have more robust variables for particular constructs. Analyses can deal with this by either focusing only on sites with high-quality variables, or by running multiple models (e.g. seeing if there are differences between one model with high-quality sites only, and a second with all of sites).

---

# 4. Methods and variable descriptions

Below follows a description of the methods used to generate each component of the datasets and more detailed descriptions of the variables.

## 4.1. n4s_hh.csv: Household demographics and wellbeing

We generated this dataset in two parts, first on household demographics and second on multidimensional human wellbeing.

### 4.1.1. Household demographics

#### Background

We used the household surveys to generate household-level variables on the age, labour profile, socio-cultural grouping and land ownership (Table 3, with example plots in Figure 3).

Variables on age (dependency) and labour profile will affect which livelihood strategies a household can engage in, and the (per capita) benefit from these livelihoods within the household. Social capital and social grouping variables are indicators of social difference that may indicate something about the way that they can interact with resource governance and other institutions. Variables on land ownership and tenure type provide household-level information on tenure, to complement the settlement-level 'Resource Tenure System' dataset.

#### Notes

- Not all variables are available across all sites because they were not collected in some of the original datasets. The degree of missingness in each dataset is summarised in Figure 4.
- Some other common variables on household demographics (e.g. health, education, assets, social capital) appear in a standardised binary form in the HWB part of the dataset (next section). If needed, unstandardised ordinal, interval and/or continuous variables are available in the original datasets.
- Household-level tenure variables are focused on private land resources. Variables on common property terrestrial, aquatic and marine resources are at the settlement level and can be found in the settlement-level components of this dataset ('Resource Tenure Systems' and 'Settlement-level SES context'). Private ownership in fisheries tends to be related to the ownership of a fishing vessel. This information is integrated into the 'productive assets' variable in the HWB dataset (next section).
- Here we have included ethnicity and religion information as nominal variables. These variables can either be used as is, or can be further interpreted to group households in some more meaningful way for a particular analysis.
- Other useful information on household occupations (e.g. the presence of 'elite' occupations; employment in different sectors) and wealth ranks can be derived from the 'Household Livelihoods' data frame.
- We have not included variables on government aid, credit/savings (source and cost) or debt because this information is not widely available throughout the surveys and the variables differ significantly where it is present. This information can be retrieved from the original surveys linked above if needed.

*Table 3. Summary of household demographic variables.*

| Short name | Long name | Type | Description |
|---|---|---|---|
| hh_size | Household size | Continuous | Household size including children |
| ppl_u15 | Occupants under 15 years old | Continuous | Household occupants under 15 years |
| ppl_o64 | Occupants over 64 years old | Continuous | Household occupants over 64 years |
| rsp_age | Respondent age | Ordinal | Age of respondent to survey |
| yr | Year | Ordinal | Year survey was conducted |
| dep_rat | Dependency ratio | Continuous | Number of people below 15 or above 64, divided by household size (UN DESA, 2019) |
| gen_hd | Gender of household head | Binary | Gender of household head |
| ethn_hd | Household head ethnicity | Nominal | Ethnicity or caste of household head |
| relig_hd | Household head religion | Nominal | Religion of household head |
| ethn_relig_cat | Socio-cultural grouping | Nominal | A marker of socio-cultural grouping as the intersection of religion and ethnicity/caste (ethnd_hd and relig_hd appended to each other) |
| soc_cult_share_pc | Within settlement prevalence of socio-cultural grouping | Continuous | The proportion of households within the villages with the same socio-cultural grouping |
| tm_vl_hd | HH head time in village | Continuous | Years household head has resided in the village |
| brn_vl_hd | HH head born in village | Binary | Household head born in village |
| tn_own | Amount of land owned | Continuous | Amount of land owned by household |
| tn_rnt | Amount of land rented in | Continuous | Amount of land rented (hectares) |
| tn_brw | Amount of land borrowed | Continuous | Amount of land borrowed (hectares) |
| tn_inf | Amount of informal land | Continuous | Amount of informally ‘private’ land (hectares) |
| tn_lsd | Amount of land leased out | Continuous | Amount of land leased out (hectares) |
| tn_unit | Unit of tenure variables | Nominal | The unit used to ask about tenure/land in the original survey |

*Figure 3. Examples of household demographic variables.*

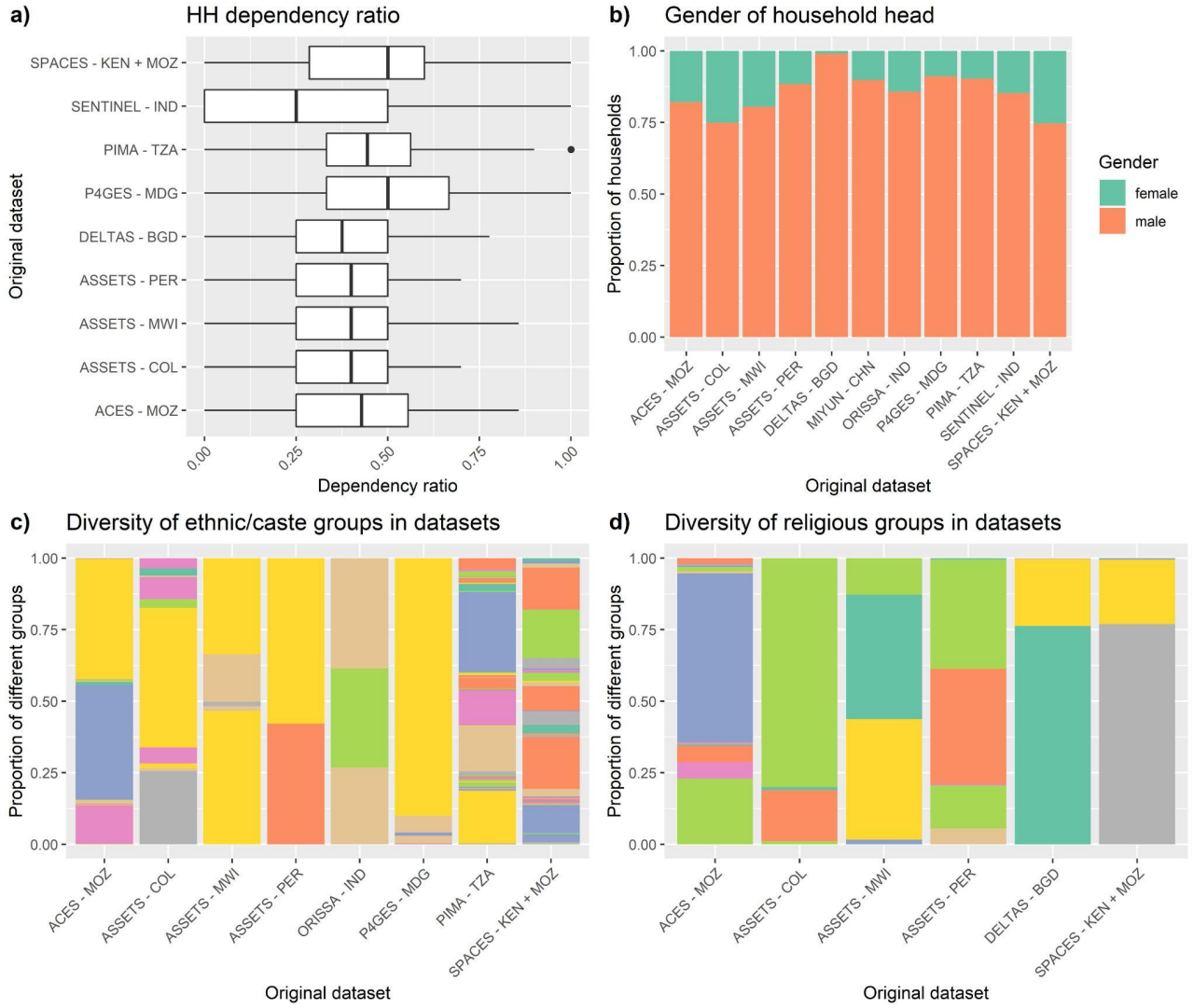
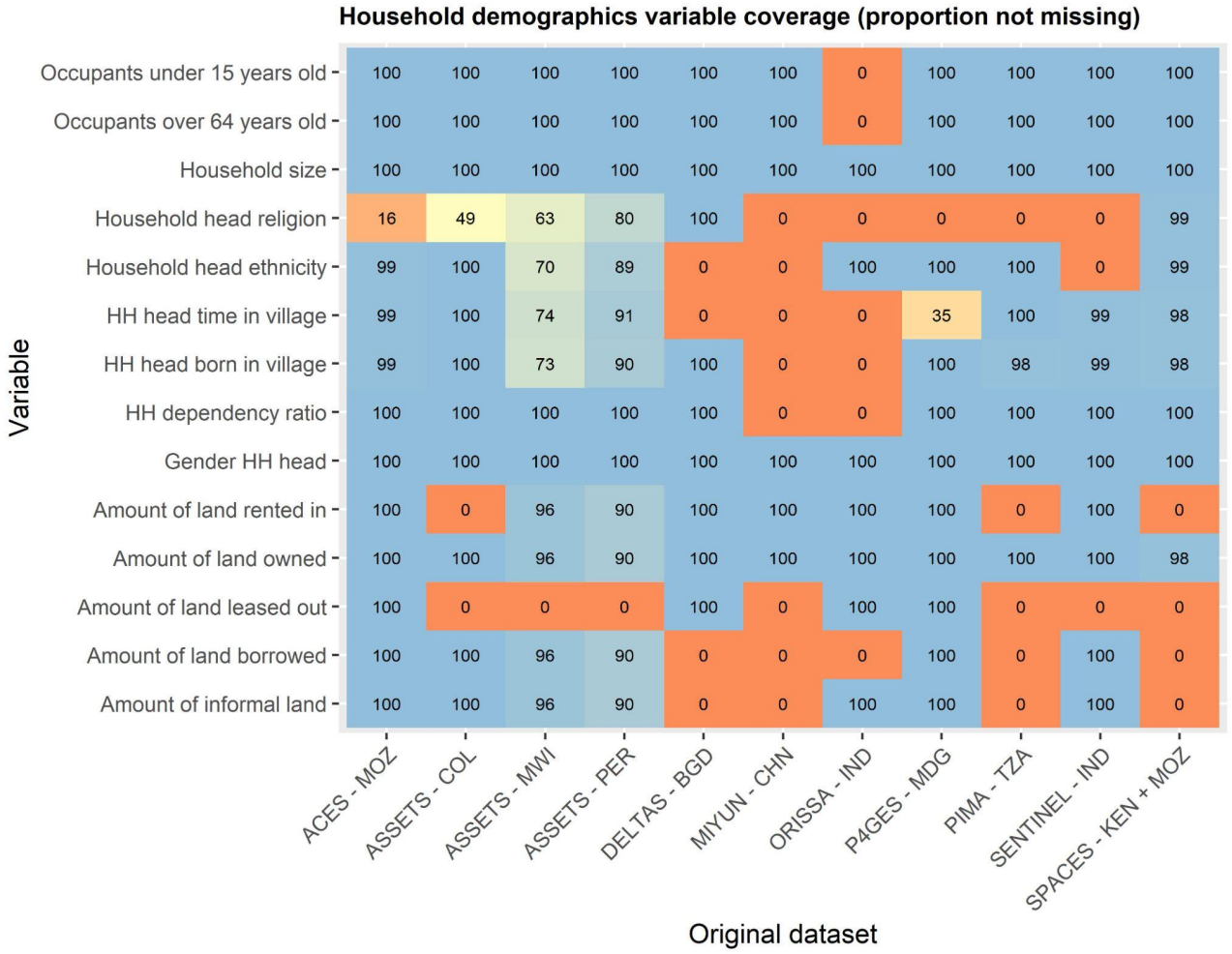


*Figure 4. Summary of missingness of household demographic variables*. 0% means the variable is not present in the original survey for that site.




### 4.1.2. Household multidimensional wellbeing

#### Background

We used the household surveys to generate a series of (mainly) binary household-level variables on different dimensions of human wellbeing (Table 4 and Figure 5). In doing so we sought to balance the need for detailed and locally contextualised measures with the need for cross-site comparability.

Broadly, we defined wellbeing as having three dimensions (material, subjective and relational) (White, 2010), and framed our measurement approaches based on the associated environment and development literature (Chaigneau et al., 2019; subjective WB: Dolan & Metcalfe, 2012; human needs: Doyal & Gough, 1991; wellbeing in developing countries: Gough & McGregor, 2007; capabilities: Sen, 1999; relational WB: White, 2015; Bodin et al., 2011). Definitions of these concepts vary (Agarwala et al., 2014). As a starting point, Coulthard et al. (2017) provide a useful summary of HWB as comprising:

> ...a material dimension that emphasizes the objective resources a person has access to; a relational dimension that considers how social relationships influence what people can (or cannot) do; and a subjective dimension that takes into account a person's level of satisfaction with the quality of life they achieve.

Within this framing ‘basic needs’ approaches aim to understand if people are deprived in different subdimensions of material, subjective and relational wellbeing (e.g. health, education, shelter, life satisfaction, social relations etc.) (Chaigneau et al., 2019).

To develop standardised measures for these different subdimensions, we adapted the methods used to generate the Oxford Poverty & Human Development Initiative’s (OPHI) multidimensional poverty indicator (MPI) (Alkire et al., 2018). The MPI approach is grounded in Amartya Sen’s capabilities approach (Sen, 1999) and is based on the ‘counting’ of different basic needs (or deprivations) that are met (or unmet) within a household (Alkire & Foster, 2011; Atkinson, 2003). It is already widely used to combine diverse data on wellbeing from different surveys, sites and countries (Alkire et al., 2017; Feeny & McDonald, 2016; Smith et al., 2019).

Typically the MPI is generated in two steps. First, for each indicator of a basic need (e.g. years of schooling) a cut-off is applied (e.g. < 6 years), below which you are deprived for that dimension (i.e. it transforms each observed indicator into a deprived/not deprived binary variable). This cutoff value can differ between datasets depending on locally contextualised cutoffs, and serves as a method of cross-dataset standardisation. In the second step, these binary indicators are added together to form a (weighted) multi-level ordinal index of relative aggregate deprivation across all households. A further cutoff is then applied to this variable to determine if a household is ‘poor’ (e.g. at least 1/3 of basic needs not met).

In generating this dataset we used the MPI approach as a starting point and made two adaptations. First, in order to maintain the richness of our wellbeing data, we only implemented the first step to provide a multivariate set of binary variables of deprivation in different basic needs. These could be combined into a weighted ordinal variable or a binary MPI if needed. Second, in addition to the material basic needs usually measured in the MPI we added further material dimensions (protein consumption, productive assets), as well as subjective (life satisfaction) and relational dimensions (institutions, autonomy) (Chaigneau et al., 2019; Howland et al., 2019).

Most variables from the original surveys were transformed into binary indicators according to common thresholds in the latest MPI (OPHI, 2018) and literature associated with the original datasets (Adams et al., 2020; Chaigneau et al., 2019; Keane et al., 2019; Smith et al., 2019). These thresholds are summarised in Table 4. The only exception is the life satisfaction variable, which was sufficiently similar between datasets that we could transform it into a four-level ordinal variable, and in doing so preserve more information on this dimension. See Appendix A for a full description of the thresholds used to generate each variable for each dataset, and for details on the transformation of the life satisfaction variable.

Notes

- Not all variables are available across all sites because they were not collected in some of the original datasets. The degree of missingness in household wellbeing variables in each dataset is summarised in Figure 6.
- We have not included a poverty line or wealth rank. These can be generated from the income information in the ‘Household Livelihoods’ dataset, and/or from the assets information in the original surveys.
- To aggregate variables into an MPI or similar, variables can be added into a weighted index, with indicators weighted so that each high-level ‘Dimension’ has equal weighting in the final indicator. To be directly comparable with the MPI *hlt_pt*, *lv_pas*, *rk_exp*, *rwb_cm_st*, *rwb_cm_ag* would need to be excluded. See OPHI 2018 for methods for

creating an MPI. An MPI may be limited by variables missing in whole datasets (see section on missingness).

*Table 4. Summary of household wellbeing variables.*

| Dimension | Short name | Full name | Type | Description of main threshold (see Annex A for alternatives for some original datasets) |
|---|---|---|---|---|
| Material - education | edu_yrs | > 6 years schooling | binary | An adult has completed six years of schooling. |
| Material - education | edu_chld_att | Children in school | binary | Any school-aged child is not attending school for at least six years |
| Material - health | hlt_chld_mt | Recent child mortality | binary | Any child below 18 has died in the family. |
| Material - health | hlt_fd_sec | Recent food shortage | binary | Reported not having enough to eat in the last 12 months |
| Material - health | hlt_pt | Recent meat consumption | binary | Did not eat meat in last week, or every week on average over period |
| Material - living std. | ls_elec | Access to electricity | binary | The household has no electricity |
| Material - living std. | ls_san | Access to sanitation | binary | No flush toilet or latrine, or ventilated improved pit or composting toilet |
| Material - living std. | ls_wat | Access to water | binary | No piped water, public tap, borehole or pump, protected well, protected spring or rainwater |
| Material - living std. | ls_hse | Access to housing | binary | At least two of the three housing materials for roof, walls and floor are natural materials |
| Material - living std. | ls_fuel | Access to cooking fuel | binary | The household cooks with dung or wood |
| Material - living std. | ls_ast | Valuable asset ownership | binary | The household does not own more than one of the following items: radio, TV, telephone, bike, motorbike computer, or refrigerator. |
| Material - living std. | ls_pas | Productive asset ownership | binary | The household does not own at least one of the following: agricultural land; fishing vessel; livestock |
| Material - risk exposure | rk_exp | Recent risk exposure | binary | Experienced a shock in the last 12 months |
| Subjective | swb | Relative life satisfaction | ordinal | Self reported life satisfaction |
| Relational | inst | Institutional membership | binary | Household is not a member of, or has not received assistance from, one of the organisations asked about in the survey |
| Relational | prel | Personal relationships | binary | Dissatisfied (or below) with friendships or community support |
| Use of wild nature for sanitation | pm_san | Use of wild nature for sanitation | binary | Presence in household of use of wild nature for sanitation |
| Use of wild nature for water | pm_wat | Use of wild nature for water | binary | Presence in household of use of wild nature for water |

| | | | | |
|---|---|---|---|---|
| Use of wild nature for housing | pm_hse | Use of wild nature for housing | binary | Presence in household of use of wild nature for housing |
| Use of wild nature for fuel | pm_fuel | Use of wild nature for fuel | binary | Presence in household of use of wild nature for fuel |

*Figure 5. Bar plots of household wellbeing variables.* 0% means the variable is not present in the original survey for that site.

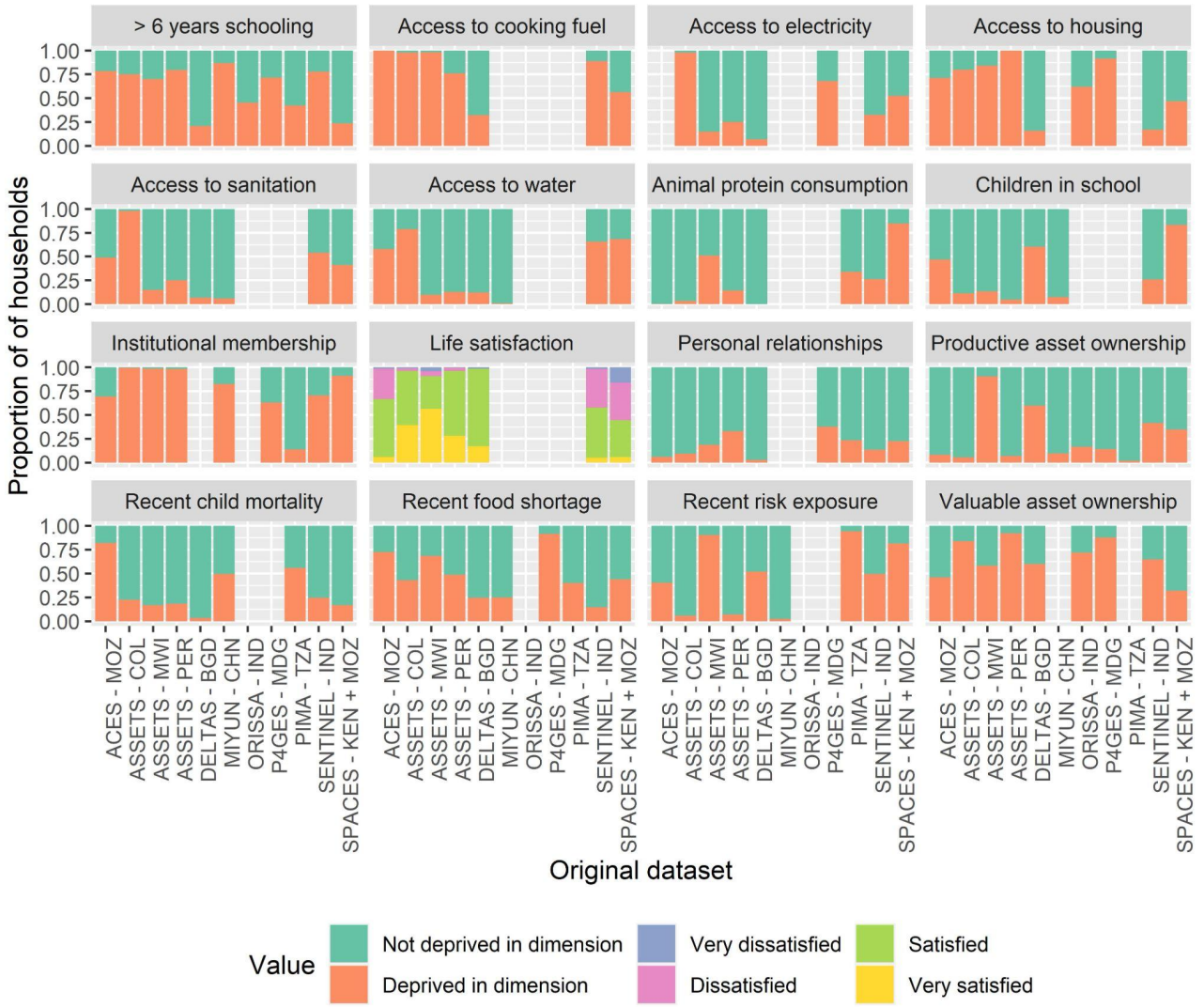
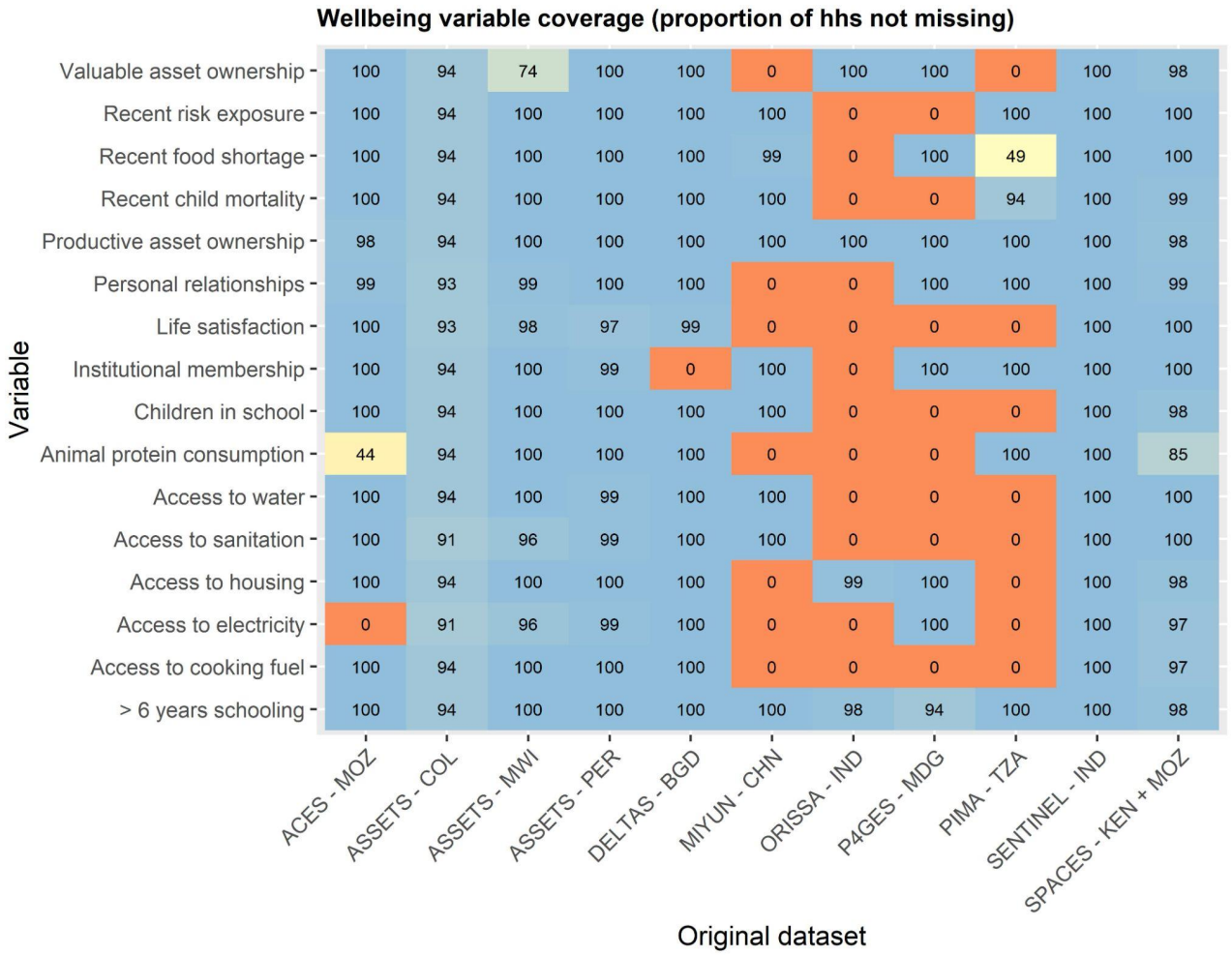


*Figure 6. Summary of missingness in household wellbeing variables.*




## 4.2. n4s_lvl.csv: Household livelihoods

### Background

Broadly the concept of livelihood includes economic as well as non-economic attributes of survival. Apart from income, it covers the “social relationships and institutions that mediate people’s access to different assets and income streams” (Ellis, 2000). This part of the dataset focuses on the economic aspect of livelihoods i.e. income of households.

We prepared a dataset containing two types of livelihoods: harvests from cultivated (e.g. farms, aquaculture) and uncultivated (e.g. forests, fisheries) sources; and cash income from employment, businesses etc. For harvests, we generated information on the annual quantity and value of harvest of each product collected by each household and shares of harvests used for subsistence and sale (Table 5, with examples in Figure 7). Cash income includes each household's earnings from non-farm businesses and other sources, wage income and remittances. To get the value of the harvest, we have used the market price of each product, including for harvests that were not traded (e.g. subsistence harvests). To tackle the problem of missing price, we assigned the median price of a product calculated at the village level (see section on imputation above). Aggregation of harvest values and other earnings give the total household gross income. We then converted the gross household income into USD using World Bank indices of Purchasing Power Parity (PPP) in the year of the survey (World Bank, 2020).

Full livelihoods data were available for all original datasets except for PIMA and PEFESPA. The PIMA survey gives the share of the harvest, instead of absolute quantities. Therefore we are not able to estimate the gross value of harvest, only the proportion. The livelihood data of PEFESPA is incomplete in the sense that it does not have data on income from the non-farm sector, wage income and remittances.

Notes

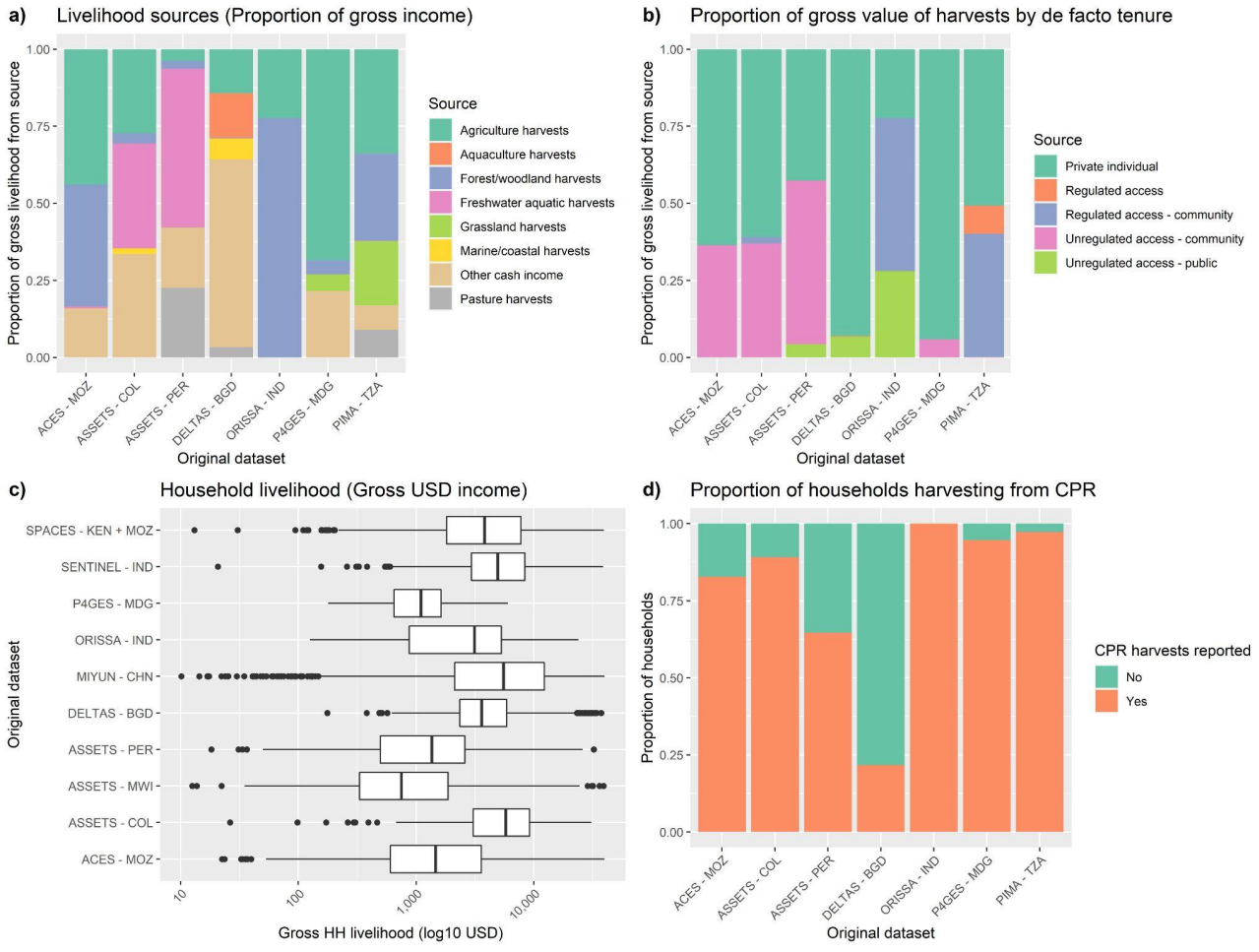
- From the product level data set, we get the gross value of harvests i.e. the monetary value of harvest without deducting the cost of labour and other input costs. The reason for considering the gross value of harvest is that most of the Nature4SDGs datasets do not have robust data on cost.

*Table 5. Summary of livelihood variables.*

| **Short name** | **Long name** | **Variable type** | **Description** |
|---|---|---|---|
| sv_sec | Survey section | Nominal | Value-labels of this variable represent different sections of the questionnaire pertaining to sources of livelihoods. |
| lvl_source | Livelihood source | Nominal | Value-labels pertain to various products harvested by households, live-stocks and non-nature sources of income. |
| h_qty | Harvest quantity | Continuous | It represents quantities of harvest of each product. |
| h_unit | Harvest unit | Nominal | Value labels represent units of products harvested. |
| h_val | Harvest value | Continuous | It pertains to the monetary value of harvests. |
| cur | Currency | Nominal | Value labels represent the currency of the country to which surveyed settlements belong. |
| usd | Value in USD | Continuous | Value in USD. Converted using World Bank PPP conversions from survey year. |
| ssn_info | Seasonality information | Nominal | Value labels pertain to the months in which a particular product is harvested. |
| subs_pc | % for subsistence | Continuous | This variable represents percentage shares of harvest used for subsistence. In projects where data on the quantity of harvest used as rent, gift are available, subs_pc means share used for household consumption. Otherwise, it gives the share of the harvest which is not sold. |
| comm_pc | % sold | Continuous | It pertains to percentage shares of sale in total value of harvest |
| rent_pc | % for rent | Continuous | It pertains to percentage shares of harvests paid as rent |
| gift_pc | % gifted | Continuous | It represents percentage shares of harvests used as gift |
| kindpay_pc | % for in kind payment | Continuous | It represents percentage shares of harvests used to pay for other goods and services (in kind). |
| subs_cat | Level for subsistence | Nominal | Value labels represent whether the household use any harvests for subsistence; 1-yes, 0-No |
| comm_cat | Level for commercial | Nominal | Value labels represent whether the household sale any harvests; 1-yes, 0-No |
| rent_cat | Level for rent | Nominal | Value labels represent whether the household use any harvests for paying rent; 1-yes, 0-No |
| gift_cat | Level for gift | Nominal | Value labels represent whether the household use any harvests in gifting; 1-yes, 0-No |
| kind_cat | Level of in kind | Nominal | Value labels represent whether the household use any |

| | | | |
|---|---|---|---|
| | payment | | harvests to pay for other goods and services ; 1-yes, 0-No |
| comments | Comments | Nominal | Comments about the household. |
| sector | Resource sector | | Higher-level economic or natural resource sector present within a settlement. Resource sectors identified include: agriculture, fisheries, forest (natural), grassland, silviculture, water bodies, built environment. |
| res_type | Resource type | | Specific type of resources or resource spaces/locations present in each settlement. Examples of resource type include: farmland, forest, rivers, lakes, home-gardens, agroforestry plots, mangroves, coastal and marine fisheries etc. |
| de_facto | De facto tenure | | Tenure in practice in the settlement for a given resource type. For non-cultivated resources, focus is on access (for appropriation). Categories include: private-ind(ividual), private-firm, share-cropping, communal (farming), regulated access-community, unregulated access-community, unregulated access-public (open access), regulated access (general), protected area (strict protection). |
| ppp | PPP conversion factor | Nominal | Purchasing power parity conversion used to convert all currency to USD in 2015. |

*Figure 7. Examples of livelihood variables*.



## 4.3. n4s_rts.csv: Settlement-level resource tenure systems

Our approach to identifying and mapping resource tenure prevalent at the site level follows existing frameworks on social-ecological systems (e.g. McGinnis & Ostrom, 2014; Ostrom, 2009), tenure and property rights (e.g. Robinson et al., 2019; RRI, 2012; Schlager & Ostrom, 1992), together with the site/project-specific information on the types of resource available to the households and *de facto* and *de jure* rights to those resources. We contextualised the settlement-level resource systems around 'service shed(s)' idea (Olander et al., 2018), and mapped key categories of rights to the identified resource types, focusing specifically on accurately mapping rights in practice (*de facto*) at all settlements based on the available qualitative data (participatory maps, land-use change, resource trends) and household survey information (land holdings, access to resources).

Resource sector and resource types

Sector is a higher-level classification of resources, similar to 'RS1 Sector' in Ostrom (2009). Within each sector, we can have multiple resource types which relate to primary use/purpose or some defining characteristics of the resource, including limitations. Each resource type could be characterised as either 'cultivated' or 'uncultivated'.

Resource tenure

The term tenure encompasses not just property rights but also wider institutions (such as who makes constitutional rules, who makes operational rules) within which resource use is embedded. While defining higher-level tenure types of resources, we limit ourselves to property rights.

All common-pool resources comprise a class of goods that shares two attributes of importance for economic activities: (1) it is costly to exclude individuals from using the good either through physical barriers or legal instruments; and (2) benefits consumed by one individual subtract from the benefits available to others (Ostrom et al., 1994). Schlager and Ostrom (1992) identify five property rights that are most relevant for use of common-pool resources, including "access, withdrawal, management, exclusion, and alienation". For our resource tenure mapping exercise, particularly for uncultivated common pool resources, we are only looking at *de facto* access and withdrawal rights from those resource types. At the higher level, we have defined four resource tenure categories based on *de facto* rights to the resource type defined earlier. Higher-level tenure classification broadly corresponds to:

1. Privately owned resources (primarily cultivated);
2. Community managed resources (regulated/unregulated at the community level);
3. Protected area (strictly protected, regulated access or *de facto* open access); and
4. Open access.

Based on the resource tenure systems mapping exercise outlined above, we have created five key variables defined at settlement level, which are summarised in Table 6, and Figures 8 and 9.

*Table 6. Summary of resource tenure system variables.*

| Short name | Long name | Description |
|---|---|---|
| sector | Resource sector | Higher-level economic or natural resource sector present within a settlement. Resource sectors identified include: agriculture, fisheries, forest (natural), grassland, silviculture, water bodies, built environment. |
| res_type | Resource type | Specific type of resources or resource spaces/locations present in |

| | | |
|---|---|---|
| | | each settlement. Examples of resource type include: farmland, forest, rivers, lakes, home-gardens, agroforestry plots, mangroves, coastal and marine fisheries etc. |
| de_facto | *De facto* tenure | Tenure in practice in the settlement for a given resource type. For non-cultivated resources, focus is on access (for appropriation). Categories include: private-ind(ividual), private-firm, share-cropping, communal (farming), regulated access-community, unregulated access-community, unregulated access-public (open access), regulated access (general), protected area (strict protection) |
| tenure_notes | Tenure notes | Additional information/notes related to tenure systems (de facto/de jure) for the resource type. Information ranges from specific detail on the type of resource (e.g. 'community forest' for resource type 'forest') to specific tenure arrangements for the resource (e.g. permit requirements and fees). |
| prod_notes | Product notes | Additional information/notes related to specific products from (primarily) non-cultivated systems, such as whether a certain product can be harvested, has restricted access and so on. For example: 'NTFP harvesting allowed, timber not permitted'. |

*Figure 8. Summary of resource tenure system sectors present across datasets.*

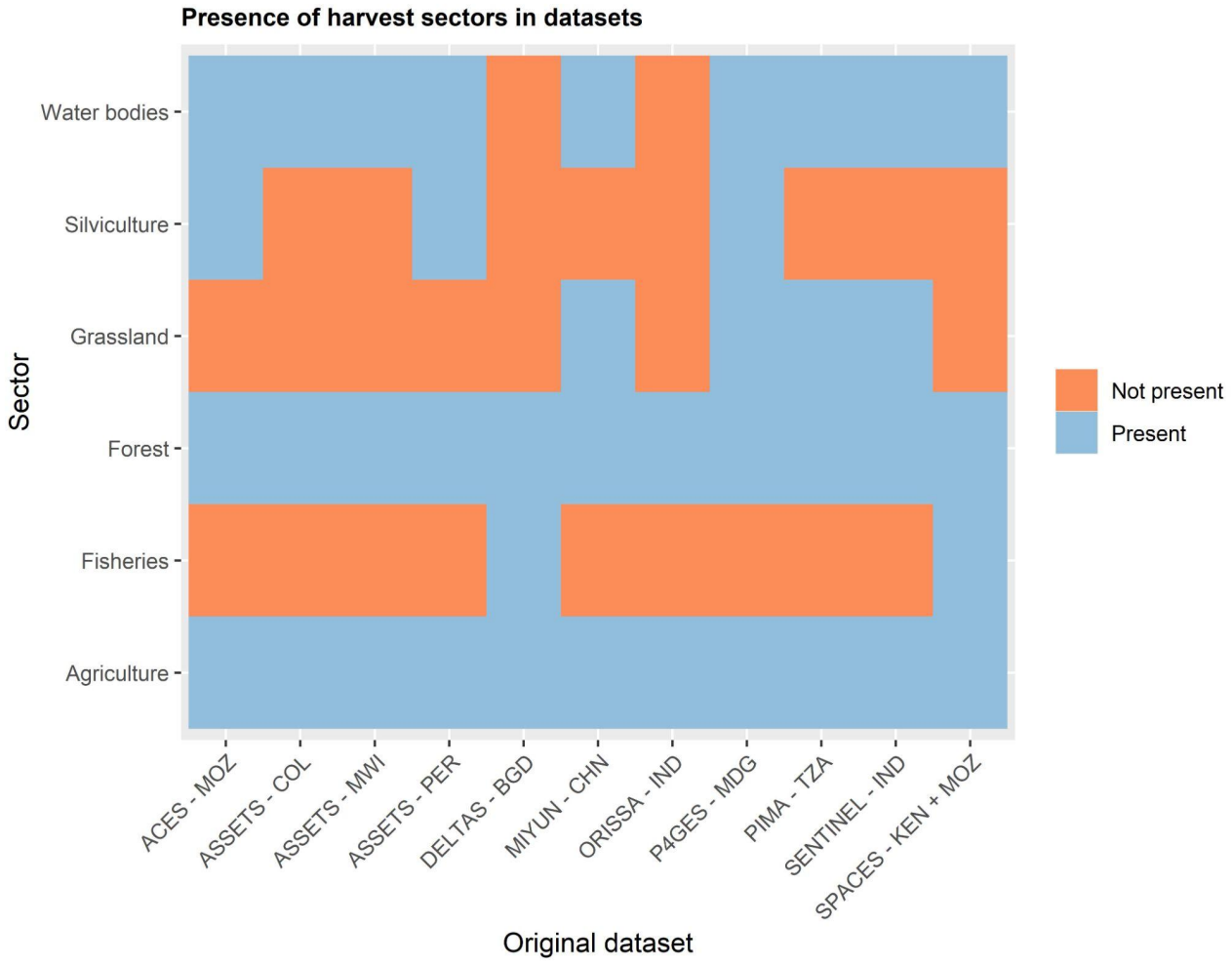
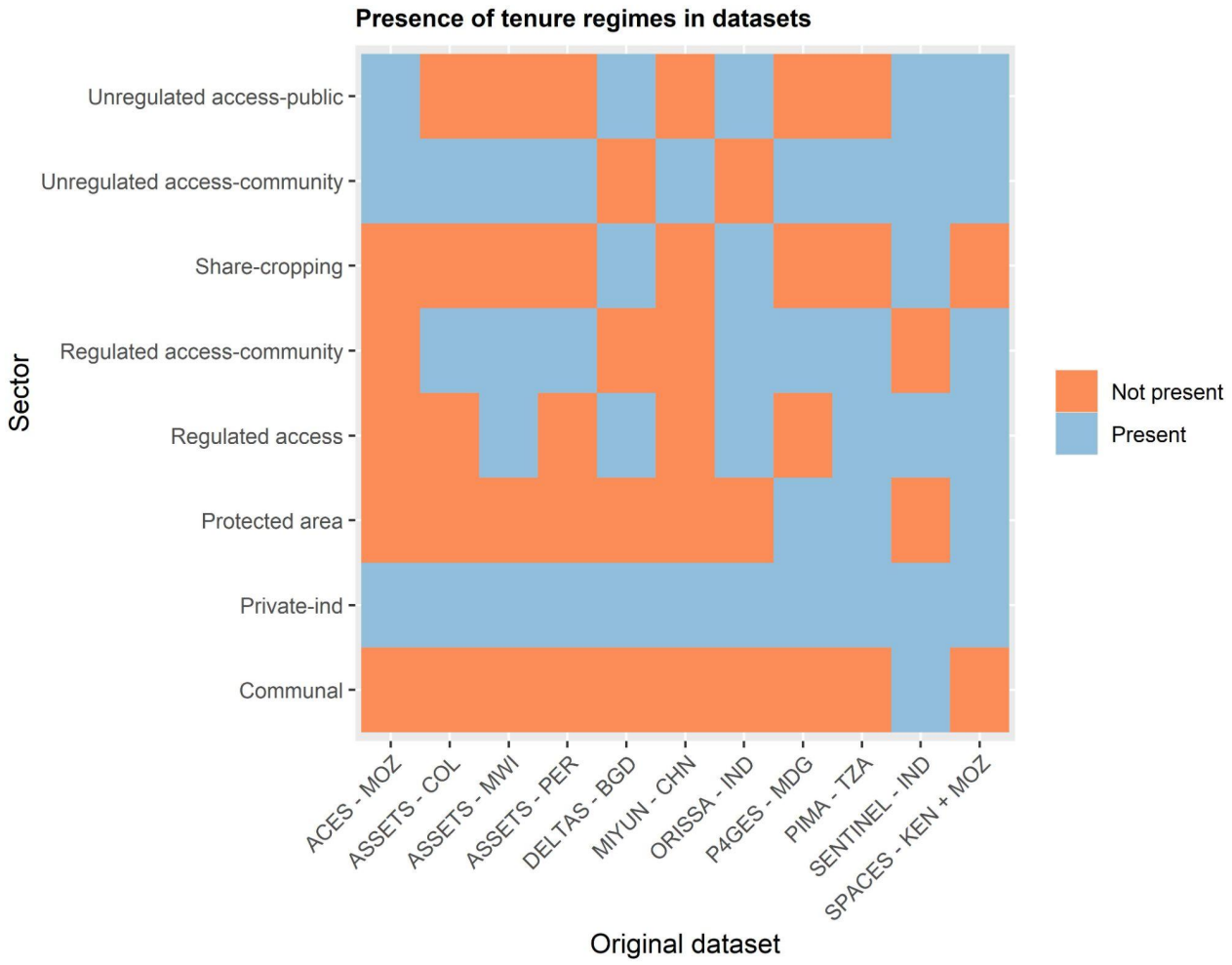


*Figure 9. Summary of de facto tenure present across datasets.*




## 4.4. n4s_setts.csv and n4s_ls.csv: Settlement-level social-ecological context

### Background

These variables aim to describe the broader social and ecological context at the settlement level (Table 8, with examples in Figure 11). The social variables are focused on the degree of market access in the village and population density. These variables were generated either directly from the surveys, or from existing spatiotemporal global datasets on travel time to cities (Weiss et al., 2018), population density (CIESIN, 2018) and estimated local area GDP in USD (Kummu et al., 2018).

The ecological variables are focused on characterising the land use land cover around each settlement, the associated *de facto* tenure regimes, as well as the relative extent of 'wild' or 'uncultivated' areas. All ecological variables are generated from existing spatiotemporal datasets.

There are two csvs associated with this part of the dataset:

- The main data frame **n4s_setts.csv**, which contains summary information about the context around each settlement
- A supplementary data frame **n4s_ls.csv**, which contains detailed information about the extent, proportion and area of different land covers and resource tenure systems around each village (see below).

## Methods

We generated the variables for each settlement according to the following steps:

1. Estimating the spatial extent, or 'landscape buffer', within which to generate ecological variables;
2. Using land use land cover products and expert knowledge to generate estimates of the extent of different resource tenure systems within this buffer (see previous section for details on resource tenure system classifications);
3. Using variables from prior steps to generate a set of simplified indices on the proportion of different tenure regimes and uncultivated resources around a settlement.

Each of these steps is explained in more detail below.

### 1. Landscape buffers

In the absence of detailed spatial information on the location of the resource catchment for each settlement, we instead focused on characterising the more general construct of the 'landscape' for all sites in a simple circular buffer around each of the settlement centroids (Simensen et al., 2018) (or technically a 'resourcescape' where this includes aquatic resources). We let the diameter of this buffer vary with each settlement (Table 7), informed by information from the surveys on the approximate size of the underlying resource catchment, and expert knowledge on the local social-ecological context. For example, the resource buffer is larger in Tanzania where pastoralist resources extend a longer distance from the village. While these buffers are not only rough representations of resource catchment boundaries, our assumption is that the landscape metrics generated within these boundaries are correlated with characteristics of the underlying resource catchments.

*Table 7. Buffer sizes used to estimate the extent of land cover around a village.*

| **Project acronym** | **Countries** | **Landscape types** | **Buffer diameter (metres)** |
|---|---|---|---|
| ACES | Mozambique | Woodland, agriculture | 5000 |
| ASSETS | Colombia<br>Peru<br>Malawi | PER & COL: Forest, agriculture, riverine<br>MWI: Forest, agriculture | 3000 |
| DELTAS | Bangladesh | Coastal, marine, mangroves, agriculture | 5000 |
| P4GES | Madagascar | Forest, agriculture | 3000 |
| PEFESPA | India (Odisha) | Forest, agriculture | 3000 |
| PIMA | Tanzania | North: Grassland pasture, savannah (arid/semi-arid)<br>South: Miombo woodland (subhumid, significant woodland canopy throughout) | 10000 |
| SENTINEL | India | Forest, agriculture | 3000 |
| SPACES | Kenya<br>Mozambique | Coastal, marine, mangroves, agriculture | 5000 |
| Miyun | China | Forest, agriculture | 5000 |

2. Extent of resource tenure systems

Figure 10 outlines the workflow for generating spatial estimates of RTS extents for each settlement.

First, we identified the best land use land cover maps available for each site. Only three of the original datasets included land cover maps (ASSETS, DELTAS and Miyun). We thus selected different land cover products for different sites, preferencing land cover data from the original datasets, followed by high-quality locally contextualised national maps where available (for sites in India, NSRC, 2016), followed by the highest resolution global product we could find where no other data was available (the 10m resolution product from Chen et al., 2019 for all sites, except for PIMA for which we used the 30m resolution product from ESA, 2016).

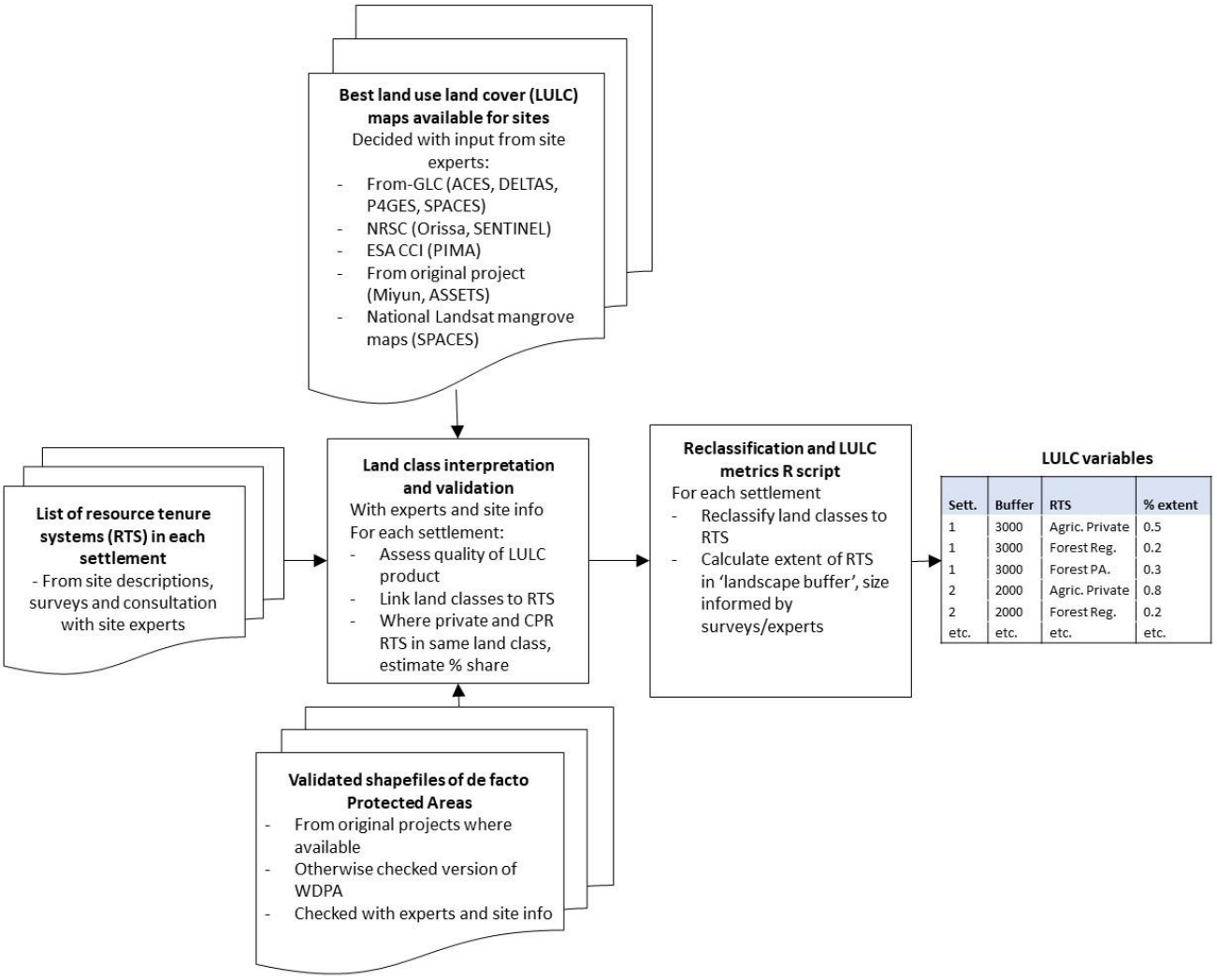
Next, to link the land classes in these different land covers to the resource and tenure types indicated in our analysis of resource tenure systems (see section on resource tenure systems above), we conducted a validation exercise with experts who are familiar with each site to assess what each land class likely means in each settlement (e.g. based on local knowledge and satellite imagery is 'grassland' actually likely to be grassland, or is it more likely agriculture or pasture, or a mix of both?). We also used this to check for and correct systematic and random errors in classification around each village. Based on this exercise we then updated the definition of each land class to better reflect the resources and associated tenure systems represented in each layer.

Following this expert assessment, we also added supplementary information on the location of mangroves (which were usually missed by land cover land use products), and on the boundaries of *de facto* protected areas (i.e. areas of the landscape where all resource use is excluded). For mangroves, we used spatial polygons of mangroves mapped during the original projects. For protected areas, we used spatial data from the original projects where available, and in other cases a version of the World Database on Protected Areas (WDPA) (Bingham et al., 2019), validated for each village to check if the boundaries were correct and if the protected area actually excluded resource use.

Finally, within the relevant landscape buffer for each village, we derived estimates of the extent of different resource tenure systems in the landscape around each settlement. Where there was more than one resource tenure system in a land class, and where these land classes reflected an important distinction for our analysis (e.g. where they include both cultivated and uncultivated systems), we used site-level knowledge to estimate the proportion by which to split the land class.

The resulting variables provide a flexible core dataset for generating different locally-contextualised variables on tenure and land cover in each settlement as needed. These variables are in **n4s_ls.csv**.

*Figure 10. Workflow for generating spatial estimates of RTS extents for each settlement.*




5. Simplified indices on the extent of tenure regimes, uncultivated resources and trends in quality

Using the above variables we generated two simplified settlement-level variables:

1. Relative extent of resource tenure systems: we used the validated land cover data on resource tenure systems to estimate the proportion of land cover under different types of tenure within the settlement buffer
2. Relative extent of uncultivated land cover: we used the validated land cover data on resource tenure systems to estimate the proportion of land cover representing uncultivated or wild nature within the settlement buffer

These variables are included in **n4s_setts.csv**, along with other settlement-level variables.

### Notes and limitations

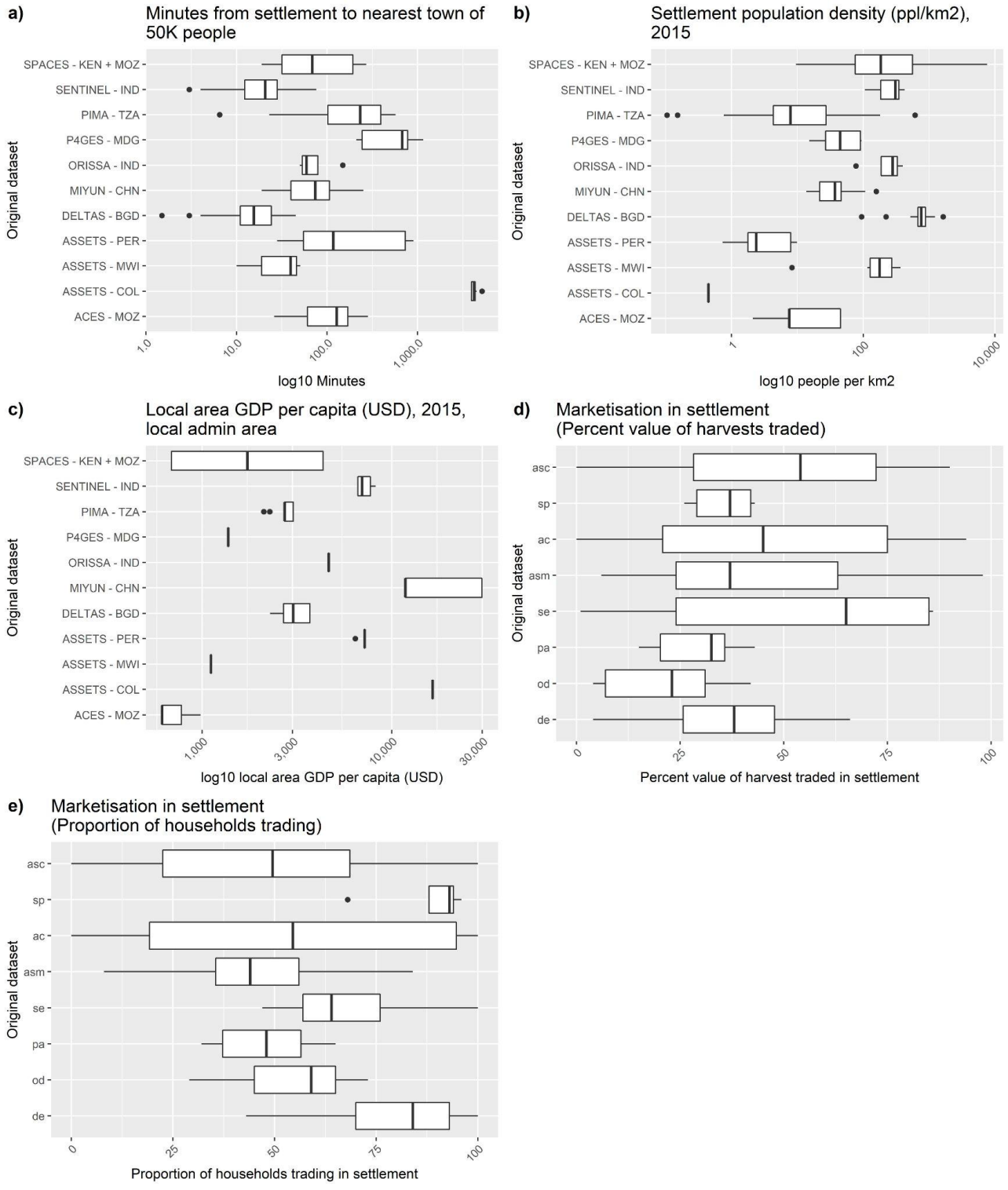
- Given the uncertainties associated with the landscape buffers and landscape data, these should only be interpreted as broad proxies for ecological context.
- Some land classes contain more than one 'resource tenure system' (e.g. 'forest' sometimes contains both private plantations and community forest). This is reflected in the variable name (see section on resource tenure systems).

*Table 8. Summary of social-ecological context variables.*

| Short name | Long name | Variable type | Description |
|---|---|---|---|
| ma_trtm | Travel time to city of 50k | Continuous | Travel time in minutes to the nearest city of 50,000 people accounting for the landscape and transport infrastructure. From Weiss et al. 2018, ~1km resolution. Average in 5km buffer around settlement centroid. |
| mkt_val | Marketisation: % harvests traded | Ratio | The share of the value of harvests sold in the settlements (from surveys) |
| mkt_hh | Marketisation:: % of households involved in selling of harvest | Ratio | The share of households selling harvests on the settlement (from surveys) |
| dg_pop_d | Population density 2015 | Continuous | People per sq. km. Derived from CIESIN 2018, ~1km resolution. Average in 5km buffer around settlement centroid. |
| ma_gdp_pc15 | Local area GDP per capita 2015 | Continuous | Local area GDP per capita, in 2015. Derived from Kummu et al. 2018, ~250m resolution. Average in 5km buffer around settlement centroid. |
| tn_ext_private | Relative Private tenure extent | ratio | The proportion of land cover under Private tenure, within the landscape buffer |
| tn_ext_unregulated_community | Relative Unregulated tenure extent | ratio | The proportion of land cover under Community-level Unregulated access tenure, within the landscape buffer |
| tn_ext_settlement | Relative Settlement tenure extent | ratio | The proportion of land cover under Settlement tenure, within the landscape buffer |
| tn_ext_regulated | Relative Regulated tenure extent | ratio | The proportion of land cover under Regulated tenure, within the landscape buffer |
| tn_ext_unknown_na | Relative unknown tenure extent | ratio | The proportion of land cover under unknown tenure, within the landscape buffer |
| tn_ext_unregulated_public | Relative Unregulated tenure extent | ratio | The proportion of land cover under Unregulated public access tenure, within the landscape buffer |
| tn_ext_protected_area | Relative Protected Area tenure extent | ratio | The proportion of land cover under Protected Area tenure, within the landscape buffer |
| tn_ext_regulated_community | Relative Regulated tenure extent | ratio | The proportion of land cover under Community-level Regulated access tenure, within the landscape buffer |
| rs_ext_uncultivated | Relative Uncultivated resource extent | ratio | The proportion of uncultivated land cover within the landscape buffer |
| rs_ext_cultivated | Relative Cultivated resource extent | ratio | The proportion of cultivated land cover within the landscape buffer |
| admn_lw_nm | Lowest admin. level (name) | nominal | The name of the lowest (before settlement) administrative level in the country (not available for all settlements) |
| admn_lw_unit | Lowest admin. | nominal | The type of the lowest (before settlement) |

| | | | |
|---|---|---|---|
| | level (type) | | administrative level in the country (not available for all settlements) |
| admn1_nm | Highest admin. level (name) | nominal | The name of the highest administrative level in the country (not available for all settlements) |
| admn1_unit | Highest admin. level (type) | nominal | The type of the highest administrative level in the country (not available for all settlements) |
| admn2_nm | 2nd highest admin. level (name) | nominal | The name of the second highest administrative level in the country (not available for all settlements) |
| admn2_unit | 2nd highest admin. level (type) | nominal | The type of the second highest administrative level in the country (not available for all settlements) |
| admn3_nm | 3rd highest admin. level (name) | nominal | The name of the third highest administrative level in the country (not available for all settlements) |
| admn3_unit | 3rd highest admin. level (type) | nominal | The type of the third highest administrative level in the country (not available for all settlements) |
| sample | Sample type | nominal | Type of sampling of households within settlement |

*Figure 11. Examples of settlement-level socio-economic context variables.*



# References


Adams, H., Adger, W. N., Ahmad, S., Ahmed, A., Begum, D., Lázár, A. N., Matthews, Z., Rahman, M. M., & Streatfield, P. K. (2016). Data Descriptor: Spatial and temporal dynamics of multidimensional well-being, livelihoods and ecosystem services in coastal Bangladesh. *Scientific Data*, *3*(November). https://doi.org/10.1038/sdata.2016.94

Adams, H., Adger, W. N., Ahmad, S., Ahmed, A., Begum, D., Matthews, Z., Rahman, M. M., Nilsen, K., Gurney, G. G., & Streatfield, P. K. (2020). Multi-dimensional well-being associated with economic dependence on ecosystem services in deltaic social-ecological systems of Bangladesh. *Regional Environmental Change*, *20*(2), 42. https://doi.org/10.1007/s10113-020-01620-x

Agarwala, M., Atkinson, G., Fry, B. P., Homewood, K., Mourato, S., Rowcliffe, J. M., Wallace, G., & Milner-Gulland, E. J. (2014). Assessing the relationship between human well-being and ecosystem services: A review of frameworks. *Conservation and Society*, *12*(4), 437–449.

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, *95*(7–8), 476–487. https://doi.org/10.1016/j.jpubeco.2010.11.006

Alkire, S., Jindra, C., Robles Aguilar, G., & Vaz, A. (2017). Multidimensional poverty reduction among countries in Sub-Saharan Africa. *Forum for Social Economics*, *46*, 178–191.

Alkire, S., Kanagaratnam, U., & Suppa, N. (2018). The global multidimensional poverty index (mpi): 2018 revision. *OPHI MPI Methodological Notes*, *46*.

Angarita-Baéz, J. A., Pérez-Miñana, E., Beltrán Vargas, J. E., Ruiz Agudelo, C. A., Paez Ortiz, A., Palacios, E., & Willcock, S. (2017). Assessing and mapping cultural ecosystem services at community level in the Colombian Amazon. *International Journal of Biodiversity Science, Ecosystem Services and Management*, *13*(1), 280–296. https://doi.org/10.1080/21513732.2017.1345981

Atkinson, A. B. (2003). Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches. *Journal of Economic Inequality*, *1*, 51–65.

Bingham, H. C., Bignoli, D. J., Lewis, E., MacSharry, B., Burgess, N. D., Visconti, P., Deguignet, M., Misrachi, M., Walpole, M., & Stewart, J. L. (2019). Sixty years of tracking conservation

progress using the world database on protected areas. *Nature Ecology & Evolution*, *3*(5), 737–743.

Bluwstein, J., Homewood, K., Lund, J. F., Nielsen, M. R., Burgess, N., Msuha, M., Olila, J., Sankeni, S. S., Millia, S. K., & Laizer, H. (2018). A quasi-experimental study of impacts of Tanzania's wildlife management areas on rural livelihoods and wealth. *Scientific Data*, *5*, 180087.

Bodin, Ö., Ramirez-Sanchez, S., Ernstson, H., & Prell, C. (2011). *A social relational approach to natural resource governance*.

Chaigneau, T., Coulthard, S., Brown, K., Daw, T. M., & Schulte‐Herbrüggen, B. (2019a). Incorporating basic needs to reconcile poverty and ecosystem services. *Conservation Biology*, *33*(3), 655–664. https://doi.org/10.1111/cobi.13209

Chaigneau, T., Coulthard, S., Brown, K., Daw, T. M., & Schulte‐Herbrüggen, B. (2019b). Incorporating basic needs to reconcile poverty and ecosystem services. *Conservation Biology*, *33*(3), 655–664. https://doi.org/10.1111/cobi.13209

Chen, B., Xu, B., Zhu, Z., Yuan, C., Suen, H. P., Guo, J., Xu, N., Li, W., Zhao, Y., & Yang, J. (2019). Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017. *Sci Bull*, *64*, 370–373.

CIESIN. (2018). *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11*. Center for International Earth Science Information Network. NASA Socioeconomic Data and Applications Center (SEDAC). https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11. https://doi.org/10.7927/H49C6VHW

Coulthard, S., Evans, L., Turner, R., Mills, D., Foale, S., Abernethy, K., Hicks, C., & Monnereau, I. (2017). Exploring 'islandness' and the impacts of nature conservation through the lens of wellbeing. *Environmental Conservation*, *44*(3), 298–309. https://doi.org/10.1017/S0376892917000273

De Jonge, T., Veenhoven, R., & Arends, L. (2014). Homogenizing responses to different survey questions on the same topic: Proposal of a scale homogenization method using a reference

distribution. *Social Indicators Research*, *117*(1), 275–300.

Devagiri, G. M., Kushalappa, C. G., Ravikumar, D., Anilkumar, K., Siddappa, S., Javed, R., Pal, S., Tomar, S., Gurav, M., & Anand, M. (2015). *Western Ghats Household Baseline*. ATREE.

Dolan, P., & Metcalfe, R. (2012). Measuring Subjective Wellbeing: Recommendations on Measures for use by National Governments. *Journal of Social Policy*, *41*(2), 409–427. https://doi.org/10.1017/S0047279411000833

Doyal, L., & Gough, I. (1991). *A theory of human need*. Macmillan International Higher Education.

Ellis, F. (2000). The determinants of rural livelihood diversification in developing countries. *Journal of Agricultural Economics*, *51*(2), 289–302.

ESA. (2016). *Tanzania Sentinel 2 Land Use Land Cover 2016*. European Space Agency Climate Change Initiative. http://geoportal.rcmrd.org/layers/servir%3Atanzania_sentinel2_lulc2016#more

Feeny, S., & McDonald, L. (2016). Vulnerability to multidimensional poverty: Findings from households in Melanesia. *The Journal of Development Studies*, *52*(3), 447–464.

Gough, I., & McGregor, J. A. (2007). *Wellbeing in developing countries: From theory to research*. Cambridge University Press.

Howland, O., Noe, C., & Brockington, D. (2019). The multiple meanings of prosperity and poverty: A cross-site comparison from Tanzania. *The Journal of Peasant Studies*, *0*(0), 1–21. https://doi.org/10.1080/03066150.2019.1658080

Keane, A., Lund, J. F., Bluwstein, J., Burgess, N. D., Nielsen, M. R., & Homewood, K. (2019). Impact of Tanzania's Wildlife Management Areas on household wealth. *Nature Sustainability*. https://doi.org/10.1038/s41893-019-0458-0

Kummu, M., Taka, M., & Guillaume, J. H. (2018). Gridded global datasets for gross domestic product and Human Development Index over 1990–2015. *Scientific Data*, *5*, 180004. https://doi.org/10.1038/sdata.2018.4

Lakerveld, R. P., Lele, S., Crane, T. A., Fortuin, K. P. J., & Springate-Baginski, O. (2015). The social distribution of provisioning forest ecosystem services: Evidence and insights from Odisha, India. *Ecosystem Services*, *14*, 56–66. https://doi.org/10.1016/j.ecoser.2015.04.001

Lipovetsky, S., & Conklin, M. (2018). Decreasing Respondent Heterogeneity by Likert Scales

Adjustment via Multipoles. *Stats*, *1*(1), 169–175.

McGinnis, M. D., & Ostrom, E. (2014). Social-ecological system framework: Initial changes and continuing challenges. *Ecology and Society*, *19*(2), art30. https://doi.org/10.5751/ES-06387-190230

NRSC. (2016). *LULC Map of India 2014-15*. National Remote Sensing Centre.

Olander, L. P., Johnston, R. J., Tallis, H., Kagan, J., Maguire, L. A., Polasky, S., Urban, D., Boyd, J., Wainger, L., & Palmer, M. (2018). Benefit relevant indicators: Ecosystem services measures that link ecological and social outcomes. *Ecological Indicators*, *85*, 1262–1272. https://doi.org/10.1016/j.ecolind.2017.12.001

OPHI. (2018). *Global Multidimensional Poverty Index 2018*. University of Oxford, UK.

Ostrom, E. (2009). A General Framework for Analyzing Sustainability of Social-Ecological Systems. *Science*, *325*(5939), 419–422. https://doi.org/10.1126/science.1172133

Ostrom, E., Gardner, R., Walker, J., Walker, J. M., & Walker, J. (1994). *Rules, games, and common-pool resources*. University of Michigan Press.

Poudyal, M., Rakotonarivo, O. S., Razafimanahaka, J. H., Hockley, N., & Jones, J. P. G. (2018). Household economy, forest dependency & opportunity costs of conservation in eastern rainforests of madagascar. *Scientific Data*, *5*(April), 1–12. https://doi.org/10.1038/sdata.2018.225

Ramirez-Gomez, S. O. I., Torres-Vitolas, C. A., Schreckenberg, K., Honzák, M., Cruz-Garcia, G. S., Willcock, S., Palacios, E., Pérez-Miñana, E., Verweij, P. A., & Poppy, G. M. (2015). Analysis of ecosystem services provision in the Colombian Amazon using participatory research and mapping techniques. *Ecosystem Services*, *13*, 93–107. https://doi.org/10.1016/j.ecoser.2014.12.009

Robinson, B. E., Zheng, H., & Peng, W. (2019). Disaggregating livelihood dependence on ecosystem services to inform land management. *Ecosystem Services*, *36*. https://doi.org/10.1016/j.ecoser.2019.100902

RRI. (2012). *What Rights? A Comparative Analysis of Developing Countries' National Legislation on Community and Indigenous Peoples' Forest Tenure Rights*. Rights and Resources Initiative.

Schlager, E., & Ostrom, E. (1992). Property-Rights Regimes and Natural Resources: A Conceptual Analysis. *Land Economics*, *68*(3), 249–262. JSTOR. https://doi.org/10.2307/3146375

Sen, A. (1999). *Development as Freedom* (Vol. 2). Oxford University Press.

Simensen, T., Halvorsen, R., & Erikstad, L. (2018). Methods for landscape characterisation and mapping: A systematic review. *Land Use Policy*, *75*, 557–569.

Smith, H. E., Ryan, C. M., Vollmer, F., Woollen, E., Keane, A., Fisher, J. A., Baumert, S., Grundy, I. M., Carvalho, M., & Lisboa, S. N. (2019a). Impacts of land use intensification on human wellbeing: Evidence from rural Mozambique. *Global Environmental Change*, *59*, 101976.

Smith, H. E., Ryan, C. M., Vollmer, F., Woollen, E., Keane, A., Fisher, J. A., Baumert, S., Grundy, I. M., Carvalho, M., & Lisboa, S. N. (2019b). Impacts of land use intensification on human wellbeing: Evidence from rural Mozambique. *Global Environmental Change*, *59*, 101976.

UN DESA. (2019). *World Population Prospects 2019: Methodology of the United Nations population estimates and projections (ST/ESA/SER.A/425)*. United Nations, Department of Economic and Social Affairs, Population Division.

Weiss, D., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., & others. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333. https://malariaatlas.org/research-project/accessibility-to-cities/

White, S. C. (2010). Analysing wellbeing: A framework for development practice. *Development in Practice*, *20*(2), 158–172.

White, S. C. (2015). *Relational wellbeing: A theoretical and operational approach*. Bath Papers in International Development and Wellbeing.

World Bank. (2020). *World Development Indicators database. PPP conversion factor, GDP (LCU per international $).* International Comparison Program, World Bank.

# Appendices

## A. Thresholds and transformations for human wellbeing indicators

### Thresholds for generating binary wellbeing indicators

Table A.1 contains descriptions of the thresholds used to generate each wellbeing variable for each dataset. Green means the exact variable exists in the respective dataset, blue means an alternative was used, while yellow means there is no coverage.

### Transforming and combining the life satisfaction variables

Five of the original household surveys asked similar questions on life satisfaction:

- *ACES*: All things considered, how satisfied are you with your life in the last 12 months? *5-point likert response: Satisfied; Neither satisfied nor dissatisfied, Dissatisfied, Very dissatisfied*
- *SPACES*: All things considered, how satisfied are you with your life as a whole these days? *4-point likert response: Very satisfied, Satisfied, dissatisfied, very dissatisfied*
- *DELTAS*: Global Satisfaction of life: Taking everything into consideration, which rung do you think you are at present? *10-point response on a 'ladder', 10 being lowest to 1 being highest*
- *ASSETS:* How satisfied are you with your life overall? *5-point likert: Very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied, very dissatisfied*
- *SENTINEL*: How would you describe your current feeling about your life as a whole? *5-point likert response: Very satisfied, satisfied, neither satisfied nor dissatisfied, dissatisfied, very dissatisfied*

There are two main issues with standardising and combining ordinal likert-type scales to compare responses across datasets. First, due to local cultural and linguistic effects we do not know whether word values (e.g. 'satisfied') have equivalent meanings between datasets, nor how people are anchoring their responses according to the minimum, central or maximum values (e.g. does 'dissatisfied' mean 'I am dissatisfied or above' or 'I am dissatisfied or below') (De Jonge et al., 2014; Lipovetsky & Conklin, 2018).

To overcome this issue, we framed the combined life satisfaction variable as 'relative within-site life satisfaction' i.e. it does not measure differences in life satisfaction between countries, but the relative satisfaction of a household within a site/dataset.

*Table A1. Wellbeing thresholds*

| Dimension | Observed indicators | Variable name | Description | SPACES (MOZ, KEN) | DELTAS (BGD) | P4GES (MDG) | PIMA (TZA) | ACES (MOZ) | ASSETS (MWI, COL, PER) | ORISSA (IND) | SENTINEL (IND) | MIYUN (CHN) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Education | Years of Schooling | edu_yrs | No household member aged 10 years or older has completed six years of schooling.<br><br>Alternative 1: Household head has not completed more than six years of schooling | Y | Y | Y | Alt. 1 | Y | Y | Y | Y | Alt. 1 |
| | Child School Attendance | edu_chld_att | Any school-aged child is not attending school for at least six years | Y | Y | N | N | Y | Y | Y | Y | Y |
| Health | Physical health | hlt_chld_mt | Any child below 18 has died in the family.<br><br>Alternative 1: has any child below 5 years died in the last year<br><br>Alternative 2: Severe illness in family (or of breadwinner) in last year | Alt. 2 | Alt. 1 | N | Alt. 2 | Alt. 1 | Y | N | Alt. 2 | Alt. 2 |
| | Food security | hlt_fd_sec | Reported not having enough to eat in the last 12 months<br><br>Alternative 1: Reported a decline in ability to provide food in the last 5 years<br><br>Alternative 2: Household is in lowest quantile of food diversity or food consumption per adult equivalent | Y | Alt. 2 | Y | Alt. 1 | Y | Y | N | Y | Alt. 2 |
| | Protein consumption | hlt_pt | Did not eat meat in last week, or every week on average over period | Y | Y | N | Y | Y | Y | N | Y | N |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Living standards | Electricity | ls_elec | The household has no electricity<br><br>A1: Main source of light is not electric mains<br><br>A2: Always has enough light in the home | Y | Alt 1. | Alt. 2 | N | N | Y | N | Alt. 1 | N |
| | Improved Sanitation | ls_san | The household's sanitation facility is not flush toilet or latrine, or ventilated improved pit or composting toilet | Y | Y | N | N | Y | Y | N | Y | Y |
| | Safe Drinking Water | ls_wat | Is not piped water, public tap, borehole or pump, protected well, protected spring or rainwater | Y | Y | N | N | Y | Y | N | Y | Y |
| | Housing | ls_hse | At least two of the three housing materials for roof, walls and floor are inadequate: the floor is of natural materials and/or the roof and/or walls are of natural or rudimentary materials<br><br>Alternative 1: exc. roof material | Y | Y | Y | N | Y | Y | Y | Alt. 1 | N |
| | Cooking Fuel | ls_fuel | The household cooks with dung or wood | Y | Y | N | N | Y | Y | N | Y | N |
| Assets | Valuable assets | ls_ast | The household does not own more than one of the following items: radio, TV, telephone, bike, motorbike computer, or refrigerator (conditional on not owning a car or truck). Or similar assets in each survey. | Y | Y | Y | N | Y | Y | Y | Y | Y |
| | | lv_pas | The household does not own at least one of the following: agricultural land; fishing vessel; livestock | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Risk | Exposure to Risk | rk_exp | Experienced a major shock in the last 12 months<br><br>Alternative 1: 6 months<br><br>Alternative 2: 5 years<br><br>Alternative 3: resorted to borrowing from loan shark in last 5 years | Y | Y | N | Alt. 2 | Y | Alt. 1 | N | Y | Alt. 3 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Life statisfaction | Reported overall life satisfaction | swb | Lower values mean lower life satisfaction (four point ordinal scale) | Y | Y | N | N | Y | Y | N | Y | N |
| Social relations | Institutional membership | inst | Household is not a member of, or has not received assistance from, one of the organisations or groups asked about in the survey<br><br>Alternative 1: Has successfully been granted community land by the local NRM institution | Y | N | Y | Alt. 1 | Y | Y | Y | Y | Y |
| | Personal relationships | prel | Disatissfied (or below) with friendships or community support<br><br>Alternative 1: No one outside of family to turn to if household needs help<br><br>Alternative 2: Feels unable to exert influence on decisions in NRM organisation asked about in the survey<br><br>Alternative 3: Feels like an outsider in the village | Alt. 1 | Alt. 3 | Alt 1. | Alt. 2 | Y | Y | N | Y | N |

Second, when combining ordinal variables with different scales (e.g. 4-point vs 10-point), we do not know if distances between ordinal values are equivalent between datasets. Thus, direct linear transformations are problematic. An established empirical method for combining such variables is the Reference Distribution method (De Jonge et al., 2014). We used this method to transform the 10- and 5-point scales to 4-point scales through the following steps: 1) evaluate the cumulative distribution functions (CDFs) of the various datasets; 2) choose as a reference distribution the 4-point CDF if it approximates all other CDFs, and accords with expected CDFs from theory; 3) recode the 10- and 5-point scales to 4-point scales so that their new transformed CDFs approximates the reference CDF (i.e. by binning values). This provides a 'relative within-site life satisfaction' variable that we can analyse across datasets.

Figure A.1 below shows the original CDFs of the life satisfaction variables, while Figure 2 shows the CDFs after converting to four points scales using the SPACES empirical CDF as a reference distribution.

*Figure A1.*

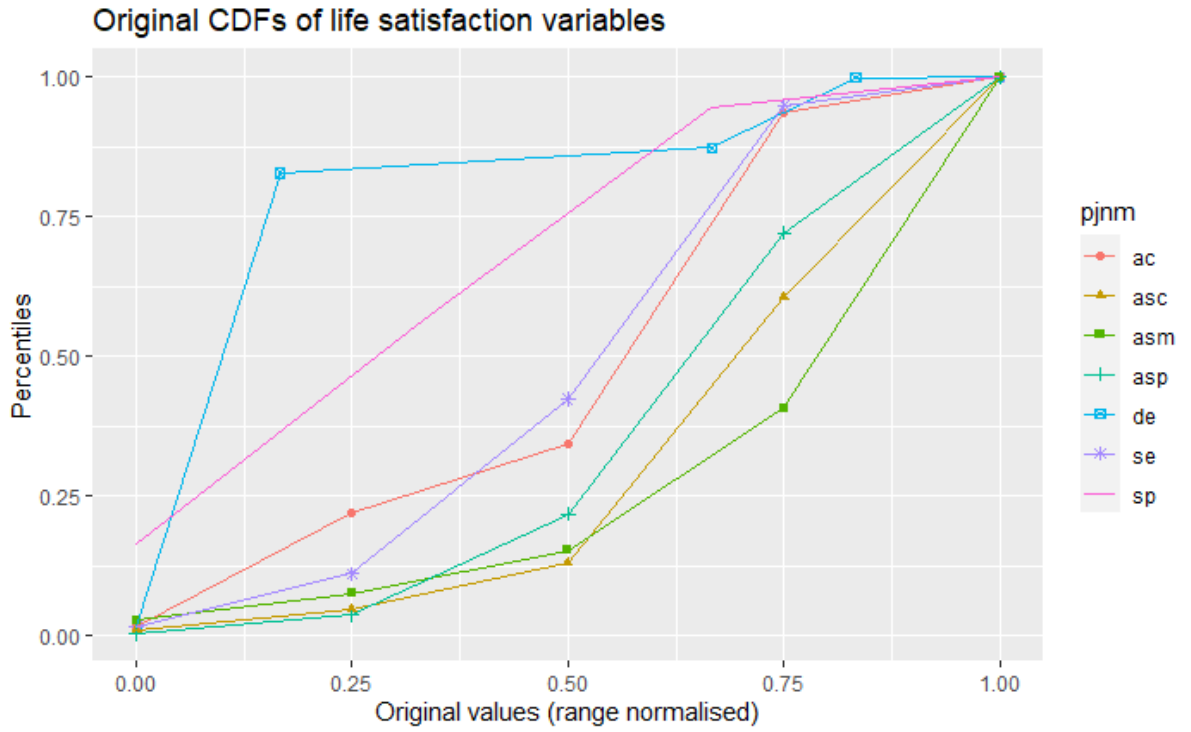
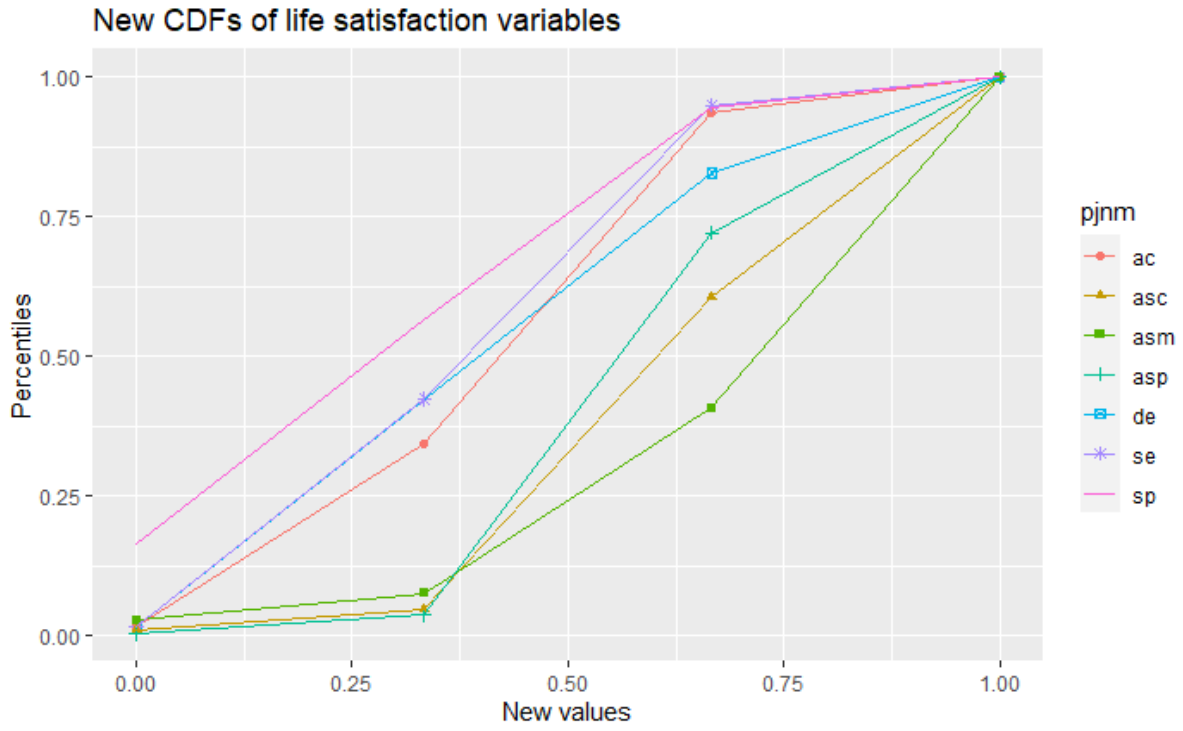


*Figure A2.*


New CDFs of life satisfaction variables
Percentiles
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
New values
pjnm
ac
asc
asm
asp
de
se
sp