**S3 supporting information**

Each of the eight model sets were selected because they were hypothesised to be potential drivers or effective predictors of forest cover (Table S3.1). The predictors within each of the sets were selected as proxies for the set because of their relevance, or because they were the best quality data that related to the set.

**Table S3.1.** **Hypothesised relationships between socioeconomic variables and forest cover.**

|  |  |  |
| --- | --- | --- |
| **Set** | **Hypotheses** | **Variable(s)** |
| Demographics | Communes/provinces with higher human populations and higher human population density will have lower forest cover due to urbanisation and agricultural expansion. Communes/provinces with higher indigenous populations will have higher forest cover because areas with high indigenous populations are more remote, and indigenous communities rely more on forests for traditional livelihoods. | Total population |
| Population density |
| Proportion indigenous |
| Education | Communes/provinces with lower levels of education will have lower forest cover because logging and forest clearance is conducted predominantly by young males of school age. Alternative hypothesis: communes/provinces with higher levels of education will have lower forest cover because education levels are likely to be higher in urban areas. | Proportion of males aged 6 – 24 in full time education |
| Employment | Communes/provinces with higher proportions of adults in the primary sector will have higher forest cover because these areas are likely to be more remote and have more natural resources such as forests. Communes/provinces with higher proportions of adults in the secondary sector will have lower forest cover as the secondary sector will be more prominent in urban/developed areas. | Proportion of adults employed in the primary sector |
| Proportion of adults employed in the secondary sector |
| Economic security | Communes/provinces with higher proportions of families with poor economic security (farmland, livestock) will have lower forest cover because rural populations in areas with high forest cover have access to land and livestock, whereas poor families in urban/developed areas do not. | Proportion of families with <1ha of rice land |
| Proportion of families who keep pigs |
| Access to services | Communes/provinces with large distances to schools are likely to be large, remote communes/provinces with high forest cover. Alternative hypothesis: areas with large distances to schools will lead to higher proportions of males out of education and engaging in forest clearing activities. Communes/provinces with higher proportions of families with access to waste collection will be in developed, urban areas and will have lower forest cover. Communes/provinces with larger distances to commune offices will be larger, more remote areas with higher forest cover. Alternative hypothesis: communes/provinces with larger distances to commune offices will have weaker governance and less law enforcement, resulting in lower forest cover. | Distance to nearest school |
| Proportion of families with access to waste collection |
| Distance to the Commune office |
| Crime and legal disputes | Communes/provinces with a higher number of criminal cases will be more urbanised area and therefore will have lower forest cover. Communes/provinces with a higher number of land conflicts will be in areas of high forest cover where land speculation and land disputes are high. Alternative hypothesis: communes/provinces with a higher number of land conflicts will be in areas with a high number of economic land concessions where forest clearance has occurred, and so will have lower forest cover. | Number of criminal cases |
| Number of land conflicts |
| Migration | Communes/provinces with a high number of in-migrants will be urban areas with large industry (i.e., high job availability) and therefore low forest cover. Alternative hypothesis: communes/provinces with a high number of in-migrants will be areas with new economic land concessions which are often in areas of high forest cover. Communes/provinces with a high number of out-migrants will have higher forest cover because they are rural, remote areas with fewer job opportunities. | Number of in-migrants |
| Number of out-migrants |
| Control | All of these variables have potential to influence forest cover within communes/provinces, yet were not specific targets for investigation. Therefore they were included as control variables. | Mean elevation |
| Distance to international border |
| Distance to provincial capital |
| Presence of economic land concessions |
| Presence of protected areas |
| Protected area category |

*Data processing*

**Table S3.2.** **European Space Agency Climate Change Initiative satellite bands. Bands highlighted in green were grouped to represent “forest cover” in both the macroeconomic and socioeconomic analyses.**

|  |  |
| --- | --- |
| Value | Label |
| 0 | No data |
| 10 | Cropland, rainfed |
| 11 | Herbaceous cover |
| 12 | Tree or shrub cover |
| 20 | Cropland, irrigated or post-flooding |
| 30 | Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%) |
| 40 | Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%) |
| 50 | Tree cover, broadleaved, evergreen, cosed to open (>15%) |
| 60 | Tree cover, broadleaved, deciduous, closed to open (>15%) |
| 61 | Tree cover, broadleaves, decisuous, closed (>40%) |
| 62 | Tree cover, broadleaves, deciduous, open (15 - 40%) |
| 70 | Tree cover, needleleaved, evergreen, closed to open (>15%) |
| 71 | Tree cover, needleleaved, evergreen, closed (>40%) |
| 72 | Tree cover, needleleaved, evergreen, open (15 - 40%) |
| 80 | Tree cover, needleleaved, deciduous, closed to open (>15%) |
| 81 | Tree cover, needleleaved, deciduous, closed (>40%) |
| 82 | Tree cover, needleleaved, deciduous, open (15 - 40%) |
| 90 | Tree cover, mixed leaf type (broadleaved and needleleaved) |
| 100 | Mosaic tree and shrub (>50%) / herbaceous cover (<50%) |
| 110 | Mosaic herbaceous cover (>50%) / tree and shrub (<50%) |
| 120 | Shrubland |
| 121 | Evergreen shrubland |
| 122 | Deciduous shrubland |
| 130 | Grassland |
| 140 | Lichens and mosses |
| 150 | Sparse vegetation (tree, shrub, herbaceous cover) (<15%) |
| 152 | Sparse shrub (<15%) |
| 153 | Sparse herbaceous cover (<15%) |
| 160 | Tree cover, flooded, fresh or brakish water |

*Socioeconomic data cleaning*

Prior to aggregation to the commune level, village data were checked for missing values. In some cases, villages had data for a subset of years but were missing data for other years. If the missing data were at the start of the study period or the end of the study period it was assumed that the village was either an old or a new village. Villages can be merged with larger villages, or two sub-villages, or “*Kroms*”, can be split into two distinct villages over time for administrative purposes. In these cases, the rows (years) with missing data were deleted, but the years with data were retained as these represent villages that existed in that year. If the missing data were in the middle of the study period (for more than one year), or if data for that village only exists for one or several years in the middle of the study period, then the data were assumed to be incomplete and the village was deleted. If the village had data for all years except one, then the missing values were estimated using linear interpolation. If the village existed in all years, but was missing data from multiple years, the village was deleted. If an entire commune was missing in some years, the commune was deleted. The above cleaning process removed 312 villages (total number of villages = 84,195), or 0.37% of the data. Data were then split into individual years, and the final village-level data were aggregated to the commune- and province level using the operations defined below in Table S5.

After aggregation, each variable was checked for obvious errors or unlikely outliers via plotting of histograms and trends. Plots were done at the province level first, to identify any communes within a province that had particularly unusual values or trends. If unusual values or trends were identified the commune was investigated in more detail. Outlier values that appeared inconsistent or implausible were removed and replaced with a value estimated via linear interpolation (Figure S1). In some cases, where data had been converted from raw values to a proportion of the total population, errors in the raw data were discovered. This became clear when the resulting proportion was >1. In these cases, the proportion was changed to 1.

**Table S3.3. Mathematical operations used to aggregate socioeconomic variables from the village to the commune and province level.**

|  |  |
| --- | --- |
| **Variable** | **Operation** |
| Total population | Sum |
| Number of families | Sum |
| Number of males aged 18-64 | Sum |
| Number of females aged 18-64 | Sum |
| Number of people aged over 61 | Sum |
| Total number of indigenous people | Sum |
| Number of families whose main occupation is farming | Sum |
| Number of land conflict cases | Sum |
| Number of in-migrants | Sum |
| Number of out-migrants | Sum |
| Number of criminal cases | Sum |
| Proportion of population that is indigenous | Mean |
| Proportion of females aged 6-24 in full time education | Mean |
| Proportion of males aged 6-24 in full time education | Mean |
| Proportion of females aged 15-45 who are illiterate | Mean |
| Proportion of males aged 15-45 who are illiterate | Mean |
| Proportion of families whose main occupation is farming | Mean |
| Proportion of people who are primarily employed in the primary sector | Mean |
| Proportion of people who are primarily employed in the secondary sector | Mean |
| Proportion of people who are primarily employed in the tertiary sector | Mean |
| Proportion of people who are primarily employed in the quaternary sector | Mean |
| Proportion of families who have less than 1ha of farmland | Mean |
| Proportion of families who have buffalo | Mean |
| Proportion of families who have pigs | Mean |
| Proportion of families who have access to waste collection | Mean |
| Number of infant (<6mo) mortality cases | Mean |
| Number of child (<5 years old) mortality cases | Mean |
| Distance to the nearest school | Median |
| Distance to the Commune Office | Median |
| Distance to the nearest health centre | Median |

A picture containing shape

Description automatically generated

**Figure S3.1.** **An example of linear interpolation for a commune with an implausible outlier. The example shows a value for the population of a commune in 2010 which is likely to be an error (solid line), and the resulting correction (dashed line).**

*Correlation*

For both analyses, correlation of predictors was assessed.

For the socioeconomic variables, correlation was assessed within each variable set. If there were incidents of high correlation, a principal component analysis (PCA) was conducted to see which variables explained the most variance. Based on these analyses, the following decisions were made:

* Total population, number of families, number of males, number of females, and population over 61 were all correlated. Following a PCA, total population was selected.
* As expected, all education variables were highly correlated. In this case, the proportion of males aged 6-24 was selected (without a PCA) because in this cultural context, males are far more likely to be engaged in activities that contribute to forest loss.
* As expected, there was a negative correlation between the proportion of people employed in the primary sector and the proportion of people employed in the tertiary and quaternary sectors, and a correlation between the proportion of people employed in the primary sector and the proportion of people whose main occupation was farming. The PCA results suggested that the proportion of people employed in the primary sector (propPrimSec) and secondary sector (propSecSec) were the most valuable predictors.
* Proportion of people with less than 1 hectare of farmland, and proportion of families who keep buffalos, were dropped due to inconsistencies in the data which suggested changes in the data collection or questions over time.
* Distance to the nearest school (dist\_sch) and distance to the nearest health centre (KM\_Heal\_cen) were correlated, and the PCA analysis was inconclusive about which variable to retain. Distance to school was retained based on the theory that forest clearance activities are more likely to be conducted by young males. An absence of accessible education is likely to be more of a driving factor in these activities than an absence of accessible health care.
* Both healthcare variables (infant mortality and child mortality) were dropped due to poor data quality.

A final assessment of correlation between predictor variables (after removal of the above) revealed no major correlations (Table S7).

**Table S3.4. Correlation matrix for the socioeconomic variables. There were no coefficients greater than 0.6 or less than -0.6.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | tot\_pop | prop\_ind | pop\_den | M6\_24\_sch | propPrimSec | propSecSec | Les1\_R\_Land | pig\_fam | dist\_sch | garbage | KM\_Comm | land\_confl | crim\_case | Pax\_migt\_in | Pax\_migt\_out | mean\_elev | dist\_border | dist\_provCap |
| tot\_pop |  | -0.33 | 0.42 | 0.19 | -0.31 | 0.04 | 0.15 | -0.28 | -0.36 | 0.10 | 0.00 | 0.32 | -0.10 | 0.39 | 0.35 | -0.34 | 0.14 | -0.10 |
| prop\_ind | -0.33 |  | -0.22 | -0.36 | 0.15 | -0.04 | -0.16 | 0.20 | 0.43 | -0.04 | 0.21 | -0.08 | 0.13 | -0.12 | -0.14 | 0.47 | -0.18 | 0.06 |
| pop\_den | 0.42 | -0.22 |  | 0.30 | -0.46 | 0.16 | 0.20 | -0.26 | -0.35 | 0.35 | -0.21 | -0.03 | -0.12 | 0.11 | 0.05 | -0.31 | 0.01 | -0.39 |
| M6\_24\_sch | 0.19 | -0.36 | 0.30 |  | -0.09 | 0.01 | 0.11 | 0.07 | -0.37 | 0.03 | -0.21 | 0.04 | -0.09 | -0.04 | -0.06 | -0.20 | 0.07 | -0.21 |
| propPrimSec | -0.31 | 0.15 | -0.46 | -0.09 |  | -0.26 | 0.09 | 0.50 | 0.24 | -0.32 | 0.14 | 0.05 | 0.01 | -0.18 | -0.21 | 0.02 | 0.08 | 0.27 |
| propSecSec | 0.04 | -0.04 | 0.16 | 0.01 | -0.26 |  | 0.02 | -0.14 | -0.08 | 0.05 | -0.05 | -0.02 | -0.02 | 0.04 | 0.03 | -0.02 | 0.04 | -0.09 |
| Les1\_R\_Land | 0.15 | -0.16 | 0.20 | 0.11 | 0.09 | 0.02 |  | 0.01 | -0.23 | -0.08 | -0.11 | -0.01 | -0.09 | 0.07 | 0.01 | -0.21 | 0.20 | -0.14 |
| pig\_fam | -0.28 | 0.20 | -0.26 | 0.07 | 0.50 | -0.14 | 0.01 |  | 0.19 | -0.14 | 0.06 | -0.02 | -0.03 | -0.21 | -0.21 | 0.02 | -0.10 | 0.15 |
| dist\_sch | -0.36 | 0.43 | -0.35 | -0.37 | 0.24 | -0.08 | -0.23 | 0.19 |  | -0.07 | 0.36 | -0.06 | 0.07 | -0.12 | -0.15 | 0.36 | -0.14 | 0.38 |
| garbage | 0.10 | -0.04 | 0.35 | 0.03 | -0.32 | 0.05 | -0.08 | -0.14 | -0.07 |  | -0.06 | -0.03 | 0.04 | 0.05 | 0.02 | 0.00 | -0.05 | -0.13 |
| KM\_Comm | 0.00 | 0.21 | -0.21 | -0.21 | 0.14 | -0.05 | -0.11 | 0.06 | 0.36 | -0.06 |  | 0.09 | 0.03 | 0.04 | 0.00 | 0.11 | -0.05 | 0.24 |
| land\_confl | 0.32 | -0.08 | -0.03 | 0.04 | 0.05 | -0.02 | -0.01 | -0.02 | -0.06 | -0.03 | 0.09 |  | 0.27 | 0.13 | 0.05 | -0.06 | 0.04 | 0.07 |
| crim\_case | -0.10 | 0.13 | -0.12 | -0.09 | 0.01 | -0.02 | -0.09 | -0.03 | 0.07 | 0.04 | 0.03 | 0.27 |  | -0.03 | -0.05 | 0.16 | -0.13 | 0.02 |
| Pax\_migt\_in | 0.39 | -0.12 | 0.11 | -0.04 | -0.18 | 0.04 | 0.07 | -0.21 | -0.12 | 0.05 | 0.04 | 0.13 | -0.03 |  | 0.42 | -0.10 | 0.01 | 0.01 |
| Pax\_migt\_out | 0.35 | -0.14 | 0.05 | -0.06 | -0.21 | 0.03 | 0.01 | -0.21 | -0.15 | 0.02 | 0.00 | 0.05 | -0.05 | 0.42 |  | -0.12 | 0.02 | -0.03 |
| mean\_elev | -0.34 | 0.47 | -0.31 | -0.20 | 0.02 | -0.02 | -0.21 | 0.02 | 0.36 | 0.00 | 0.11 | -0.06 | 0.16 | -0.10 | -0.12 |  | -0.26 | 0.15 |
| dist\_border | 0.14 | -0.18 | 0.01 | 0.07 | 0.08 | 0.04 | 0.20 | -0.10 | -0.14 | -0.05 | -0.05 | 0.04 | -0.13 | 0.01 | 0.02 | -0.26 |  | -0.05 |
| dist\_provCap | -0.10 | 0.06 | -0.39 | -0.21 | 0.27 | -0.09 | -0.14 | 0.15 | 0.38 | -0.13 | 0.24 | 0.07 | 0.02 | 0.01 | -0.03 | 0.15 | -0.05 |  |

*Modelling*

**Table S3.5. Within-set models for the commune-level socioeconomic analysis. Maximal within-set models were run followed by subsequent, less complex models, to identify the most important variables within each set. The variables with the largest effects were taken forward to the final candidate set. Only variables with negligible effects were dropped. If a variable set only had one variable it was automatically taken forward. All models included an offset term which was the logged commune area (km2), and a random effects structure of the form ~(year|Province/Commune).**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model set / model | Variables | | | | |
| *Population demographics* |  |  |  |  |  |
| popdem.m1 | Population density | Proportion indigenous |  |  |  |
| popdem.m2 | Population density |  |  |  |  |
| *Education* |  |  |  |  |  |
| edu.m1 | Proportion males in school |  |  |  |  |
| *Employment* |  |  |  |  |  |
| emp.m1 | Proportion primary sector | Proportion secondary sector |  |  |  |
| emp.m2 | Proportion primary sector |  |  |  |  |
| *Economic security* |  |  |  |  |  |
| econ.m1 | Proportion no farmland | Proportion with pigs |  |  |  |
| econ.m2 | Proportion with pigs |  |  |  |  |
| *Access to services* |  |  |  |  |  |
| acc.m1 | Distance to school | Access to waste collection | Distance to commune office |  |  |
| acc.m2 | Distance to school | Access to waste collection |  |  |  |
| acc.m3 | Access to waste collection |  |  |  |  |
| *Crime and legal disputes* |  |  |  |  |  |
| jus.m1 | Criminal cases | Land conflicts |  |  |  |
| jus.m2 | Criminal cases |  |  |  |  |
| *Migration* |  |  |  |  |  |
| mig.m1 | In-migration \* | Out-migration |  |  |  |
| mig.m2 | In-migration | Out-migration |  |  |  |
| mig.m3 | Out-migration |  |  |  |  |
| mig.m4 | In-migration |  |  |  |  |
| *Control* |  |  |  |  |  |
| env.m1 | Elevation |  |  |  |  |
| hum.m1 | Distance to International border | Distance to provincial capital | Presence of ELC | Presence of PA | PA category |
| hum.m2 | Distance to International border | Distance to provincial capital | Presence of ELC | Presence of PA |  |

\* Indicates interaction

**Table S3.6. Final candidate model set for the commune-level socioeconomic analysis. Variables were selected based on the results of the within-set models (see table S8). All models included an offset term of the logged commune area (km2), and a random effects structure of the form ~(year|Province/Commune). Model selection was done using an Information Theoretic approach. Shaded model (m1) was the top model and was used for predictions and inference.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Delta AIC** | **Variables** | | | | | | |
| m1 | 0 | Population density | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m2 | 47.91 | Proportion males in school | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m3 | 47.54 | Proportion primary sector | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m4 | 46.79 | Proportion with pigs | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m5 | 49.36 | Distance to school | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m6 | 49.52 | Criminal cases | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m7 | 48.61 | Out-migration | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |  |
| m8 | 49.86 | Proportion males in school | Distance to school | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |
| m9 | 48.35 | Proportion primary sector | Out-migration | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |
| m10 | 18.82 | Population density | Proportion males in school | Elevation | Distance to Int’l border | Distance to provincial capital | Presence of ELC | Presence of PA |

Chart

Description automatically generated

**Figure S3.2. Quantile-quantile plots for the random effect “Province” of the final socioeconomic model.**

Chart, histogram

Description automatically generated

**Figure S3.3. Quantile-quantile plots for the random effect “Commune” of the final socioeconomic model. Plots suggest the assumption of normality of deviations of the conditional means of the random effects from the global intercept is violated.**

Chart, scatter chart

Description automatically generated

**Figure S3.4. Plot of residuals versus fitted values for the final socioeconomic model.**

**Table S3.7. Final candidate model set for the province-level socioeconomic analysis. Variables were selected based on the results of the within-set models (see table S8). All models included a random effects structure of the form ~(year|Province). Model selection was done using an Information Theoretic approach. Shaded model (m8) was the top model and was used for predictions and inference.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Delta AIC** | **Variables** | | | | | | |
| M1 | 57.38 | Population density | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M2 | 58.41 | Proportion males in school | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M3 | 58.13 | Proportion primary sector | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M4 | 59.46 | Proportion with pigs | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M5 | 5.04 | Distance to school | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M6 | 61.55 | Criminal cases | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M7 | 61.65 | Out-migration | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |  |
| M8 | 0.00 | Males in school | Distance to school | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |
| M9 | 58.38 | Primary sector | Out-migration | Elevation | Distance to int’l border | Distance to provincial capital | Presence of ELCs | Presence of PAs |
| M10 | 203.95 | Population density |  |  |  |  |  |  |
| M11 | 210.35 | Land conflicts | Criminal cases |  |  |  |  |  |
| M12 | 214.36 | In-migration | Out-migration |  |  |  |  |  |
| M13 | 209.36 | Proportion males in school |  |  |  |  |  |  |
| M14 | 209.48 | Proportion primary sector | Proportion secondary sector |  |  |  |  |  |
| M15 | 207.60 | Proportion no farmland | Proportion with pigs |  |  |  |  |  |
| M16 | 212.50 | Distance to school |  |  |  |  |  |  |
| M17 | 100.59 | Elevation |  |  |  |  |  |  |
| M18 | 132.49 | Distance to int’l border | Distance to provincial capital |  |  |  |  |  |
| M19 | 181.77 | Presence of ELCs | Presence of PAs |  |  |  |  |  |

Chart

Description automatically generated

**Figure S3.5. Cambodian provinces clustered using unweighted pair-group using arithmetic averages (UPGMA). Clustering was based on a selection of the socioeconomic variables used during the modelling. Data were averaged across the study period 2007 – 2012. Variables included were total population, population density, number of land conflict cases, number of criminal cases per capita, number of in- and out-migrants, the proportion of the population classified as indigenous, proportion of males aged 6 – 24 in school, proportion of the population employed in the primary and secondary sectors, proportion of families with no access to agricultural land, proportion of families who kept pigs, distance to the nearest school, proportion of families with access to waste collection, and distance to the commune (administrative) centre.**