**Results**

***Macroeconomic analysis***

*Rate of forest loss response*

Models revealed that there were no strong effects of the macroeconomic predictors on forest loss between 1993 and 2015 (Figures S2 – S4). For each predictor set there were between 5 and 28 models in the top model set and final coefficients were calculated using full averages (Tables S9 – S17, Burnham and Anderson, 2007). The largest effects were from two of the control variables (population density, time). The largest effect overall was for population density with a one-year time lag (full averaged coefficient = -632.9, SE = 64.8, Table S10). The largest effect excluding control variables was for agricultural proportion of GDP with a one-year time lag (full averaged coefficient = -14.9, SE = 7.9) suggesting that there is a small reduction in the rate of forest loss as the contribution of agriculture to national GDP increases, although this effect is very weak (Figure S2, Table S10) and is not considered definitive.

*New economic land concession response*

There were 287 new ELCs allocated within the study period, with the majority (51%) being designated for rubber production (Table S18). The most valuable crop in terms of commodity price during the study period was rubber, with a mean market price of $1743/ton, followed by rice ($348/ton) and sugar ($282/ton, Table S18). The most valuable crop in terms of producer (farmgate) prices was sugar with a mean price over the study period of $2115/ton, followed by rubber ($317/ton) and corn ($197/ton, Table S18). A greater number of effects were revealed in the macroeconomic analysis with new economic land concession allocation as the response. The largest effect was for the economic control variable population density, where there were very strong negative effects across all time lags (rate ratios for one-year lag = 0.012, two-year lag = 0.002, three-year lag = 0.0005, Table 3), indicating that new ELCs do not get allocated in areas of high human population density. The largest overall effect excluding control variables was for changes in agricultural proportion of GDP with no time lag and a one-year time lag (no time lag rate ratio = 1.310, and one-year time lag rate ratio = 1.284, Table 3, Figure 2).

From an economic perspective there were positive relationships between the allocation of new ELCs and increases in the agricultural proportion of GDP and increases in foreign direct investment (one-year time lag rate ratio = 1.004, Table 3, Figure 2). These effects suggest ties between both the development of new industrial-scale concessions and the growth of the agricultural sector, and the injection of foreign wealth into the sector via the purchasing of concessions by international companies. For example, when the agricultural sector’s proportion of national GDP decreases by 3% in a given year relative to the previous year, the number of new ELCs allocated that year is predicted to be approximately 2, whereas when the sector’s proportion of national GDP increases in a given year by 1% relative to the previous year, the number of new ELCs is predicted to be 6. When the amount of foreign investment decreases by approximately $10 million relative to the previous year, the number of new ELCs one year later is predicted to be 3. Conversely, when foreign investment in a given year increases by approximately $300 million relative to the previous year, then one year later the number of new ELCs is predicted to be 10. The one-year time lag of the effect of foreign investment suggests that it takes approximately one year from the time of investment for a company to see the creation of their land concession. There was also a positive relationship between new ELC allocation and increases in development flows to the environment sector (no time lag rate ratio = 1.031). This suggests that in the short-term, investments into the environment sector via development funding (predominantly from international donors) does not reduce the number of new ELC allocations.

There was a negative relationship between new ELC allocation and increases in per capita GDP (one-year time lag rate ratio = 0.985 and two-year time lag rate ratio = 0.974, Table 3, Figure 2). The reduction in ELC allocation as GDP increases, over a period of one and two years, potentially suggests that there is a positive economic effect of ELCs. New concessions inject money into the national economy at various scales, for example at the national level via taxes to the government, and to the local level via employment opportunities and infrastructure development. Thus, as the economy grows, the need for new ELCs diminishes. For example, when GDP per capita in a given year falls by approximately $6 relative to the previous year, the number of new ELCs is predicted to be 8, whereas when the GDP per capita rises in a given year by approximately $60 relative to the previous year, the number of new ELCs predicted is only 3.

The largest effect within the commodity set was for the change in market price of rice in the same year as the response (no time lag) with a rate ratio of 1.009 (Table 3). There were further strong positive relationships between the changes in the market price of rubber (no time lag rate ratio = 1.001), the changes in the non-food production index (one-year time lag rate ratio = 1.007), and changes in the market price of sugar (two-year time lag rate ratio = 1.009). Economic land concessions in Cambodia are predominantly agro-industrial concessions, and therefore the positive relationships between the market price of agricultural commodities and new ELC allocations is not surprising. Rubber and rice are the most valuable market commodities within the variable set, and we can see this reflected in the model; if rubber market prices do not change between years *t* and *t+1* then approximately 4 new ELCs are predicted in year *t+1*, whereas if the price of rubber increases by $1500/ton in year *t*, then approximately 29 new ELCs are predicted in year *t+1*. Similarly, if there is no change in the market price of rice between two given years, then approximately 5 new ELCs are predicted. If the market value increases by $300/ton then in year *t+1* approximately 80 new ELCs are predicted. Interestingly the effect of changes in sugar price were weak when there was no time lag, but the effect was stronger when both a one-year and two-year time lag were introduced (Figure 3).

There were three negative relationships between ELC allocation and commodity variables, all of which were in the same year as the response (no time lag, Figure 3). There were weak negative effects of changes in the market prices of corn (no time lag rate ratio = 0.997) and sugar (no time lag rate ratio = 0.999). Considering the stronger positive effects of sugar price on ELCs after one- and two-year lags, it is unlikely that the very weak negative effect with no time lag is meaningful. The non-food production index had a much stronger negative effect on ELC allocation when there was no time lag (rate ratio = 0.990). The change in direction of the effect of the non-food production index between no time lag and a one-year time lag suggests a complex relationship between the index and ELCs.

The producer price variable set, which reflects the farmgate prices of the commodities, had both positive and negative relationships with ELC allocation (Figure 4, Table 3). The strongest positive relationship was with changes in the producer price of rubber (no time lag rate ratio = 1.035). The effect of positive changes (i.e., net increases) in the price a farmer will get for rubber production can be seen in the predictions of new ELCs (Figure 4). The difference between the number of ELC allocations when the producer price of rubber changes from a decrease of $30/ton (from year *t* to year *t+1*) to no change at all (i.e., the price remains constant) is approximately 3. In contrast, the difference in ELC allocation between no change in price and a positive change of $30/ton is more than 12. This suggests that producers are highly influenced by sale prices of commodities, particularly of high value products such as rubber, and that they will act quickly when there is the potential for financial gain. There were also positive relationships between ELC allocation and changes in the producer price of corn (one-year time lag rate ratio = 1.011) and the producer price of rice (two-year time lag rate ratio = 1.013, Figure 4, Table 3). Corn and rice are less valuable in terms of absolute producer prices than sugar and rubber, and this may be reflected in the time lag that exists between positive changes in the prices and increases in new ELCs.

There were two negative relationships between producer price variables and new ELC allocations (Figure 4). Increases in the producer prices of rice and cassava resulted in fewer predicted ELCs in the same year (no time lag rate ratio = 0.976) and two years later (two-year time lag rate ratio = 0.982), respectively. The difference in the direction of the effect of rice producer prices in year *t* and year *t+2* (Figure 4) suggests that there is a complex relationship between rice production and new ELC allocation. Rice production is the dominant agricultural crop in Cambodia and is the second most valuable commodity in terms of market value (Table S18). Yet only 1.7% of ELCs created during the study period were designated for rice production (Table S18), suggesting that rice production and price were not driving forces behind ELC allocation. The negative relationship between the producer price of cassava and new ELC allocation was strong (two-year time lag rate ratio = 0.982, Figure 4). Cassava is not a valuable crop, yet it was the third most designated crop for new ELCs during the study period (4.9% of new ELCs, Table S18). It is unclear what is driving the negative relationship between cassava and new ELCs after two years.

**Table 3. Parameter coefficients, standard errors, and rate ratios from the top model(s) in the macroeconomic analysis with rate of economic land concession allocation response. Missing values denote predictor variables that were not selected in the top model(s) for that lag period.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***No time lag*** | | |  | ***1 year time lag*** | | |  | ***2 year time lag*** | | |
| **Variable** | **Coefficient** | **SE** | **Rate ratioa** |  | **Coefficient** | **SE** | **Rate ratioa** |  | **Coefficient** | **SE** | **Rate ratioa** |
| ***Macroeconomic*** |  |  |  |  |  |  |  |  |  |  |  |
| GDP | - | - | - |  | -0.01500 | 0.00340 | 0.985 |  | -0.02600\* | 0.00390 | 0.974 |
| Agricultural proportion of GDP | 0.27000 | 0.07000 | 1.310 |  | 0.25000 | 0.06600 | 1.284 |  | -0.03400\* | 0.07600 | 0.967 |
| Development flows - agriculture | - | - | - |  | - | - | - |  | -0.00005\* | 0.00020 | 1.000 |
| Development flows - environment | 0.03100 | 0.00400 | 1.031 |  | - | - | - |  | -0.00260\* | 0.00450 | 0.997 |
| Foreign direct investment | - | - | - |  | 0.00360 | 0.00050 | 1.004 |  | 0.00040\* | 0.00060 | 1.000 |
| Population density | -4.43000 | 0.85000 | 0.012 |  | -6.09000 | 0.81000 | 0.002 |  | -7.68000\* | 0.95000 | 0.000 |
| Forest remaining | -0.00030 | 0.00004 | 1.000 |  | -0.00004 | 0.00004 | 1.000 |  | 0.00004\* | 0.00005 | 1.000 |
| ***Commodity / production*** |  |  |  |  |  |  |  |  |  |  |  |
| Change in median market price - corn | -0.00330 | 0.005697 | 0.997 |  | 0.00704\* | 0.00647 | 1.007 |  | -0.00365\* | 0.00329 | 0.996 |
| Change in median market price - rice | 0.009324 | 0.00198 | 1.009 |  | -0.00429\* | 0.00272 | 0.996 |  | 0.00004\* | 0.00058 | 1.000 |
| Change in median market price - rubber | 0.001247 | 0.00024 | 1.001 |  | 0.00019\* | 0.00022 | 1.000 |  | -0.00004\* | 0.00009 | 1.000 |
| Change in median market price - sugar | -0.00005 | 0.001931 | 1.00 |  | 0.00708\* | 0.00127 | 1.007 |  | 0.00877\* | 0.00124 | 1.009 |
| Non-food agricultural production index | -0.00995 | 0.00175 | 0.990 |  | 0.00672\* | 0.00264 | 1.007 |  | -0.00149\* | 0.00203 | 0.999 |
| Crop production index | - | - | - |  | 0.00042\* | 0.00144 | 1.000 |  | -0.00328\* | 0.00427 | 0.997 |
| Total production from forestry | - | - | - |  | 0.00000\* | 0.00000 | 1.000 |  | 0.00000\* | 0.00000 | 1.000 |
| Forest remaining | -0.00014 | 0.00002 | 1.000 |  | -0.00017\* | 0.00003 | 1.000 |  | -0.00013\* | 0.00003 | 1.000 |
| ***Producer prices*** |  |  |  |  |  |  |  |  |  |  |  |
| Producer price of corn | 0.00415 | 0.00355 | 1.004 |  | 0.01093\* | 0.00240 | 1.011 |  | 0.00014\* | 0.00081 | 1.000 |
| Producer price of rice | -0.02465 | 0.00436 | 0.976 |  | 0.00452\* | 0.00564 | 1.005 |  | 0.01258\* | 0.00474 | 1.013 |
| Producer price of rubber | 0.03424 | 0.00401 | 1.035 |  | -0.00075\* | 0.00228 | 0.999 |  | -0.00431\* | 0.00467 | 0.996 |
| Producer price of sugar | 0.00004 | 0.00010 | 1.000 |  | 0.00016\* | 0.00018 | 1.000 |  | 0.00000\* | 0.00006 | 1.000 |
| Producer price of cassava | 0.00032 | 0.00123 | 1.000 |  | 0.00006\* | 0.00076 | 1.000 |  | -0.01791\* | 0.00214 | 0.982 |
| Forest remaining | -0.00023 | 0.00002 | 1.000 |  | -0.00015\* | 0.00002 | 1.000 |  | -0.00013\* | 0.00002 | 1.000 |

\* Coefficients derived from full averaging of models within dAIC < 6.

A Rate ratio = exp(coefficient)

Chart

Description automatically generated

**Figure 2. Modelled relationships between economic predictors and the allocation of new economic land concessions in Cambodia between 1993 – 2015. Top row: no time lag between predictor and response; middle row: 1-year time lag between predictor and response; bottom row: 2-year time lag between predictor and response.**

A picture containing diagram

Description automatically generated

**Figure 3. Modelled relationships between commodity price predictors and the allocation of new economic land concessions in Cambodia between 1993 – 2015. Top two rows: no time lag between predictor and response; third row: 1-year time lag between predictor and response; bottom row: 2-year time lag between predictor and response.**

Diagram, engineering drawing

Description automatically generated

**Figure 4. Modelled relationships between producer price predictors and the allocation of new economic land concessions Cambodia between 1993 – 2015. Top row: no time lag between predictor and response; middle row: 1-year time lag between predictor and response; bottom row: 2-year time lag between predictor and response.**

***Socioeconomic analysis***

*Commune-level model*

Initial within-set model selection resulted in a final candidate set with 10 models and 13 unique variables (Table S19). There was a single top model according to AIC (m1), with all other models having delta AIC values of more than 18 (Table S19). The top model only had one non-control variable - population density (Table 4). The random effects term with the highest variance was Commune (10.45 [SD = 3.23], 60% of the total random effect variance), followed by Province (6.77 [SD = 2.60], 39% of the total random effect variance, Table 4). The variance explained by year at both the commune and province level was low (0.005 [SD = 0.068] and 0.0005 [SD = 0.022] respectively), contributing approximately 1% of the total random effect variance (Table 4). The marginal R2 (fixed effects only) was 0.78 (78%), and the conditional R2 (fixed and random effects) was 1, suggesting that most of the model variance was explained by the fixed effects. The largest positive effect was from mean elevation (rate ratio = 2.861, Table 4) which relates to 0.6 forest pixels (0.06 km2) predicted within an “average” commune (i.e., all other fixed and random effects set to their mean) when mean elevation is at the minimum within the country. When the mean elevation is at the maximum found within the country (and all other terms are set to their mean), the number of forest pixels predicted is 13,380 (1,204 km2). This highlights that higher elevation areas of Cambodia are much more likely to be forested than lower elevation areas. The strongest negative effect was from population density (rate ratio = 0.001, Table 4) which relates to approximately 1.5 predicted forest pixels (0.14 km2) at the minimum value of population density found within the country, contrasting with a prediction of effectively zero (2.22 × 10-16) forest pixels at the highest value of population density within the country. All other model terms, excluding the presence of ELCs, had positive effects on forest cover (Table 4). These effects suggest that remote communes (large distances to provincial capitals) that are centrally located within the country (far away from international borders) are predicted to have high forest cover. Interestingly, although the effects are weak, communes that contain ELCs are predicted to have lower forest cover than those without, and communes with protected areas are predicted to have higher forest cover than those without (Table 4).

The results from the final commune-level model must, however, be viewed with extreme caution because model validation revealed some serious underlying issues. As is suggested by the variance associated with the commune-level random effect term, there was extreme variation between communes for all variables (predictors and response, Figure 5). This between-group variance results in the model being unsuitable for generalised (i.e., ‘global’) predictions (Figure 5). Intercept and slope estimates between communes, even within the same province, varied hugely (Figure 6), and this issue was highlighted in diagnostic plots where we see that the assumption of normality of deviations of the conditional means of the random effects (for commune) from the global intercept is violated (Figure S6). Furthermore, the model residuals displayed heteroskedasticity, with the model predicting particularly poorly for lower values of the response (Figure S7). Therefore, drawing general inferences about the relationships between forest cover and socioeconomics at the country level using this model is inappropriate.

**Table 4. Model outputs and rate ratios from the top models from the socioeconomic analysis. Outputs are for the commune-level analysis and the province-level analysis. Reported coefficients are on the link (log) scale.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Variance** | **Std.Dev** | **Coefficient** | | **SE** | **Rate ratio*a*** |
| ***Commune-level final model*** |  |  | |  |  |  |
| *Random effects* |  |  |  | |  |  |
| Commune (intercept) | 10.4500 | 3.2334 | - | | - |  |
| Year/Commune (slope) | 0.0046 | 0.0680 | - | | - |  |
| Province (intercept) | 6.7730 | 2.6025 | - | | - |  |
| Year/Province (slope) | 0.0005 | 0.0220 | - | | - |  |
| *Fixed effects* |  |  |  | |  |  |
| Intercept | - | - | -4.6240 | | 0.5620 |  |
| Population density | - | - | -7.5140 | | 1.1270 | 0.001 |
| Mean elevation | - | - | 1.0510 | | 0.1220 | 2.861 |
| Distance to In'tl border | - | - | 0.5805 | | 0.2036 | 1.787 |
| Distance to Provincial capital | - | - | 0.6929 | | 0.1114 | 2.000 |
| ELC presence | - | - | 0.0000 | | 0.0025 | 0.999 |
| PA presence | - | - | 0.0093 | | 0.0143 | 1.009 |
| ***Province-level final model*** |  |  |  | |  |  |
| *Random effects* |  |  |  | |  |  |
| Province (intercept) | 1.1762 | 1.0845 | - | | - |  |
| Year/Province (slope) | 0.0058 | 0.0765 | - | | - |  |
| *Fixed effects* |  |  | - | | - |  |
| Intercept | - | - | -2.9900 | | 0.4497 |  |
| Males in school (low) | - | - | 0.0051 | | 0.0019 | 1.002 |
| Distance to school (low) | - | - | -0.0174 | | 0.0022 | 1.002 |
| Mean elevation (low) | - | - | -0.0223 | | 0.0024 | 1.002 |
| Distance to border (low) | - | - | 0.0061 | | 0.0019 | 1.002 |
| Distance to Prov capital (low) | - | - | -0.0072 | | 0.0019 | 1.002 |
| Presence of economic concessions (1) | - | - | 1.9974 | | 0.4090 | 1.505 |
| Presence of PAs (1) | - | - | 2.8063 | | 0.4965 | 1.643 |

*a* Rate ratio = exp(coefficient)

Chart, scatter chart

Description automatically generated

**Figure 5. Predicted relationships (red lines) between socioeconomic variables and forest cover in Cambodia between 2007 – 2012 from the top commune-level model. Predictions are ‘global’ i.e., all random effects were set to their mean values, and thus predictions are not for any specific commune. Black dots are the raw data points of each predictor versus forest cover.**

Diagram, shape, arrow

Description automatically generated

**Figure 6. Predicted relationships between population density and forest cover within Cambodian provinces between 2007 – 2012 using the top commune-level model. Faded grey lines are the predictions for each individual commune within each province. Black lines are the mean provincial predictions, which were computed using the 50% quantile from all commune predictions. Plot panels have non-standard y axis ranges.**

*Province-level model*

The province-level models were run to eliminate the commune-level variation and to identify any broader relationships between forest cover and socioeconomics. A candidate set of 19 models was built and an evaluation of AIC selected a single model (m8) as the top model (Table S20). Model m5 had some support (delta AIC = 5, Table S20) but was a simpler version of m8 and therefore inferences were drawn from m8 alone. The random effects term with the highest variance was Province (1.18 [SD = 1.08], which constituted 99% of the total random effects variance), followed by year (0.006 [SD = 0.077], which was 1% of the total random effects variance). The marginal R2 (fixed effects only) was 0.71 (71%) and the conditional R2 (fixed and random effects) was 0.99 (99%), suggesting that the majority of model variance was explained by the fixed effects. Presence of ELCs and presence of PAs had the largest two positive effects relative to their refences levels (no ELCs, no PAs), suggesting that provinces that have those two features are predicted to also have higher forest cover (rate ratios = 1.51 and 1.64 respectively). In provinces where the proportion of males in school and distance to school are both low, higher levels of forest cover are predicted compared with provinces where these variables are high. Furthermore, in provinces where elevation, distance to an international border, and distance to the provincial capital are low, forest cover is predicted to be higher than in provinces where these variables are high. However, all the above effects are weak (Figures 7 & 8). For example, the difference in the predicted number of forest pixels between a province with a low proportion of males in school and a province with a high proportion (with all other variables set to low), is 200 (18 km2). The difference in the number of predicted forest pixels between a province with low median distances to schools and a province with high median distances (with all other variables set to low), is 689 (62 km2). As standalone figures these appear large, but in the context of the range of the response variable (minimum value of 54 forest pixels to a maximum of 146,876 forest pixels), the effects are relatively weak. Presence of PAs had the largest effect on predicted forest pixels. The number of forest pixels predicted for a province with PA presence is 36,890 (3,320 km2) higher than for a province with no PA presence. This emphasises the relationship between forested land and protected areas in Cambodia. The size of the effects for the two socioeconomic predictors (proportion of males in school, and distance to school) in the top model suggest that these variables have little power to predict forest cover at the provincial level in Cambodia, but that the presence of protected areas and economic land concessions do.

***Cluster analysis***

The UPGMA clustering had the highest cophenetic correlation (0.79) and the lowest Gower distance (254.14) and was therefore selected. The matrix correlation statistic suggested that 4 clusters were optimal, but that between 3 and 7 clusters had very similar support. When divided by 4 clusters, there was a large group (16) of provinces that fell into a single cluster, and so 5 clusters were chosen to add further nuance (Figure 9). The provinces within clusters were geographically contiguous (Figure 10), although clusters that had smaller cophenetic distances (i.e., were closer on the dendrogram, Figure 9) were not necessarily geographically contiguous. The largest cluster (cluster 5) dominated a central strip of the country, separating the smaller, and more similar clusters (Figure 10). Only clusters 2 and 4 were contiguous with each other. These results suggest that provinces often have similar socioeconomic conditions to that of their neighbours, but that there are also distinct regions within the country that can be characterised by their socioeconomics rather than their geography. A heatmap of the socioeconomic variable values for each cluster revealed some distinguishing patterns (Figure 11). The largest cluster (cluster 5) was distinguished by high or very high values of all variables, which translates to generally large provinces with high population density, high education levels, high proportions of primary and secondary sector workers, and high migration (Table 5). This contrasts with cluster 2, which has predominantly low values for the socioeconomic variables which translates to very small provinces with low population density, low levels of education, low levels of primary sector employment (higher secondary sector employment), and low levels of migration (Table 5). Clusters 3 and 4 had the highest levels of migration (and interestingly the highest levels of land conflict), education, and population density, reflecting the presence of two of the three largest cities and significant urban development. Cluster 1 had the lowest population density, education, proportion of secondary sector workers, and migration, reflecting the clusters remote geography and rural character. Provinces within cluster 1 were also the most forested but had also lost the most forest during the study period (Figure 12). Provinces within cluster 5 were generally the next most forested after cluster 1 and had also lost large areas of forest during the study period (Figure 12). Cluster 3 had the least amount of forest, which was expected due to high levels of urbanisation and agriculture. Clusters 1 and 2 had the highest elevation, and clusters 1 and 5 had the highest mean distance to a provincial capital (Figure 12).

A picture containing diagram

Description automatically generated

**Figure 7. Predicted forest cover within each Cambodian province given high and low levels of school attendance (males aged 6 – 24 in school) from the top province-level model. All other variables in the model were set to their reference level (distance to school = low, elevation = low, distance to international border = low, distance to provincial capital = low, economic land concession = yes, protected area = yes).**

A picture containing diagram

Description automatically generated

**Figure 8. Predicted forest cover within each Cambodian province given high and low distances to the nearest school from the top province-level model. All other variables in the model were set to their reference level (school attendance = low, elevation = low, distance to international border = low, distance to provincial capital = low, economic land concession = yes, protected area = yes).**

Chart

Description automatically generated

**Figure 9. Cambodian provinces clustered based on socioeconomics. 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. The clustering method was unweighted pair-group using arithmetic averages (UPGMA).**

Map

Description automatically generated

**Figure 10. Map of Cambodia showing the clusters resulting from the unweighted pair-group using arithmetic averages (UPGMA) method. Provinces are labelled.**

Chart, bar chart

Description automatically generated

**Figure 11. Heatmap showing the variable values for each cluster. Variables were categorised as “v.low” if the mean (across provinces within that cluster) was below the 25% quantile for that variable across the whole country, “low” if the mean was above 25 and below 50%, “high” if the mean was above 50% but below 75%, and “v.high” if the mean was above the 75% quantile. Pax\_migt\_out = numbers of out-migrants, Pax\_migt\_in = numbers of in-migrants, land\_confl = number of land conflicts, crim\_case = criminal cases per capita, KM\_Comm = distance to commune office, garbage = proportion of families with access to waste collection, dist\_school = distance to nearest school, pig\_fam = proportion of families who keep pigs, Les1\_R\_Land = proportion of families with no rice land, propSecSec = proportion of adults employed in the secondary sector, propPrimSec = proportion of adults employed in the primary sector, M6\_24\_sch = proportion of males aged 6-24 in education, prop\_ind = proportion of the population that is indigenous, pop\_den = population density.**

Diagram

Description automatically generated

**Figure 12. Boxplots showing the distribution of environmental variables for each cluster: *a* = mean forest area, *b* = mean area (km2), *c* = change in forest cover (between 2007-2012), *d* = mean elevation (masl), *e* = mean distance to international border, *f* = mean distance to a provincial capital. Boxplots show the median (centre line within boxes), 25 and 75% percentiles (box edges), and minimum and maximum values (upper and lower whiskers, not exceeding 1.5 × interquartile range). 5 UPGMA clusters.**

**Table 5. Descriptive typology of the provinces and clusters within Cambodia, clustered using socioeconomic variables and the unweighted pair group using arithmetic mean (UPGMA)**

|  |  |  |
| --- | --- | --- |
| **UPGMA cluster** | **Provinces** | **Description** |
| 1 | Mondulkiri, Ratanikiri | Very large provinces with very high elevations. Very low population density, and very high proportion of indigenous people. Very low education levels, very high proportion of primary sector workers and very low proportion of secondary sector workers. Economic security provided by rural livelihoods - few people have no farmland and livestock ownership is common. Very low access to services, high crime per capita, low land conflict, and very low migration levels. |
| 2 | Pailin | Very small provinces with very high elevations. Low population density and low proportion of indigenous people. Low levels of education, low proportion of people in the primary sector but higher proportion of people in the secondary sector. Very few people with no farmland, but very little livestock ownership. High access to services and high crime per capita. Low land conflict and low migration. |
| 3 | Kampong Cham, Kandal, Prey Veng, Takeo | Small provinces with very low elevations. Very high population density and high proportion of indigenous people. Very high levels of education, high proportion of people in the primary sector, but very high proportion of people in the secondary sector. High proportion of people with no farmland, but high levels of livestock ownership. High access to services and low crime per capita. But very high migration levels and very high rates of land conflict. |
| 4 | Banteay Meanchey, Battambang | Large provinces with low elevations. Very high population density and very low proportion of indigenous people. Very high levels of education, and relatively low proportion of workers in the primary and secondary sectors (suggesting higher proportions in the other sectors e.g., tertiary). High proportion of people with no farmland, and low levels of livestock ownership (suggesting very urban). Low access to services, but this may be explained by the mean size of the provinces in this cluster (there is high access to garbage collection). Low crime per capita, but very high migration and very high rates of land conflict |
| 5 | Kampong Chhnang, Kampong Speu, Kampong Thom, Kampot, Kep, Koh Kong, Kracheh, Otdar Meanchey, Preah Sihanouk, Preah Vihear, Pursat, Siem Reap, Stung Treng, Svay Rieng | Large provinces with high elevations. High population density and very high proportion of indigenous people. High levels of education, and a high proportion of workers in both primary and secondary sectors. Very high proportion of people with no farmland, but also very high proportion of people with livestock. Low access to services (although very high access to garbage collection) - this may be an artefact of the very large mean area of the provinces in this cluster. Very high crime rates, very high migration, and very high rates of land conflict. |