**Methods**

***Study area***

This study investigated macroeconomic and socioeconomic predictors of forest cover and loss in Cambodia at two different spatial resolutions over two time periods. The macroeconomic analysis was at the national scale between 1993 – 2015, and the socioeconomic analysis was at the scale of a) the commune, and b) the province between 2007 – 2012. Cambodia is in mainland SEA and is bordered by Laos (NE), Thailand (NW), Vietnam (E), and the Gulf of Thailand (SW) (Figure x). The country has a surface area of 176,520 km2 (UNCTAD, 2020) and is located at latitudes 10-14° north of the equator and thus has a tropical monsoon climate (McSweeney et al. 2010).

***Variable selection***

The response variables for the macroeconomic analysis were 1) change in forest cover (forest loss) from time *t* to time *t+1* and 2) the number of new economic land concession (ELC) allocations in year *t*. The response variable for the socioeconomic analysis was forest cover area. Both forest cover response variables were produced from the same data source (see “Data sources” below). Macroeconomic, socioeconomic, and control variables for both sets of analysis were selected based on a combination of previous studies, data availability, and the authors’ knowledge of Cambodia.

Macroeconomic variables were selected to create three sets of predictors, each targeting a different driver: economic development (n=8), commodity prices (external market forces, n=8), and producer prices (internal market forces, n=5) (Table Sx, Nelson et al. 2006; Ewers 2006; Gong et al. 2013; Kuang et al. 2016; Fan & Ding 2016; Bonilla-Bedoya et al. 2018). Each predictor was hypothesised to be a driver of forest loss (Table Sx). Human population density was included as a control variable for the economic set and total forest remaining was included as control variable across all sets, as both were expected to influence forest loss. Both per capita Gross Domestic Product (GDP) and amount of forest remaining were included to reflect the economic development path and the forest scarcity path respectively (Rudel et al. 2005; Lambin & Meyfroidt 2010). After pre-analysis checks for errors and correlation, the resulting variable set contained 20 variables (Table 1).

Socioeconomic variables were selected to create 8 variable sets reflecting different aspects of socioeconomic status and development, each of which was hypothesised to be either a driver or predictor of forest cover (Table Sx, Luck et al. 2009; Ty et al. 2012; Kristensen et al. 2016; Bonilla-Bedoya et al. 2018). The variable sets were population demographics (n=8), education (n=4), employment (n=5), economic security (n=2), access to services (n=4), social justice (n=2), migration (n=2), and control variables (n=6). Control variables were included to account for the effects of environmental and other human factors including economic land concessions (Abdullah & Nakagoshi 2007; Davis et al. 2015), protected areas (Bonilla-Bedoya et al. 2018), elevation (Ty et al. 2012), and distance to human infrastructure (Ty et al. 2012). A habitat control variable was excluded because the response variable (forest cover) was extracted from a land cover layer and represented a specific type of habitat, resulting in non-independence between the response and habitat.

***Data sources***

National macroeconomic variables were acquired from publicly available sources (Table 1) for the period 1993 – 2015. Fine-scale socioeconomic variables were extracted from the Cambodian Commune Database for the years 2007 – 2012 (Table 2) which are available from Open Development Cambodia ([www.opendevelopmentcambodia.net](http://www.opendevelopmentcambodia.net)). Data on economic land concessions, protected areas, and elevation (digital elevation model), and shapefiles for the country, provinces, and communes were provided by the Wildlife Conservation Society. Forest cover layers were taken from the publicly available European Space Agency Climate Change Initiative (ESACCI) satellite data for the years 1993 – 2015.

Table 1. Variables selected for the macroeconomic analysis. Variables range from 1993 – 2015.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictor variable** | **Units** | **Resolution** | **Source** | **Details** |
| *Economy* |  |  |  |  |
| GDP per capita | Billions USD | National | World Bank | Constant 2010 rates |
| GDP growth | % | National | World Bank | Annual percentage growth rate of GDP at market prices based on constant local currency |
| Foreign Direct Investment | Millions USD | National | UNCTAD | Inward and outward flows and stock |
| Agricultural sector proportion of GDP | % | National | CNIS | Proportion of national GDP |
| Development flows to agriculture | Millions USD | National | FAO | Donor development investment flows, other official flows, and private donor flows at constant 2016 prices to all agriculture and forestry sub-sectors |
| Development flows to environment | Millions USD | National | FAO | Donor development investment flows, other official flows, and private donor flows at constant 2016 prices to general environment protection |
| *Commodity prices* |  |  |  |  |
| Crop Production | Index | National | FAO | Relative level of the aggregate volume of agricultural production for each year in comparison with the base period 2004-2006 |
| Non-food agricultural production | Index | National | FAO | Relative level of the aggregate volume of non-food agricultural production for each year in comparison with the base period 2004-2006 |
| Forestry production | m3 | National | FAO | Total production values for industrial roundwood, non-coniferous tropical wood, other industrial roundwood, sawlogs and veneer logs (coniferous and non-coniferous), and sawnwood (coniferous and non-coniferous |
| Price of rice | USD/ton | Global | World Bank | Median annual global market price of rice |
| Price of corn | USD/ton | Global | World Bank | Annual global market price of corn |
| Price of rubber | USD/ton | Regional | RASCE | Monthly regional market value of rubber on the Singapore Exchange |
| Price of sugar | USD/ton | Global | World Bank | Annual global market price of sugar |
| *Producer prices* |  |  |  |  |
| Producer price of Rice | USD/ton | National | FAO | Farmgate prices for Cambodian producers |
| Producer price of rubber | USD/ton | National | FAO | Farmgate prices for Cambodian producers |
| Producer price of cassava | USD/ton | National | FAO | Farmgate prices for Cambodian producers |
| Producer price of corn | USD/ton | National | FAO | Farmgate prices for Cambodian producers |
| Producer price of sugar | USD/ton | National | FAO | Farmgate prices for Cambodian producers |
| *Control* |  |  |  |  |
| Population density | pax/km2 | National | FAO | People per km2 |
| Forest remaining | Km2 | National | ESACCI | Raw value of forest remaining |

Table 2. Variables selected for the socioeconomic analysis. Variables range from 2007 – 2012.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Set** | **Variable** | **Transformation for analysis** | **Province-level class** | **Details** |
| Demographics | Total population |  | NA | Includes women, men, and children of all ages |
|  | Population density |  |  |  |
|  | Number indigenous | Proportion of total population | Categorical | Total number of people who are indigenous/ethnic minority |
| Education | Males aged 6 – 24 in school | Proportion of total number of males aged 6 - 24 |  | Number of males aged 6 - 24 in full time education |
| Employment | Number of adults employed in primary sector | Proportion of total adult population | Categorical | The primary sector includes agriculture (rice and other crop farming), fishing, livestock farming, forestry, and non-timber forest product collection (Kenessey 1987) |
|  | Number of adults employed in secondary sector | Proportion of total adult population | Categorical | The secondary sector includes wood-based production (e.g. furniture), metal- and glass-based production, foodstuff production, plastic- and rubber-based production, textiles production (Kenessey 1987) |
| Economic security | Number of families with <1ha rice land (including no rice land) | Proportion of total number of families | Categorical |  |
|  | Number of families who keep pigs | Proportion of total number of families | Categorical |  |
| Access to services | Distance to nearest school |  | Categorical | Median distance from any village in the commune to the nearest school (primary or secondary) |
|  | Number of families with access to waste collection | Proportion of total number of families |  |  |
|  | Distance to the Commune Office |  |  | Median distance from any village in the commune to the Commune Office (government administration office) |
| Social justice | Number of criminal cases | Criminal cases per capita | Categorical | Includes murder, theft, and other criminal cases |
|  |  |  |  |  |
|  | Number of land conflict cases |  | Categorical | In the previous 12 months |
| Migration | Number of in-migrants |  | Categorical | Migration into the commune |
|  | Number of out-migrants |  | Categorical | Migration out of the commune |
| Control | Mean elevation (masl) |  | Categorical | Mean elevation for the commune |
|  | Distance to international border (km) |  | Categorical | Distance from the centre of the commune to the nearest international border |
|  | Distance to Provincial Capital (km) |  | Categorical | Distance from the centre of the commune to the centre of the provincial capital (town or city) |
|  | Presence of economic land concessions |  |  | Binary. 1 = part or all of an economic land concession falls within the boundary of the commune, 0 = no economic land concession falls within the commune boundary |
|  | Presence of protected area |  |  | Binary. 1 = part or all of an protected area falls within the boundary of the commune, 0 = no protected area falls within the commune boundary. "Protected area" includes Wildlife Sanctuary, National Park, Protected Landscapes, Multiple-use areas, RAMSAR sites |
|  | Protected area category |  |  | None = no protected area falls within commune, MULTI = more than one category of protected area falls within commune, WS = wildlife sanctuary, NP = national park, PL = protected landscape, MUA = multiple-use area, RMS = RAMSAR |

***Data processing***

The forest cover variable (response) for both analyses was extracted from the ESACCI product by totalling the number of pixels (1 Pixel = 0.09km2) in each year classified as bands 50, 60, 61, 62, 70, 71, 72, 80, 81, 82, 90, and 100 (Table Sx). For the macroeconomic analysis, the total forest cover for the whole country was used, and for the socioeconomic analysis the forest cover layer was further stratified into forest cover per commune and forest cover per province. Forest cover data processing was done in QGIS (QGIS Geographic Information System v3.16). For both analyses, predictor variables were checked for collinearity, and if two variables in the same set had a correlation coefficient of >0.6 then generally one was removed (Supporting Information).

*Macroeconomic analysis*

Forest cover was converted to change in forest cover using *forest covert+1 − forest covert*, where *t* represents year *t.* There were no periods of forest gain during the study period, and so the response can be considered as rate of forest loss. All predictors were converted from raw values to change in values using *Xt+1 – Xt,* where *t* represents year *t* (Barrett et al. 2006). The variable *forest remaining* was left as raw values (km2). Cambodia’s first general election and subsequent adoption of a free market economy occurred in 1993, resulting in unreliable GDP-related values for 1993 (Chhair & Ung 2013) and subsequent change values in 1994, and so these were removed. To simplify interpretation, predictor variables were not centred or scaled prior to change calculations or modelling.

*Socioeconomic analysis*

Data from the Commune Database were at the resolution of individual village, and so the selected variables (Table 2) were aggregated to the commune and province level after error checking and cleaning (Supporting Information). Some variables were converted from raw values to proportional data to account for large differences in commune and province size and human population (Table 2). Data were checked for errors in R (Supporting Information, R Core Team, version 4.0).

***Modelling***

*Macroeconomic models*

This analysis aimed to model the relationships between changes in macroeconomic predictors and 1) the change in forest cover at a national level and 2) the allocation of new ELCs, for the time period 1993 – 2015. Models were run for both response variables with each of the three variable sets: economic development, commodity prices, and producer prices. To account for the effect of time, a linear model of the response as a function of time (year) was run and the model residuals were extracted and used as a control predictor in all subsequent models. The amount of forest remaining (km2) was also included as a control variable in all models. Modelling was done using Generalised linear models (GLM) and followed an information theoretic approach (Burnham & Anderson 2007). For the models with rate of forest loss as the response both gaussian and gamma distributions were tested, and for the models with ELC allocation a Poisson distribution was used. Resulting models were compared using Akaike’s Information Criterion (AIC). Final rate of forest loss models used gaussian distributions. All predictors in each model set had been selected because of a priori hypotheses (Table Sx), and so within each set all combinations of possible models were run and compared using AIC. Models with ∆AIC < 6 were considered to have sufficient support and retained in the final model set. Model averaging was implemented for the final model set, resulting in model-averaged coefficients for all model terms (Burnham & Anderson 2007). Models were run and averaged using the MuMIn package in R (Version 1.43.17, Bartoń 2020). This modelling procedure was repeated for a one-year time lag and two-year time lag as follows:

No time lag:

Where is the response at time , and is predictor variable at time.

One year time lag:

Where is the response at time , and is predictor variable at time.

Two year time lag:

Where is the response at time , and is predictor variable at time.

*Socioeconomic models*

This analysis aimed to model the relationships between forest cover and socioeconomic variables within communes between 2007 – 2012. The results of initial commune-level modelling prompted further aggregation of the data to the province-level and models were built to investigate the relationships between forest cover and socioeconomic variables within provinces for the same time period.

*Commune-level models*

Generalised linear mixed models (GLMM) with Poisson errors were used with commune nested within province as random intercept terms to account for repeat measurements and the hierarchical data structure, and year as a random slope term to account for temporal autocorrelation (Zuur et al. 2009). The natural logarithm of commune area (km2) was used as an offset term in all models to account for large variation in commune size. Due to the large number of available predictor variables, maximal within-set models were run first for each of the 8 variable sets (Table Sx), and variables with very weak, or no effect were dropped. Simplified models were compared with maximal models using likelihood ratio tests and analysis of variance tests. If a variable set had only one variable, this was automatically taken forward. Because assessment of term significance in GLMMs is complex, predictions and plots were made for all terms before being dropped to ensure noteworthy effects were not being missed. This process resulted in a final set of 13 variables which were used to create a candidate set of 10 models (Table Sx). Following an information theoretic approach (Burnham & Anderson 2007) models were compared via AIC to select the top model or models. The resulting final model fit was assessed via diagnostic plots (residuals versus fitted, quantile-quantile of random effects, Supporting Information, Harrison et al (2018)). Marginal (fixed effects only) and conditional (fixed and random effects) pseudo-R2 values were calculated based on Nakagawa & Schielzeth (2017) using the R package ‘MuMIn’ (Bartoń 2020). To investigate the variation in effects between provinces, predictions were made for each variable within each commune and the 50% quantile from all commune-level predictions within each province was extracted as the provincial mean prediction.

*Province-level models*

The same GLMM model formulation was used for the province-level models except that commune was removed from the random effects structure. Based on provincial-level histograms of predictor variables, 14 predictors were converted to categorical variables by splitting the data by the mean, resulting in “high” and “low” values (Table 2). Following an information theoretic approach, a candidate set of models was created (Table Sx) and model comparison was done using AIC.

***Cluster analysis***

Agglomerative clustering was conducted to create a typology for provinces based on the socioeconomic variables used in the analysis above. Several agglomerative clustering approaches were assessed. These were single linkage, complete linkage, unweighted pair-group using arithmetic averages (UPGMA), unweighted pair-group using centroids (UPGMC), Ward’s minimum variance, and flexible clustering. The methods were compared using cophenetic correlation and Gower distance metrics, and the appropriate number of clusters (k) was selected using the matrix correlation statistics (Borcard et al. 2018). The capital city of Phnom Penh, which is technically a province in itself, was removed prior to clustering because it has extreme values for many of the variables and is thus an outlier that affects the clustering.