**Introduction**

Southeast Asia (SEA) is characterised by complex biogeography and extensive tropical forest cover resulting in exceptional biological diversity, yet it has one of the highest rates of deforestation in the world (Hughes, 2017). Deforestation rates in SEA are comparable only with those of Latin America (Estoque et al., 2019), and the resulting habitat loss is arguably more damaging to biodiversity (Sodhi et al., 2010, 2004). The loss of SEA’s forests has potentially severe consequences for climate change (Ceddia et al., 2015), ecosystem-based adaptation (Estoque et al., 2019), local people (Culas, 2007; Frewer and Chan, 2014; Gaughan et al., 2009; Poffenberger, 2006), and biodiversity (Chapman et al., 2018; Hearn et al., 2018). The drivers of tropical deforestation vary both by location and by scale, ranging from broader drivers such as population pressure and weak institutions (Geist and Lambin, 2002), to proximate causes at a local level such as the expansion of cash crops, agriculture, and other food production (Estoque et al., 2019; Imai et al., 2018; Stibig et al., 2014; Wilcove et al., 2013; Zeng et al., 2018), the associated expansion of roads and infrastructure that facilitate such expansion (Hughes, 2018), and civil unrest and war (Kaimowitz and Fauné, 2003; Price, 2020). Globally, land conversion for commodity production is the single largest driver of deforestation (Curtis et al., 2018).

Deforestation is accentuated in low income countries where poverty, debt, and overpopulation are high, and thus so are the demands for economic growth via agricultural commodities and other natural resource-based products (Culas, 2012). Research has shown that this trend is reversed in high income countries where economic activity shifts to become more service-driven, and demand for environmental services increases, resulting in afforestation (Culas, 2012). The change from deforestation (low income) to afforestation (high income) is termed forest transition (Mather, 1992). There is also evidence that remaining forest area is an effective predictor of deforestation rates, with countries undergoing forest transition when remaining forest cover across the country is low (Lambin and Meyfroidt, 2010). Within SEA, countries such as Vietnam have undergone forest transitions and reduced deforestation rates dramatically over the last 20 years (Meyfroidt and Lambin, 2008). Yet often this progress is at the expense of poorer, less developed countries with weaker institutions and policy frameworks such as Cambodia, where activities including logging and forest clearance leak across borders (Meyfroidt and Lambin, 2009).

There are various broad pathways which can result in a land use transition within a country (Lambin and Meyfroidt, 2010). Despite the terminology, these pathways are not linear or deterministic; they are driven by complex endogenous and exogenous drivers and feedback loops between economic, political, social, and environmental factors operating at a variety of scales (Lambin and Meyfroidt, 2010). To facilitate the development of sustainable land use policies that can guide countries towards more desirable forest transition pathways, and to support targeted interventions that are effective at reducing forest loss, it is critical to identify and understand macroeconomic and socioeconomic drivers and predictors of forest cover and loss at different scales. Describing relationships between forest cover and macroeconomic and socioeconomic variables, rather than simply biophysical variables, is particularly important in developing countries that are undergoing social transition, as these are the conditions under which deforestation is often accelerated (Imai et al., 2018). Successfully isolating the signals of these relationships is however, challenging, due to the complexity of social-ecological systems, the non-linear feedback loops, and the heterogeneity in system dynamics at different scales.

***Modelling approaches***

Modelling the processes that drive land use change (LUC), including forest loss, is challenging due to the complexity of the systems within which it occurs (Basse et al., 2014). Advanced modelling frameworks have developed which are allowing greater understanding of the processes underlying LUC, and subsequently more accurate predictions into the future (Basse et al., 2014; Bonilla-Bedoya et al., 2018). However, the modelling approach taken depends on several factors including 1) the research question, e.g., whether you want to model *patterns* of LUC or the *processes* of LUC, 2) the scale of the analysis (Brown et al., 2013), e.g., understanding human decision-making that drives land conversion at the local level, or understanding how global macroeconomics affect agricultural expansion at a regional or national level, 3) expertise and interests, which govern both the research questions and the statistical approaches that are likely to be used, and 4) data availability, which although has seen dramatic improvements in recent decades thanks to freely available, high resolution remote sensing imagery and long-term global data sets on socioeconomics, can still limit analyses, particularly when there are scale mismatches between data (Ceddia, 2019).

Methodological approaches fall broadly into two groups governed largely by the aims of the study. First, modelling the spatial processes of LUC is a common goal, as this allows researchers to use patterns of past LUC to predict which areas are at higher risk of land conversion in the future, with the potential to explore a number of plausible future scenarios (Basse et al., 2014). There are several spatially explicit cell-based modelling frameworks that can achieve these aims, including maximum entropy (Bonilla-Bedoya et al., 2018; de Souza and De Marco, 2014), and cellular automata (Stevens and Dragićević, 2007; Yang et al., 2012), which rely on discrete spatial units that have associated variable values and tend to be spatially correlated. Models are trained on past LUC, thus developing rules which govern the likelihood of a given cell being converted under various future conditions. These rule-based approaches can be improved with advanced machine learning techniques such as artificial neural networks (Basse et al., 2014). More traditional statistical models have also been used in a cell-based framework, whereby models such as logistic regression are used to predict the likelihood of a cell being converted (Aguiar et al., 2016; Aspinall, 2004). The above methods require spatially explicit data at the scale of the cell, and so are unable to model broader relationships and drivers that operate at a different level within the system. These approaches rely on geographical and biophysical predictor variables that have fixed values in space, precluding the investigation of national or regional drivers. Socioeconomic drivers can be modelled, provided they can be represented in discrete space and are at the correct scale (de Souza and De Marco, 2014; Estoque et al., 2019).

Second, researchers may want to model the relationships between LUC and trends in predictor variables over time rather than space. These approaches are generally less deterministic than the spatial process modelling above, and are often at much larger scales (e.g., Bhattarai and Hammig, 2004; Ceddia, 2019; Ewers, 2006). These analyses are often targeting broader economic, socioeconomic, cultural, political, and institutional drivers of LUC, which are less amenable to spatial sampling. Generalised linear mixed models (GLMMs, also known as multilevel or hierarchical models) are often employed in such analyses, as GLMMs can account for temporal autocorrelation and hierarchical data structures (Zuur et al., 2009). Studies have used these, and other regression-type models, to investigate the relationships between LUC and national income and forest policies (Bhattarai and Hammig, 2004), income, land, and wealth inequalities (Ceddia, 2019), indigenous land tenure (Ceddia et al., 2015), macroeconomics and economic development (Culas, 2007; Ewers, 2006), and urban socioeconomics (Gong et al., 2013). Studies that use GLMMs almost exclusively use data from multiple countries, taking advantage of the ability of these models to harness large longitudinal data sets with few “subjects” without succumbing to pseudoreplication (Gelman and Hill, 2006, see Ewers, 2006 for an example of where pseudoreplication may be an issue). Another advantage of GLMMs is the ability to quantify between-group variance, which not only offers crucial insight about the differences between groups (e.g., countries) from which inference can be drawn (Zuur et al., 2009), but can also highlight potential problems with ‘global’ predictions (i.e., predictions that are made with all random effect terms set at their mean). Yet very few studies that use these models for LUC report any values for variance associated with the random (group-level) effects. For example, Bhattari and Hammig (2004) use data from 63 countries to produce a single effect for GDP per capita on deforestation, yet do not report any value for country-level variance. The effect size is relatively small, and therefore if there was large between-country variance then the country-level effects could be vastly different, rendering the single global effect misleading. Furthermore, some studies fail to even specify the statistical or modelling approaches used to arrive at their conclusions (e.g., Gao and Liu, 2012).

Modelling complex, non-linear systems that interact and feedback at various temporal and spatial scales is difficult (Dawson et al., 2010), and conservation challenges can arise when there are mismatches in the interactions and scales between social and ecological systems (Beever et al., 2019). It is therefore important for researchers to both accept and report high levels of uncertainty and variation within models that attempt to simplify complex systems. For example, without appropriate estimates of variance at the different levels within GLMMs, it is difficult to reliably assess model coefficients. This is particularly true when hierarchical model levels represent different spatial scales (e.g., countries within regions, cities within states).

***Cambodia***

Between 1975 and 1992 Cambodia suffered enormous civil unrest, war, and foreign occupation, which resulted in almost complete economic collapse. Yet by 2006 Cambodia’s economy was one of the fastest growing economies in the world and represents a good example of a country’s ability to move from post-conflict status to full integration within a dynamic regional economy (Hughes and Un, 2011). This remarkable economic recovery has, however, come at a cost for the country’s natural environment (Davis et al., 2015).

In the decades following the 1991 Paris Peace accords, which officially ended civil war and foreign occupation in Cambodia, political stability and rapid economic growth resulted in an economic transformation (Hughes and Un, 2011). Between 1990 and 2006 the economy transitioned from 85% subsistence agriculture, a small garments sector, and generally low economic productivity, to 55% of the population involved in subsistence agriculture and a private sector that was growing exponentially, thanks to investments in tourism, manufacturing, and mining (Eliste and Zorya, 2015; Hughes and Un, 2011). Between 2000 and 2006 the economy grew by an average of 8.7% - one of the highest rates in the world - driven primarily by manufacturing, (especially garment manufacturing), construction, services, and tourism. These industries were geographically limited to the two major cities: Phnom Penh, Siem Reap, and their surrounds. Outside of the large cities, the rural economy was also undergoing significant change. The adoption of the 2001 Land and Forestry Laws saw a major drive towards industrial-scale agriculture in the form of private and state-owned land concessions targeting commercial crops such as rubber and sugar (Eliste and Zorya, 2015). These concessions drove most of the growth within the agricultural sector, and despite having positive economic impacts, have received heavy criticism for violations of local land rights and illegal deforestation within protected areas (Global Witness, 2013; Vrieze and Kuch, 2012; Watson et al., 2014).

The rapid economic development around the large urban centres has increased the gap in development and socioeconomic status between the urban and rural populations. Despite the national economic success since the 1990s, Cambodia is still one of the least developed countries in the world, with more than 33% of the population either in severe poverty or vulnerable to severe poverty (UNDP, 2020). In one study of a rural province in Cambodia, it was found that 70% of the population were either low-skilled non-permanent wage employees, permanent farmers, or were involved in resources extraction (Nguyen et al., 2015). The authors further found that all households, even those with high skilled or permanent wage employees, engaged in personal agriculture or livestock rearing, and were in some part dependent on environmental resources (Nguyen et al., 2015). Rural poverty and reliance on natural resources have been exacerbated by decades of insecure land tenure, followed by several ambitious, yet poorly governed, land tenure policies such as Directive 01 (Grimsditch and Schoenberger, 2015; Milne, 2013). These land tenure policies resulted in widespread land speculation and rapid deforestation across the country, including within protected areas (see Thesis appendix).

At the start of the century 41.9% of Cambodia’s land area was forested, and by 2012 the total forested area had been reduced by 19.8%, equating to over 1.3 million hectares (Davis et al 2015). Only 25 other countries lost more forest than Cambodia between 2000 – 2012 (Hansen et al 2013). Such high rates of deforestation, in the context of Cambodia’s economic status and the rural population’s reliance on natural resources (Nguyen et al., 2015), suggests that the country is on the increasing trajectory of the Environmental Kuznet’s Curve, whereby national economic development is improved at the expense of forest cover (Stern, 2004). This theory is supported by evidence of increased forest loss within agro-industrial land concessions, highlighting these economic concessions as key drivers of deforestation. These Economic Land Concessions (ELCs) have been an important, and controversial, tool in the economic recovery of Cambodia. Nevertheless, the importance of natural resources for the country’s large rural population, and in particular the reliance on forest resources by the country’s indigenous populations (Nguyen et al., 2015), means that reducing forest loss and directing the country towards a forest transition is important (Rudel et al., 2005). Therefore, understanding the relationships between macroeconomics, socioeconomics, and forest cover at different scales across the country will be critical to inform sustainable economic policies and effective forest conservation programmes.

***Purpose of the study***

In this study we investigate the relationships between forest cover and loss, and macroeconomic and socioeconomic variables in Cambodia. Specifically, we: 1) model the relationships between three sets of macroeconomic predictors and forest loss at the national scale, 2) model the relationships between three sets of macroeconomic predictors and the creation of new ELCs at the national scale, 3) model the relationships between forest cover and eight sets of socioeconomic predictors, at two different sub-national scales, and 4) conduct a cluster analysis to group provinces by socioeconomics in order to further understand the socioeconomic typologies within the county, and their relationships to forest cover.