Understanding the drivers and proximate causes of deforestation is critical for the development of sustainable environmental policies and forest conservation initiatives. In this study, I have modelled the relationships between variables describing socioeconomic development and forest cover at multiple scales and have investigated these relationships using two different approaches. This study has revealed some important relationships from which we can make inferences regarding the socioeconomic, geographical, and biophysical predictors of forest cover across Cambodia. Furthermore, I have revealed key methodological issues, particularly around scale and model variance, that are likely to be common in analyses such as this, which focus on large spatial scales and fine resolutions, but which often remain unexplored or unreported in the literature. Studies investigating the socioeconomic drivers of deforestation need to target multiple scales to build a cohesive picture of the social-ecological systems within which deforestation occurs, so that policy development targets the appropriate drivers at each scale (e.g., at different administrative levels).

**The effect of scale on predicting forest cover**

This analysis highlights the importance of scale when modelling complex social-ecological systems; researchers must not only select the scale of the analysis carefully but must also be aware of underlying variation which may be affecting estimates, requiring cautious interpretation of results. The results from the GLMMs have highlighted the effect of scale on predictors of forest cover. The direction of the effect of distance to an international border changes depending on whether you are looking at the commune-level or the province-level; it was positive within communes and negative within provinces. This reversal of effect direction between scales also occurs for the distance to the provincial capital (positive within communes and negative within provinces). Taken together, the two models can add important nuance to the interpretation of results. Provinces that are close to international borders have higher forest cover, but within those provinces, the communes that are furthest away from the border and the provincial capital are predicted to have the highest forest cover. Provinces furthest away from the major the urban centres of Phnom Penh, Siem Reap, and Battambang tend to be the large, rural provinces that have an international border (e.g., Mondul Kiri, Ratanak Kiri, Stung Treng, Koh Kong) and have high forest cover. Increases in human population density over time, including from in-migration, often result in agricultural expansion, exploitation of forest resources (e.g., timber), and increased urbanisation, all of which could be reducing forest cover around the provincial capitals. International borders promote the movement of people, commodities, economic activity, and all the associated infrastructure that is required to maintain such activity. When combined with illegal cross-border activities such as logging, land clearance, and the wildlife trade (see Evans et al. 2013), it is plausible that communes closer to the international borders are more likely to have reduced forest cover (Grogan et al., 2015).

**Socioeconomic typography of provinces in Cambodia**

The cluster analysis revealed an interesting pattern of distinct socioeconomic regions across Cambodia, suggesting that in many cases provinces that are adjacent to each other tend to have similar socioeconomic characteristics, resulting in clusters that are comprised of spatially contiguous provinces. The two cluster that generally display the largest differences are clusters 1 and 3. Cluster 1 contains the provinces of Mondul Kiri and Rattanak Kiri which are large, remote, and some of the least developed provinces in the country. They are home to the Eastern Plains Landscape which is one of the most important areas in SEA for biodiversity (chapter 4; Gray et al., 2012; Nuttall et al., 2017, 2021). Mondul Kiri and Rattanak Kiri (cluster 1) have the highest forest cover, low population density, low access to services, and low migration. Economic development in the first two decades after the civil war was focused almost entirely on the major cities: Phnom Penh (cluster 3), Sihanoukville (cluster 5), and Battambang (cluster 4), with rural provinces remaining underdeveloped, inaccessible, and poor (Hughes and Un, 2011). The lack of infrastructure and access, coupled with low population density and few employment opportunities that limited in-migration, have all likely contributed to forest cover remaining high (Evans et al., 2013).

Conversely, Kampong Cham, Kandal, Prey Veng, Takeo (cluster 3) have the lowest levels of forest cover and the cluster contains the capital city of Phnom Penh and the surrounding provinces which are the hubs for industry and economic activity (such as the garment sector). Cluster 5, which contains the provinces of Kampong Chhnang, Kampong Speu, Kampong Thom, Kampot, Kep, Koh Kong, Kracheh, Otdar Meanchey, Preah Sihanouk, Preah Vihear, Pursat, Siem Reap, Stung Treng, Svay Rieng, is interesting because it contains the largest number of provinces. The expectation was that the provinces that most closely resembled cluster 1 (i.e., large, rural provinces with high forest cover) such as Stung Treng, Preah Vihear, and Koh Kong, would have been clustered either with cluster 1, or within a separate cluster. However, they were clustered with the central belt of provinces (e.g., Kampong Speu, Kampong Chhnang, Kampong Thom) which are almost exclusively low elevation agricultural provinces that are geared towards rice production. The inclusion of Stung Treng, Preah Vihear, and Koh Kong within this cluster and the resulting cluster typologies, suggest that there has been some success in increasing the socioeconomic status of rural, highly forested provinces without excessive loss of forest cover.

**Methodological approach**

*Mixed models*

The commune-level model revealed that population density was the only non-control variable with any effect on forest cover, and the effect was weak. I was limited in the socioeconomic variables that were available, and it is possible that the variables selected were simply poor predictors of forest cover. However, the modelling process revealed very large between-commune variation in both predictor and response variables, in addition to many random effect levels (between 1,317 and 1,512). Model predictions from the final model, and from preliminary models, showed that the parameter estimates (intercepts and slopes) for a given socioeconomic variable (see Figure 3 for an example from population density) varied widely from commune to commune, even within the same province. Therefore, it is possible that the difficulty in estimating a single parameter from the surrounding “noise” resulted in the detection of weak, or no effects, rather than a genuine lack of effects. The province-level model was built to counter the issue of excessive between-commune variance by approaching the analysis from a different scale. Two socioeconomic variables remained in the final province-level model but again, the effects were relatively weak. It is still possible that the weak effects represented a genuine lack of correlation between socioeconomics and forest cover, however, modelling the effects at a larger scale will simply mask the large variation that exists at the finer scale, rather than eliminating it. An 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.

The inherent complexity within social-ecological systems results in significant challenges when researchers attempt to model them (Basse et al., 2014). Taking this study as an example, a researcher has a choice between modelling at a large scale (e.g., national, regional) where effects may be weak or unrepresentative of much of the country or region, or modelling at a fine scale where effects may be swamped by variation resulting in the loss of the true signal. By reframing analytical goals and aiming for description of the data, for example using cluster analyses, over statistical hypothesis testing and attempts at explanation, researchers can reduce the need for increasingly complex data and models.

*Cluster analysis*

The purpose of the cluster analysis was to explore an approach that was different to the traditional statistical modelling I had done using GLMMs, and to remove the above issues of variance. I was interested to see what patterns would emerge when the underlying goal of statistical hypothesis testing (i.e., the effect of the predictor *x* on response *y* is significantly different from 0) was removed. The cluster analysis revealed patterns beyond those produced using the GLMMs and was therefore a worthwhile addition to the study. The advantage of clustering techniques such as UPGMA is that although there are metrics that can suggest optimal numbers of clusters, the researcher can select the number of clusters that is most useful for their particular investigation (Borcard et al., 2018). Unlike statistical models, cluster analysis does not produce estimates of effect sizes, nor can predictions be made. Nevertheless, by altering the number of clusters, investigating different clustering approaches, followed by considered exploratory analysis and plotting, a comprehensive picture of the study system can be produced. This may be a sensible first step in a larger analysis which can increase understanding of the system before modelling approaches are decided upon. Furthermore, methods such as cluster analysis are in some cases, conceptually simpler than advanced statistical and mechanistic modelling, making interpretation and explanation simpler.

**Policy implications**

The results of this study have highlighted that the regions of Cambodia that have the highest forest cover also tend to be the rural, remote, poor provinces with high proportions of indigenous people. It is people living within these areas that will be reliant on natural resources and forest products for their subsistence and livelihoods. In these circumstances, the efforts of an individual actor to increase their socioeconomic status is likely to include agricultural expansion, resulting in forest loss. Therefore, to avoid forests being the price of socioeconomic development, national and sub-national government need to develop economic policy frameworks that deliver economic benefits whilst encouraging forest protection, such as payments for ecosystem services schemes, and support for agricultural improvement technologies and diversification (Eliste and Zorya, 2015). The Cambodian government, however, has shown enthusiasm in the past for economic development via private land leases for industrial-scale commercial agriculture (chapter 2), many of which have been awarded in rural, remote, forested land. These economic land concessions (ELCs) have frequently been placed on traditional lands of indigenous people, on the lands of the rural poor who had yet to be awarded legal land titles, and in areas of high forest cover, including protected areas (Beauchamp et al., 2018; Davis et al., 2015; Magliocca et al., 2019; Neef et al., 2013; Oldenburg and Neef, 2014; Vrieze and Kuch, 2012). Remote provinces with a low density of relatively poor inhabitants, low levels of land tenure security, and plentiful forests, are particularly vulnerable to the allocation of new ELCs, particularly if the economic policies of the last decade are pursued. Despite legal requirements to the contrary, ELCs often contribute very little to local economies, and are sources of land conflict, illegal settlement, and extensive, unregulated, and often illegal deforestation (Beauchamp et al., 2018; Davis et al., 2015; Global Witness, 2013; Milne and Mahanty, 2015; Neef et al., 2013; Oldenburg and Neef, 2014; Vrieze and Kuch, 2012; Watson et al., 2014). There has been a reduction in new ELC allocations in recent years (www.opendevelopmentcambodia.net), which may suggest that new avenues for economic development and growth in the agriculture sector are being developed.

Since the end of civil conflict in the early 1990s, there has been significant migration and resettlement into rural provinces as people move back into traditional homelands or move in search of new land to settle (Milne and Mahanty, 2015). This post-war migration has come at a cost to forest cover, as families look to establish and expand their agricultural land (Hought et al., 2012; Kong et al., 2019). The rural provinces with high forest cover are still vulnerable to in-migration and land speculation, as access to the provinces has improved significantly over the last decade and poor, landless families seek to establish themselves in frontier areas (Evans et al., 2013). This, and other studies, have demonstrated that increases in human population density can predict forest loss (Dasgupta et al., 2005; Krishnadas et al., 2018). In the context of poor environmental governance and weak institutions, as in Cambodia (Milne and Mahanty, 2015; Riggs et al., 2018), rural in-migration could continue to drive forest loss. Government policies for rural settlement and land titling need to pre-empt increased migration into rural areas, particularly those with protected areas, and ensure forest loss is minimised. The two most prominent settlement initiatives – social land concessions and Directive 01 – have been widely criticised for poor management and implementation, and both have resulted in the loss of forests inside protected areas (Milne 2013; Oldenburg & Neef 2014; Grimsditch & Schoenberger 2015, also see Appendix).

The cluster analysis placed the provinces Preah Vihear, Stung Treng, and Koh Kong into cluster 5, suggesting that these provinces have socioeconomic conditions similar to the wealthier, more developed provinces such as Siem Reap, and to provinces with extensive agriculture such as Kampong Chhnang. This placement was despite many similarities with the provinces in cluster 1 (Mondul Kiri, Rattanak Kiri), including being large, rural, and with high forest cover. This clustering suggests that Preah Vihear, Stung Treng, and Koh Kong have made progress in increasing the socioeconomic conditions of the population, without extensive forest loss (median forest loss within cluster 5 is less than for cluster 1). These improvements may have been driven by the expansion of economic sectors that do not rely on natural resource extraction or agricultural expansion, or indeed by the growth of nature-friendly sectors, such as ecotourism. These results warrant further investigation to identify whether there are specific policies, initiatives, economic conditions, or social movements that have improved socioeconomic conditions with minimal deforestation, and which could be replicated in other poor, forested provinces.