An ensemble of spatially explicit land-cover model projections: prospects and challenges to retrospectively evaluate deforestation policy

Bradley, Andrew V.; Rosa, Isabel M. D.; Brandao, Amintas, Jr.; Crema, Stefano; Dobler, Carlos; Moulds, Simon; Ahmed, Sadia E.; Carneiro, Tiago; Smith, Matthew J.; Ewers, Robert M.

Modeling Earth Systems and Environment

DOI:
10.1007/s40808-017-0376-y

Published: 01/12/2017

Peer reviewed version

Title:

An ensemble of spatially explicit land-cover model projections: prospects and challenges to retrospectively evaluate deforestation policy.

Authors:

Andrew V. Bradley[1], Isabel M. D. Rosa[1,2], Amintas Brandão Jr.[3, 4], Stefano Crema[5], Carlos Dobler[6], Simon Moulds[7,8], Sadia E. Ahmed[9,10], Tiago Carneiro[11,12], Matthew J. Smith[13], Robert M. Ewers[1]

Affiliations and addresses:

[1] Department of Life Sciences, Imperial College of London, Silwood Park, Ascot, SL5 7PY, UK.

[2] Biodiversity Conservation, German Centre for Integrative Biodiversity Research (iDiv), Deutscher Pl. 5E, 04103 Leipzig, Germany; Martin Luther University Halle-Wittenberg, Halle, Germany.

3 Imazon, Travessa Dom Romualdo de Seixas, 1698, Ed Zion Business, Umarizal, Belém, Brazil.

[4] Center for Sustainability and the Global Environment (SAGE), Nelson Institute for Environmental Studies, University of Wisconsin, 1710 University Avenue, Madison, WI 53726, USA

[5] Clark Labs, Clark University, 950 Main Street, Worcester, 01610, MA, USA.

[6] Graduate School of Geography, Clark University, 950 Main Street, Worcester, 01610, MA, USA.

[7] Centre for Water Systems, University of Exeter, North Park Road, Exeter EX4 4QF, UK.

[8] Department of Civil and Environmental Engineering, Imperial College of London, Silwood Park, Ascot, SL5 7PY, UK.

[9] Computational Science Laboratory, Microsoft Research, 21 Station Road, Cambridge, CB1 2FB, UK.

[10]AB Sustain, 64 Innovation Way, Peterborough, PE2 6FL, UK.

[11] Environmental Change Institute, Oxford University Centre for the Environment, South Parks Road, Oxford, OX1 3QY, UK.

[12] TerraLAB, Computer Science Department, Federal University of Ouro Preto, Ouro Preto, Brazil.

[13]Microsoft Digital, Microsoft, 2 Kingdom Street, London, Cambridge, CB1 2FBW2 6BD, UK

Corresponding Author:

Dr Andrew V. Bradley, Department of Life Sciences, Imperial College of London, Silwood Park, Ascot, SL5 7PY, UK. e-mail: avbradley4@gmail.com ORCID: 0000-0002-6702-9751

Abstract

Ensemble techniques, common in many disciplines, have yet to be fully exploited with spatially explicit projections from land-change models. We trial a land-change model ensemble to assess the impact of policies designed to conserve tropical rainforest at the municipality scale in Brazil, noting the achievements made and challenges ahead. Four spatial model frameworks that were calibrated with the same predictor variables produced 21 counterfactual simulations of the actual landscape. Individual projections with a uniform calibration period gave estimates that between 29% and 68% of the simulated deforestation was saved, but lacked an uncertainty estimate, whilst batch projections from two different model frameworks provided a more dependable mean estimate that 38% and 49% deforestation was prevented with an uncertainty range of 1900 km^2 and 1000 km^2. The consensus ensembles used agreement between the projections and found that the 7 examples with a uniform calibration period produced an error margin of ± 435.94 km^2 and a prevented forest loss estimate of 50%. Using all 21 projections with diverse calibration periods improved these errors to ±179.26 km^2 with a 53% estimate of prevented forest loss. Whilst we achieved a method of combining projections of different frameworks to reduce uncertainty of individual modelling frameworks, demonstrating a control model and accounting for non-linear conditions are challenges that will provide better confidence in this method as an operational tool. Such retrospective evidence could be used to make timely rewards for efforts of governments and municipalities to support tropical forest conservation and help mitigate deforestation.

Key words:

Land-cover modeling, deforestation, environmental policy, SimiVal., Brazil, Green Municipality.

1. Introduction

Ensemble modelling is common in many disciplines such as economics, management, systematics, biomedicine, ecology, meteorology and climatology and involves the comparison and/or combination of several model simulations to give a summarised outcome (Araújo & New, 2007, Knutti *et al.*, 2010). However, one such discipline that has yet to exploit this procedure is spatially explicit land-change modelling. Here, we investigate the potential of applying ensemble modelling techniques to this discipline, where land-cover estimates play a fundamental supporting role in the development of building global climate (e.g. Pielke *et al*; 2002; Pielke *et al.*, 2011) and biodiversity scenarios (e.g. Nelson *et al.*, 2010). There are good reasons for generating ensembles. In climatology, combining several models (Tebaldi & Knutti, 2007) has improved climate predictions (Laprise *et al.*, 2013) by reducing noise or unforced variability that may be present in each of the input models (Taylor *et al.*, 2012). With many categories of output data organised as grids in a time series, data is well suited for climate model ensembles, presenting robust summaries such as precipitation with confidence intervals and reported errors over large spatial areas (e.g. Sylla *et al.*, 2015). In principle, this could be true in land-change modelling. In practice, however, detailed regular time series data are rare with land-change maps, models are run on categorical data at discrete time points, such as a binary forest and non-forest map for a given year (Rosa *et al.*, 2014) often produced from supporting datasets which themselves have discrepancies in their classification (e.g. Bai *et al.*, 2014). Thus climatological ensemble methods are difficult to apply to land-change modelling without adaptation. Ecological modellers, who also have limited data, have investigated the pros and cons of using different ensemble approaches including, bounding box, average, consensus, committee, and Bayesian methods (Araujo & New 2007). Out of these, the consensus ensemble was most promising, as it summarises variability (Buisson *et al.*, 2010) and provides a measure of central tendency that shows the signal above the noise of errors and uncertainty (Araujo & New, 2007). Mapping the consensus model can then give a geographical dimension to direct policy attention and modelling to areas of uncertainty (Diniz-Filho *et al.*, 2009); a useful application for spatially explicit land-change modelling. Consensus modelling is already becoming more prevalent in ecology for resource management and planning (Zhang *et al.*, 2015); a trend which land-change models may well follow if their potential can be demonstrated.

Spatially explicit multiple land-change model investigations are rare, and their focus has been on understanding the relative performance and different outcomes from different modelling frameworks. This has been done using different study locations for nine models (Pontius *et al.*, 2008), using the same input data for four different models (Mas *et al.*, 2014), and post-hoc comparison of different model projections against observed maps of

change (Rosa *et al.*, 2014). Other approaches include the effects of different calibrations within the same model (Soares-Filho *et al.*, 2013) and the variation in land-change projections with different combinations of models (Olmedo *et al.*, 2015). Prestele *et al.*, (2016) used the differences in the characteristics of 11 different global model projections with different input data, to search for uncertain areas of land-use change where models did not agree.

Currently spatially explicit land-change model frameworks are often parameterised to run a range of scenarios producing multiple outcomes. Often restricted in that they conform to the linearity of processes in the calibration period and rarely account for non-linear processes in the projection period (Brown *et al*., 2013), which could be caused by climate, political change, commodity process and natural disasters (Muller *et al*., 2014), they are set up to project examples such as 'business-as-usual' and 'strict-governance' (e.g. Soares-Filho *et al.*, 2006), 'moderate-' and 'severe-climate-change' (e.g. Lapola *et al.*, 2011), or define a probabilistic outcome calculated from multiple model iterations (e.g. Rosa *et al.*, 2013). To mitigate issues with non-linearity some methods have been explored by combining model projections from multiple calibration periods (Rosa *et al.*, 2015) or recalibrating a model several times over (Soares-Filho *et al.*, 2013), but these methods do not combine different model framework projections into an ensemble.

In this work we trial an ensemble of projections from several spatially explicit land-change models to investigate the outcome and how this contrasts to the individual calibrations of the same model and batch model projections. Rather than attempt crystal ball projections of different policy scenarios, we look at an alternative application to make a retrospective analysis of policy intervention on a form of land-change, deforestation. Our design was to calibrate these models with conditions prior to policy intervention and project forwards to investigate what the land-cover would have been in a non-policy outcome (the counterfactual) and compare against what actually happened after policy implementation. In this work we consider the achievements and further challenges that spatially explicit land-change modellers face to appraise policy intervention with this method. Our focus is on deforestation in South American tropical forests where land-change is prevalent (Fernside *et al*., 2005). Here, a method to make a rapid appraisal of counter deforestation policies would help to quickly direct payments and investment to the right areas for forest conservation (Strassbourg *et al*., 2012). This would be a more timely strategy than the complex process of setting and monitoring reference level compliance over longer time scales proposed by Reducing Emissions from Deforestation and forest Degradation (REDD+) programmes (Meridian Institute, 2011; Herold *et al.*, 2012; Angelsen *et al*, 2013) and a simpler way than

modelling policy impacts with complex statistical approaches requiring a number of time dependent variables (Cisneros *et al*., 2015; Sills *et al*., 2015).

2. Method

We undertook this trial by producing land-change projections with a suite of different model structures that are unified only in that they employ the same set of predictor variables at the same study site. With these constraints we were unable to attract agent based and econometric modellers to participate and so we restricted the experiment to spatially explicit models implemented by willing participants. We defined a set period, 2000 to 2004, prior to aggressive deforestation mitigation policies, for the models to be calibrated within. The modellers were to assume that whatever calibration interval was set within this period, future conditions would continue as they were in the calibration period. Assuming this linearity, each projection would be to a fixed year, 2012, producing a counterfactual business-as-usual scenario. With no maps to validate a counterfactual scenario, we tested that patterns of change in the projections were non-random and spatially acceptable. The counterfactual maps were combined into an ensemble. The difference between actual and modelled deforestation was then calculated for the individual projections, batch projections and ensembles to work out the percentage of forest saved from policy intervention. The results were compared to consider the uncertainties of each approach. We then assessed the availability of a control model, where policy is absent or had no impact, to estimate potential error margins of the models and ensembles to increase support for the predicted impact of these policies. The work used data from Paragominas, a municipality located in the state of Pará, Brazil, where a reduction of deforestation rates has been attributed to the enforcement of environmental policy (see study site).

2.1 Study site

Paragominas is a 19, 465 km^2 Brazilian municipality located in the North-East region of the State of Pará (3.2°S, 47.5°W) (Fig. 1a). Established in the 1960s, it is one of many municipalities in Brazil which has undergone extensive deforestation yet still has large areas of forest available for exploitation. Colonisation of the area was assisted by the construction of the Brasília to Belém Highway in 1959 running north-south through the centre of the region. Expansion of cattle pasture (Hecht, 1985), followed by timber extraction in the 1990s (Verissimo *et al.*, 2002; Brito *et al.*, 2010; Nuñes *et al.*, 2014) and then a return to agriculture encouraged by mechanisation (Pinto *et al.*, 2009) has driven deforestation. By 2004, the federal level Action Plan for the Prevention and Control of Deforestation in the Legal Amazon (PPCDAm) (Brazil, 2004) was enforced to

provide better deforestation monitoring and controls in an effort to reduce deforestation rates. In 2008, government campaigns became much more aggressive aiming to reduce the deforestation rates by 20% of the annual average observed between 1996 and 2005 (Nepstad *et al.*, 2009). Paragominas and 36 other municipalities were placed on a 'black-list' for deforestation by the Ministry of the Environment with penalties with respect to credit, access to markets, and additional uncompromising enforcement from IBAMA, the National Environmental Enforcement Agency (Nepstad *et al.*, 2009; Piketty *et al.*, 2015). Liberation from the embargo list required a reduction of annual deforestation to below 40 km^2 per year with at least 80% of the rural properties in the Environmental Registry Database (CAR), forcing landowners to declare their production activities. In 2010, Paragominas was first to meet these criteria to become classed as a 'Green Municipality' (Whately & Campanili, 2013; Viana *et al.*, 2016). Monitoring of deforestation by PRODES shows continued forest clearance since 2000 with a general fall in deforestation rates, with some negative reactions to new policy, often manifest as land grabbing, shown as pulses of increased deforestation rates (Fig. 1b).

[Fig. 1]

2.2 Data

The modelling was confined within the municipality of Paragominas, demarked by the study area bounds (Fig. 1). The model predictor variables were: distance to rivers, distance to roads, settlement boundaries, indigenous lands, altitude, protected area importance, soil type, a binary land-cover classification of forest and non-forest for 2000, 2001, 2002, 2003 and 2004, water bodies, and a large number of socioeconomic data layers referring to income, education and data from the 2000 census (full details and sources in the Online resource). All data was resampled to 200 m resolution for all years between 2000 and 2004 inclusive. We also prepared a binary land-cover classification of forest and non-forest for 2012, representing actual land-cover conditions at the end of the projection period. This 2012 map was withheld from the modelling groups to ensure each model prediction was blind. The calibration period needed to fit into a short period at the beginning of the data set and predict within the remainder of the data set.

For practical purposes using a calibration period from 2000 to 2004 allowed calibration time steps of 1, 2, 3, and 4 years, multiples of 2012, making maximum use of the data available at the time of research. Where a grid cell showed change to non-forest between 2004 and 2012 in each projection, the cell was categorised as deforested and used to calculate the cumulative score of deforestation in that projection. In the ensemble, the proportion of

deforestation in each grid cell was calculated on the level of agreement in each grid cell as opposed an all or nothing calculation for individual projections.

2.3 Participating land-change models

All modelling approaches were spatially explicit and used deterministic, probabilistic and stochastic methods to select pixels for land-cover change. Models were constructed and run by modelling groups familiar with, and involved in the development of, the various platforms. Our volunteer groups were by no means exhaustive of all land-change modelling frameworks, for instance there were no agent based or econometric models, although the ensemble did provide a variety of modelling approaches. We used allocation models, with some able to account for characteristics in the landscape, where each pixel is allocated a probability based on an extrapolation in rates of change, and if capable based on spatial proximity to another variable. We did not include variations in commodity prices which can improve spatial choices in the modelled area, and we acknowledge that these models are weaker when it comes to ascribing amount and location of change (Pontius 2000; vanVliet *et al*., 2006).

Modelling groups were free to choose the transition period(s) they used to calibrate their models and the (sub)set of variables they incorporated into that calibration. The only restriction placed on modelling groups was a requirement to generate binary forest and non-forest prediction maps for the year 2012. In this short calibration period of four years the variability in annual rates means particular choices of beginning and end time periods of the calibration may contain extreme rates of change. For example, a calibration using 2003 data (Fig. 1b) may exaggerate the calibration of land-change rate resulting in excessive or reduced deforestation projections which, when measured against reality, may be considered the least credible projections. The participating model frameworks were Lulcc-R (Moulds *et al.*, 2015), StocModLCC (Rosa *et al.*, 2013), TerrSet-Imazon (Clark Labs, 2009) and TerrSet-Clark Labs (Clark Labs, 2009; Eastman, 2014)

2.3.1 Lulcc-R (LUCR)

All deforested cells were identified and an R machine learning algorithm, randomForest, was used to relate observed land-cover change against all the available 81 predictor variables combined. A training partition with 10% of the data to fit the model was applied, with the remaining data used to test the models with the Receiver Operator Characteristic (ROC). Using the obtained statistical calibration, a simple demand scenario based on a

continuation of the observed deforestation rate was constructed. Then, an ordered allocation procedure was followed whereby cells were arranged in order of suitability for deforestation. Working from high suitability to low suitability, the cells were compared to a random number between 0 and 1 drawn from a uniform distribution. The cell in question was deforested if the suitability was greater than the random number (thus the results are one instance of multiple possible realisations), producing a forest, non-forest map.

2.3.2 StocModLCC (SM)

SM uses, for each calibration period, a stepwise logistic regression to determine the best combination of predictor variables, using a maximum likelihood score determined from multiple iterations with different variable parameterisations. These predictor variables (plus the contagion parameter characteristic of the model) were then used to simulate land-cover change. The model also includes a stochastic element and avoids deterministic behaviour by using a Markov Chain Monte Carlo sampling technique. Consisting of a posterior mean and confidence range for each of the predictor variables a draw from this distribution adds a random element on a pixel by pixel basis (Rosa *et al.*, 2013). The model is run 100 times for each calibration period, producing binary forest/non-forest maps, which when combined provide a probability that a pixel was deforested. For the ensemble, a threshold probability score derived from the average number of change pixels of the 100 runs was then applied to the probability scores to create a binary forest and non-forest map.

2.3.3 TerrSet Land-change modeler – Imazon (TS-IM)

For this study, the TerrSet Land-change Modeler (LCM) with the Multi-Level Perceptron (MLP) machine learning algorithm was used. LCM estimated a combination of predictor input variables that best represent the occurrence of non-forest in the forested areas. Based on the input predictor variables, the MLP calculated the potential of a pixel to change, to produce a probability map of change. The model then calculated the proportion of each land-cover transition and allocated change accordingly by imposing the rate calculation from the calibration. No dynamic or constraint variables were used and the rate of change to non-forest between 2004 and 2012 was calculated using the Markov chain option.

2.3.4 TerrSet Land-change modeler – Clark Labs (TS-CL)

This model also used the TerrSet LCM machine learning method described above. The group operating it, however, selected a wider range of predictor variables and calibration periods. For each calibration period, four

sets of conditions were produced representing, no dynamic variable and no constraint, no dynamic variable with constraint, dynamic variable with no constraint, and dynamic variable and constraint. The dynamic variable was calculated by recalculating the distance to non-forest from the result of the previous iteration. This new distance to non-forest was used to recalculate the transition potential of the pixels using the weights from the original MLP neural network model. The constraint was a 70 % reduction in modelled transition potential image values within indigenous areas. The proportion of each land-cover transition was then allocated by imposing the rate calibration.

2.5Validation and error quantification

The enforcement of policy imposes non-linear conditions on the study area and the actual deforestation outcome will be different from modelled linear conditions. As the assumption of process linearity is broken there was no reasonable way to validate each of the model land-cover projections against observed land-cover change. This poses challenges in presenting the robustness of the method, how realistic the model projections are and quantification of errors and uncertainty, important requirements to accept this method, particularly if the outcome is to evaluate and reward policy activity. Instead, to make sure the projections are realistic, we tested if the model projections were better than a paired random model for each of the individual model projections, in which exactly the same number of change pixels (forest to non-forest) were randomly distributed. Each of the pairs were then tested for realism by calculating the Hit (H - predicted change where there was observed change), Miss (M - observed change which was not predicted), False alarm (F - prediction of change where there was no observed change) and Correct persistence (C - predicted as no change where there was also no change in the observed) statistics (referred to as HMFC) relative to the observed 2012 map. For model projections that performed better than their random counterparts, the HMFC statistics were expected to be worse for the random case, and for model forecasts with more similarity to the 2012 map it was expected to find more hits, fewer misses and fewer false alarms. To verify the HMFC results the Jaccard (2012) Similarity Index, $JSI = H / (H + M + F)$, was used to provide a one value summary of the similarity between the observed and predicted data, although the contribution of hits and misses cannot be discriminated. As most of the landscape actually persists including C in JSI causes high agreement, implying that all models are very good. We are concerned with agreement in change so correct persistence C is excluded. In a further test on realism of the predicted landscape, all projections were analysed with SimiVal, the Similarity Validation tool (Bradley *et al.*, 2016). This tool quantifies how dissimilar each projected land-cover change map was from the reality of 2012,

i.e. if the projection is a perfect case, more random or more systematically-biased in terms of landscape dimensions such as allocation, structure, and spatial autocorrelation.

Understanding the errors and uncertainty in the methodology also has challenges. Since the model assumes linearity then it would be ideal to test the procedure against linear conditions using a control model with independent data to understand if the model over- or under-predicts. In order to find this condition, a search for a location where deforestation rates were not supressed was made, ideally with no policy enforcement and within reasonable distance of the study site where controls are likely to be similar. For each adjacent municipality around Paragominas average rates were calculated from PRODES deforestation statistics for the period 2000 to 2004, for the pre-policy conditions, and 2005 to 2012, for the policy enforcement period. An ANOVA test was applied to find significantly different cases.

2.6 Ensemble

Using the consensus technique that ecological modellers have found to be desirable (Araujo & New 2007), model agreement was determined by stacking together the model projections and summing the number of times each pixel had been selected to deforest. Where at least one model predicted change in the pixel the mean and variance was then calculated. Ensemble results comprised a stack of binary data containing 1's and 0's, so we used a bootstrapping technique to quantify uncertainty around the mean prediction. We resampled the population of values in each pixel 1000 times to create a 95% confidence interval. The new mean, or level of agreement, was used as a value to calculate the proportional loss of forest in each pixel. These scores were then aggregated for the whole of the study area. This was repeated for the upper and lower 95% confidence limits, to give the upper and lower bounds for the forest loss estimate. The forest loss estimates were compared to forest loss estimates for the individual projections and against actual forest lost within the study area.

3. Results

3.1 Projection and ensembles

Overall, there were 21 projections from the modelling groups, with seven projections sharing the same calibration interval of four years and a further 14 projections of overlapping annual, biannual and triple year time intervals (Table 1). Two of the models had a batch of projections and the rest were stand-alone projections. Using the variety of calibration intervals, two consensus ensembles were created to investigate how calibration

interval would affect the ensemble. One where the calibration period was consistent for all of the projections, from 2000 to 2004 called 'Uniform cal', and another ensemble using all projections regardless of the calibration period called 'Diverse cal'.

[Table 1]

3.2 Validation

Individual models had low JSI values (from 0.06 to 0.18), which was better than their random counterparts (0.03 to 0.06). The HMFC statistics show a high percentage of false alarms (3.3 % to 10.3 %) for individual models, with just 0.4 % to 2.3 % hits and 1.8 % to 3.6 % misses. This arose primarily because most of the projections allocated more pixels for change than actually occurred, raising false alarms. Again, however, the model projections performed better than their random counterparts which had a lower percentage of hits (0.2 % to 0.6 %) and a greater percentage of misses (3.2 % to 3.9 %). The two consensus ensembles, Uniform cal, and Diverse cal, had JSI scores of 0.13 and 0.14 and also performed better than their random counterparts. Before correction for agreement, the ensembles projected a much greater absolute change (almost seven times more area than 2012, 3884.0 and 5191.2 km^2) than actual. Consequently, HMFC statistics had the greater numbers of hits (3.0 and 3.4%) and false alarms (16.9 and 23.2 %) with fewer misses (1.0 and 0.6%). However, once the data was bootstrapped to account for the probability of agreement between projections the proportional loss of forest fell for Uniform cal to 1565.5 km^2 and for Diverse cal to 1653.2 km^2. The ensemble probabilities are shown (Fig. 2).

[Fig. 2]

In terms of allocation, structure and spatial autocorrelation, SimiVal validation (Fig. 3), indicated that the model projections demonstrated outcomes with no direct replication of 2012, the extreme cases of systematic-bias or randomness. Overall the different combinations of calibrations and models have, as expected, produced a variety of modelled outcomes. The ensembles Uniform cal and Diverse cal that combine this variety are found within the cloud of points described by the individual model projections that comprise each ensemble, enforcing their ability to provide a consensus opinion. The smaller ensemble, Uniform cal, was further away from the actual perfect case of 2012 as the component projections were distant from perfect, whereas Diverse cal encompassing all 21 individual projections had a greater spread with some projections closer to the perfect case 2012.

[Fig. 3]

The average aggregate deforestation rates for 2000 to 2004 and 2005 to 2012 show how, although forest loss rates are different in different municipalities, there was a marked decrease in rates for virtually all municipalities between the two time points (Fig 4 (a)). No municipalities in 2005-2012 continued to deforest at the same rates as 2000-2004 and none of the rate differences were significantly different from the rest. A suitable control case to quantify errors and uncertainty of the projections and ensemble could not be found, even when the search was expanded to all municipalities in Brazil. It seems that events after 2004 resulted in a net decrease in deforestation. The only way to identify an example resembling control conditions was to reduce the aggregate time step (Fig 4 (b)). In this case, coincidental to the 2008 black listing policy where average deforestation rates increased in three municipalities. Although this indicated potential for control conditions for this method this was an inappropriate interval for the model calibration and target date used.

[Fig. 4]

3.3 Forest loss estimates

The actual amount of deforestation lost from 2004 to 2012 for Paragominas was recorded as 19 662 pixels or 786.5 km^2 spread across the municipality (Fig. 1), a value lower than projected by most model projections ranging from 508.6 km^2 through to 2425.0 km^2 and the ensembles projecting around 1600 km^2.

Single model projections from all four modelling frameworks, SIN (Fig. 5), calibrated between 2000 and 2004 produced a range of estimates, with the most similar estimates from the TS-CL projections (lowest ~1104.96 km^2) through to SM which predicted 2425.04 km^2, over three times more loss than reality. This indicates the influence different frameworks can have on rate calculations and spatial choices (if accommodated in the model) even when the calibration length and predictive variables are similar. Conversely the TS-CL series calculated similar amounts as they were projected from similar rate calculations. However, these projections were observed to allocate change in different places because of the enforced calibration conditions of protected areas and changing proximity to non-forest. Using the extremes of these single projections the difference saved between the business-as-usual projections and actual deforestation as a result of policy enforcement is quantified to be between 29 % (TS-CL 7) and 68 % (SM).

The individual batch projections from SM (7) and TS-CL (12) give quite variable amounts and ranges of carbon stock differences, BX (Fig. 5). The SM framework predicted a minimum of 508.64 km^2 and maximum of 2425.04 km^2 with an average loss of 1555.19 km^2. The TS-CL framework predicted a minimum of 861.44 km^2 and maximum of 1862.16 km^2 with an average loss of 1277.54 km^2. These results highlight the variance on the model outcomes by using short and different calibration periods, even within a single modelling framework. For TS-CL the effect of using a calibration with a lower deforestation rate is that those projections have the least change. Those runs from 2000 to 2004 have moderate amounts of change and those calibrations using the 2003 to 2004 interval have the greatest amounts of change. This is not as clear cut in SM as the largest 4 year and 2 year intervals have the greatest amount of change and single year intervals mostly result in the least amount of change, despite sometimes including the 2003 low rate, suggesting that in this framework the choice of different predictor variables and spatial component for each projection have a strong influence on the model outcomes. The difference in forest loss from actual was: +786.69 km^2 for the average of SM (7), +491.04 km^2 for the average of TS-CL (12). The difference saved between the business-as-usual projections and actual deforestation as a result of policy enforcement is quantified as 49 % (SM) and 38 % (TS-CL).

The bootstrapped consensus ensembles gave similar projections of forest loss Uniform cal, 1567.75 km^2 and, Diverse cal, 1656.66 km^2, with the main difference being that the Diverse cal had a narrower confidence interval around the ensemble prediction, EN (Fig. 5). The difference in forest loss from actual was: +781.25 km^2 for Uniform cal and +870.16 km^2 for Diverse cal indicating policy enforcement can be quantified to have saved 50 % (Uniform cal) and 53 % (Diverse cal) of the projected deforestation.

[Fig.5]

4. Discussion

In this ensemble trial we were able to approach some of the challenges of building a land-change model ensemble. We built the ensemble using the consensus approach which has been successfully applied in other disciplines such as ecology (Araujo and New, 2008; Buisson *et al.*, 2010; Zang *et al.*, 2015). We combined outputs of four model frameworks and were able to constrain the modelling conditions to a standardised set of predictive variables and limit the calibration constraints to a time before new deforestation mitigation policies were enforced. The outcome was a series of counterfactual scenarios in 2012 as if policies were not implemented. Most projections and the ensembles showed that in the absence of these policies, the study area

modelled would have suffered greater deforestation than reality in 2012, although there were contrasts in terms of the amounts, the range of results, and certainty between the single projections, batch projections and the ensembles. We emphasise that these are experimental results to illustrate this trial in a land-change modelling ensemble and there are further methodological considerations before this approach can become an operational tool for a retrospective assessment of policy impacts to mitigate deforestation.

With no independent data to prove that the ensembles provide reliable results, we executed two methods to assess the realism of the model projections that fed into the ensembles: the HMFC statistics and SimiVal comparison. HMFC statistics for the random versions of the models were different to the model projections. We found that all individual model projections were more realistic than the random projections which produced many more false alarms and fewer hits, because random projections disregard the probability and likely locations of land-cover change that a calibrated model will calculate. SimiVal showed different degrees of variation in structure and outcomes of land-cover change, the projections were neither strongly random nor systematically-biased, and structural changes of land-cover change were realistic, albeit in different manners and with outcomes that differed in their specifics (Fig. 3). Importantly, the high level of dissimilarity among projections clearly demonstrates that the ensemble was not built from a set of model clones.

Out of all the projections and methods, single projections from different frameworks produce widely different results. Using a stand-alone single model projection would be unreliable without verification. It is impossible to know if individual counterfactuals are anywhere within range of a likely non-policy scenario and there is no conventional form of validation to test the final projection. On the other hand, batch projections from individual frameworks can provide a range of outcomes, along with uncertainty marked by extreme individual projections. Analysts can be selective with outcomes that are realistic and appropriate for their needs or use a mean outcome within an envelope of range. However, different frameworks can produce a different mean outcome and range in comparison to another framework. In fact, projections from different frameworks may never overlap and without running two models analysts would never know. Thus, analysts using a single model framework with many projections are still uncertain of the reliability of their outcome without a second opinion. In contrast, the ensemble produces a single result with an error margin based on the bootstrapped consensus provided by several opinions, in this case from four model frameworks. This was a significant outcome given the differences in internal structures, calibrations and parameterisations of the four parent modelling frameworks. With this in mind, combining all the projections into an ensemble to find the areas of agreement and variability (uncertainty)

helps compensate for the lack of validation data. We also found that increasing the number of projections in the ensemble, regardless of their calibration interval, reduced the error margin bringing greater certainty to the predicted estimate. On the flip side, by bootstrapping into a single result, an element of the exploratory nature of modelling is absent as extreme cases and individual projections are no longer represented.

We identified a number of hurdles to make the ensemble, and in presenting the outcomes the process also raised further challenges that will strengthen the operational capability of this method. Before embarking on this trial, we needed to find and focus on a study location, Paragominas, where policy impacts have been reported to slow land-cover change rates through various control policies. This would test the theory that a projection based on the pre-policy conditions would predict greater magnitudes of land-cover change in comparison to policies that supressed deforestation. We identified a time period to project into after hard line environmental policy changes were implemented from 2004. This included the PPCDAm 2004 national policy and the 2008 localised black-listing. By making the start conditions equal and insisting on a constrained calibration period using the same calibration data we aimed for a new and inclusive approach to include and account for model variability, rather than simply compare model differences (Pontius *et al.*, 2008; Mas *et al.*, 2014).

One of the biggest challenges we have found is being able to provide confidence in the ensemble outcome in the absence of a parallel scenario where deforestation continued as normal. This would provide error margins if the actual processes prove to be slightly non-linear and the parallel model does not quite match. Without this quality control it would be difficult to justify an evaluation of policy impacts and distribute rewards. In this case national level policies, the PPCDAm of 2004 and actions of IBAMA up to 2008, or some other intervening factor (Müller *et al*., 2014) influenced many municipalities as indicated by the overall fall in aggregate percentage forest losses (Fig. 4(a)). Furthermore, we cannot be sure which policy drove the change, as one of the policies may have been completely ineffective. Concentrating the model calibration and projections on shorter time scales and single policies, perhaps specific to some municipalities and not others, may eradicate these issues. For example, focus on a single policy such as the 2008 blacklisting (Fig. 4(b)).

Our pre-policy era was intended to represent linear conditions with near uniform deforestation rates. In reality, as the PRODES data shows, land-change rates are volatile and vary inter-annually over a short time period (Fig. 1b), and linear conditions are in reality rare over such short time scales. This was reflected in the decisions of some modellers to make multiple calibrations within the 2000 to 2004 calibration interval to try and account for

continuously changing rates of deforestation. Using data within a short calibration interval introduces a greater probability of calculating projections from extreme, but real rates, as greater intervals may average out deforestation rates. Excessive rates over short time intervals are unlikely to have been sustained during the projection period, questioning the credibility of some counterfactuals. Longer time steps found in many modelling studies (van Vliet *et al*., 2016) may be less susceptible to year-on-year rate differences but it is still possible to miss-represent future conditions if calculated rates between the selected calibration time points are too high or too low. A general issue in land-change modelling is that calibration has to use the data that is available, and using short time steps as we have done is not unheard of (Lapola *et al*., 2011; Maeda *et al*., 2011; Yanai *et al*., 2012, Rosa *et al*., 2013). Additionally, using historical data to calibrate models cannot account for unpredictable non-linear shifts caused by climatic events, economic fluctuations, political shifts and natural disasters (Müller *et al*., 2014). Nevertheless, whatever rates were used in the calibrations, rather than treating each projection separately, as may be done with exploratory modelling (Deadman *et al*., 2004) bootstrapping into a consensus ensemble found the common ground through agreement because the most extreme projections generated a weak consensus of opinion with low probability scores. The ensemble smoothed out the less credible results by reducing the biases of projections with extreme calibration results.

Isolating the projections with the longest calibration period (Uniform cal) showed how the outcome of the ensemble was influenced by short variable calibration rates. We found the apparently paradoxical result that the higher variability among model projections led to tighter confidence limits around the ensemble mean (Diverse cal). This outcome is, however, typical of climate ensembles as noise and unforced variability are reduced (Taylor *et al.*, 2012), which has also been found to be best summarised with the consensus ensemble (Buisson *et al.*, 2010). An ensemble with more models provides a more statistically sound outcome regardless of the increased variance, a result that is concurrent with many other disciplines where more models are often shown to provide better outcomes (e.g. Taylor *et al.*, 2012; Matre *et al.*, 2014; Laprise *et al.*, 2013; Zang *et al.*, 2015).

Other intervening factors that contribute to non-linear conditions that our approach did not allow for may be changing commodity prices, switches in land-use, actors that may receive certain sanctions or benefits, and the spatial footprint that these temporal changes will act on, often accounted for with agent based and econometric models. Our projections relied on a weaker element of allocation models to project location and amounts of change that are consequently seldom validated (Pontius 2000; van Vliet *et al*., 2006). These additional superimposed factors may also drive spatial patterns of change; tricky to calculate with allocation models. There

were some spatial choices in the modelling frameworks, SM predicted deforestation based on proximity to previously deforested pixels, and ID-CL 5, 6 and 7 considered protected land and proximity to non-forest pixels. Although allocation models are not so reliable at predicting amounts, a spatial representation is important as many modelling exercises would not have any impact without flagging the proximity of threats to untouched forests, protected areas and so on (e.g. Soares-Filho *et al*., 2006; Lapola *et al*., 2010; Alcamo *et al*., 2011). As driving forces of deforestation are spatially variable (Geist & Lambin 2002) modelling variations at municipality scale would be an extremely useful method for quantification of spatial similarity (Zang *et al.*, 2015). Modelling policy impacts countrywide may show the degree of influence that national policy enforcement has on different municipalities depending on their spatial proximity to agencies and law enforcing infrastructure and may provide an alternative approach to using countrywide historical rates to set reference levels in REDD+, particularly for smaller scale detailed 'tier3' estimates (Angelsen *et al.*, 2013). Governments would be able to provide near real time evidence for prevented deforestation to international committees and receive their rewards, which may be monitory or as part of carbon trading deals. Local state level governments would be able to provide the same evidence to national governments and take their share of international rewards, and if spatial detail can be resolved, individual stakeholders may be able to claim from national or international funds.

Having overcome a series of challenges to create an ensemble of land-cover model projections we have highlighted the current limitations and further improvements of this method. If operational, the method could compliment tools that typically require long term monitoring and compliance, like REDD+. This method could reduce the necessity, for intricate measurements (Meridian Institute, 2011), advance plans and predictions (Herold *et al.*, 2012) and time lags in payment (Angelsen *et al.*, 2013) and perhaps contribute more quickly to the estimated cost of preventing deforestation (Nepstad *et al*, 2009). With fine tuning, this new methodological approach of ensemble modelling has potential to provide a retrospective benchmark to immediately reward efforts to avoid the effects of land-cover change (Strassburg *et al.*, 2012). Furthermore, this method could be applied to evaluating preservation of biodiversity and ecosystem services, and identifying specific areas that require more protection.

Acknowledgements

Spatial data sets were supplied by Imazon, Brazil, the socioeconomic data from Instituto Nacional de Pesquisas Espaciais, Brazil, and prepared at Imperial College by Igor Lysenco. Funding for the land-cover modelling Tansley working group was from Natural Environment Research Council Biodiversity & Ecosystem Service Sustainability grant, NERC/PR100027. AVB, IMDR and RME were supported by European Research Council grant 281986. This paper represents a contribution to Imperial College's initiative in Grand Challenges in Ecosystems and the Environment. We also thank the feedback of the anonymous referees on this paper.

References

Angelsen A, Ainembabazi JH, Bauch S C, Herold M, Verchot L, Hänsel G, Schueler V. Toop G, Gilbert A. Eisbrenner K. (2013) Testing methodologies for REDD+: Deforestation drivers, costs and reference levels. Technical Report, UK Department for Energy and Climate Change, DECC.

Araújo MB, New M, (2006) Ensemble forecasting of species distributions. Trends in Ecology and Evolution 22:42-47

Bai Y, Feng M, Wang H, Zhu Y, Liu Y (2014) Assessing Consistency of Five Global Land Cover Data Sets in China. Remote Sensing, 6:8739-8759, doi:10.3390/rs6098739

Batty M, Torrens PM, (2005) Modeling and prediction in a complex world. Futures 37:745-766

Bradley AV, Rosa IMD, Pontius RG, Ahmed SE, Araújo MB, Brown DG, Brandão Jr. A, Câmara G, Carnerio TGS, Hartley AJ, Smith MJ, Ewers RM (2016) SimiVal, a Multi-Criteria Map Comparison Tool for Land-Change Model Projections. Environmental Modelling and Software 82:229-240

Brazil (2004) Plano de Ação para a prevenção e controle do desmatamento da Amazônia Legal, Grupo Permanente de Trabalho interministerial para a redução dos índice s de desmatamento da Amazônia Legal (Decreto de 3 de Julho de 2003). Presidência da República Casa Civil, Governo Federal, Brazil. pp 156 http://www.fundoamazonia.gov.br/FundoAmazonia/export/sites/default/site_pt/Galerias/Arquivos/Publicacoes/PPCDAM.pdf

Brito B, Souza Jr. C. Amaral P (2010) Reducing emissions from deforestation at municipal level: a case study of Paragominas, Eastern Brazilian Amazon. In: Everything is Connected: Climate and Biodiversity in a Fragile

World, ed. UK Department for Environment, Food, and Rural Affairs, Brasil: DEFRA, British Embassy Brasilia pp. 29-31

Brown DG, Verburg PH, Pontius Jr. RG, Lange MD (2013) Opportunities to improve impact, integration and evaluation of land-change models. Current Opinion in Environmental Sustainability 5:425-457

Buisson L, Thuiller W, Casajus N, Lek S, Grenouillet G (2010) Uncertainty in ensemble forecasting of species distribution. Global Change Biology 16:1145–1157, doi: 10.1111/j.1365-2486.2009.02000.x

Cisneros E, Zhou SL, Börner J (2015) Naming and Shaming for Conservation: Evidence from the Brazilian Amazon. PLoS ONE 10(9): e0136402. doi:10.1371/journal.pone.0136402

Clark Labs, (2009) The Land-change Modeler for ecological sustainability. IDRISI Focus Paper, Worcester, M: Clark University. http://www.clarklabs.org/applications/uplaad/Land-Change-Modeler-IDRISI-Focus-Paper-pdf

Deadman P, Robinson D, Moran E, Brondizio E (2004) 'Colonist household decision making and land-use change in the Amazon Rainforest: an agent-based simulation' Environment and Planning B: Planning and design 2004 31:693-709

Diniz-Filho JA, Bini LM, Rangel TF, Loyola RD, Hof C, Nogués-Bravo D, Araújo MB (2009) Partitioning and mapping uncertainties of forecast species turnover under climate change. Ecography 32:897-906.

Eastman R, Solorzano LA, Fossen ME (2005). Transition potential modelling for land-cover change. D. Maguire, M. Batty; M. Goodchild (Eds). GIS, Spatial Analysis, and Modeling, ESRI Press, California, pp. 357-385.

Geist HJ, Lambin EF, (2002) Proximate Causes and Underlying Driving Forces of Tropical Deforestation. BioScience 52:143-150.

Hecht S (1985) Environment, development and politics: capital accumulation in the livestock sector in eastern Amazonia. World Development 13:663-684.

Herold M, Angelsen A, Verchot LV, Wijaya A, Ainembabazi JH (2012) A stepwise framework for developing REDD+ reference levels. In Angelsen A, Brockhaus M, Sunderlin WD, Verchot, LV (eds) 2012 Analysing REDD+: Challenges and choices, Ch 16. CIFOR, Bogor, Indonesia.

Jaccard P, (1912) The distribution of the flora in the alpine zone. New Phytol. 11:37-50

Knutti R (2010) The end of model democracy? An editorial comment. Climate Change 102:395-404.

Lapola DM, Schaldach R, Alcamo J, Bondeau A, Msangi S, *et al.* (2011) Impacts of Climate Change and the End of Deforestation on Land Use in the Brazilian Legal Amazon. Earth Interact. 15:1-29.

Laprise R, Hernández-Díaz L, Tete K, Sushama L, Šeparović L, Martynov A, Winger K, Valin M, (2013) Climate projections over CORDEX Africa domain using the fifth-generation Canadian regional climate model (CRCM5). Clim Dyn 41:3219–3246.

Mas J-F, Kolb M, Paegelow M, Olmedo MTC, Houet T (2014) Inductive pattern-based land use/cover change models: A comparison of four software packages. Environmental Modelling & Software 51:94-111

Maeda EE, de Almeida CM, de Carvalho Ximenes A, Formaggio AR, Shimabukuro YE, Pellikka P (2011) Dynamic modelling of forest conversion: simulation of past and future scenarios of rural activities expansion in the fringes of the Xingu National Park, Brazilian Amazon. Int J Appl Earth Obs Geoinf. 13:435–446. doi: 10.1016/j.jag.2010.09.008.

Meridian Institute (2011) Guidelines for REDD+ Reference Levels: Principles and Recommendations. Prepared for the Government of Norway, by Arild Angelsen, Doug Boucher, SandranBrown, Valérie Merckx, Charlotte Streck, and Daniel Zarin. Available at www.REDD-OAR.org.

Matre P, Wallach D, Asseng S, Ewart F, Jones JW, *et al.*, (2014) Multimodel ensembles of wheat growth: many models are better than one. Global Change Biology, doi: 10.1111/gcb.12768

Moulds S, Buytaert W, Mijic A (2015) An open and extensible framework for spatially explicit land use change modelling: the lulcc R package, Geosci. Model Dev. 8:3215-3229. http://www.geosci-model-dev.net/8/3215/2015/gmd-8-3215-2015-metrics.html

Müller D, Sun Z, Vongvisouk T, Pflugmacher D, Xu J, Mertz O (2014) Regime shifts limit the predictability of land-system change. Global Environmental Change 28:75-83

Nelson E, Sander H, Hawthorne P, Conte M, Ennaanay D, Wolney S, Manson S, Polasky S (2010) Projecting Global Land-Use Change and Its Effect on Ecosystem Service Provision and Biodiversity with Simple Models. PLoS One 5(12) https://doi.org/10.1371/journal.pone.0014327

Nepstad D, Soares-Filho BS, Merry F, Lima A, Moutinho P, Carter J, Bowman M, Cattaneo A, Rodrigues H, Schwartzman S. McGrath DG, Stickler CM, Lubowski R, Piris-Cabezas P, Rivero S, Alencar A, Almeida O, Stella O (2009) The end of deforestation in the Brazilian Amazon. Science 326:1350-1351.

Pan Y, Birdsey RA, Phillips OL, & Jackson RB (2013) The Structure, Distribution, and Biomass of the World's Forests. Annual Review of Ecology, Evolution, and Systematics 44:593-622. doi:10.1146/annurev-ecolsys-110512-135914.

Picketty M-G, Poccard-Chapuis R, Drigo I, Coudel E, Plassin S, Laurent F, Thâles M (2015) Multi-level Governance of Land Use Changes in the Brazilian Amazon: Lessons from Paragominas, State of Pará. Forests 6:1516-1536; doi:10.3390/f6051516

Pielke RA, Marland G, Betts RA, Chase TN, Eastman JL, Niles JO, Niyogi DS, Running SW (2002) The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. Phil. Trans. R. Soc. Lond. A 360:1705-1719, DOI: 10.1098/rsta.2002.1027

Pielke R A, Pitman A, Niyogi D, Mahmood R, Alpine C, Hossain F, Goldewijk KK, Nair U, Betts R, Fall S, Reichstein M, Kabat P, de Noblet N (2011) Land use/land cover changes and climate: modeling analysis and observational evidence. Wiley Interdisciplinary Reviews: Climate Change 2:1757-7799. DOI: http://dx.doi.org/10.1002/wcc.144

Pinto A, Amaral P, Souza Jr., *et al.* (2009) Diagnostico Socioeconomico e Florestal do Municipio de Paragominas. Report. Imazon, Belem, Para, Brasil pp 65

Pontius RG, (2000) Quantification error versus location error in comparison to categorical maps. Photogramm. Eng. Remote Sens. 66:1011-1016

Pontius RG, Boersma W, Catella J-C, Clarke K, de Nijs T, Dietzel C, Duan Z, Fotsing E, Goldstien N, Kok K, Koomen E, Lippit C. D, McConnell W, Sood AM, Pijanowski B, Pithadia S, Sweeney S, Trung TN, Veldkamp

AT, Verburg PH (2008) Comparing the input, output, and validation maps for several models of land-change. Annals of Regional Science 42:11-37

Portela R, Rademacher I (2001) A dynamic model of patterns of deforestation and their effect on the ability of Brazilian Amazonia to provide ecosystem services. Ecol. Modell. 143:115-146

Prestele R, Alexander P, Rounsevell, MDA, Arneth A, Calvin K, Doelman J, Eitelberg DA, Engström K, Fujimori S, Hasegawa T, Havlik P, Humpenöder F, Jain AK, Krisztin T, Kyle P, Meiyappan P, Popp A, Sands RD, Schaldach R, Schüngel J, Stehfest E, Tabeau A, Van Meijl H, Van Vliet J. and Verburg PH, (2016) Hotspots of uncertainty in land-use and land-cover change projections: a global-scale model comparison. Global Change Biology. doi:10.1111/gcb.13337

Rosa IMD, Ahmed SE, Ewers RM (2014) The transparency, reliability and utility of tropical rainforest land-use and land-cover change models. Global Change Biology 20:1707-1722.

Rosa IMD, Purves D, Carreiras JMB, Ewers RM (2015) Modelling land cover change in the Brazilian Amazon: temporal changes in drivers and calibration issues. Regional Environmental Change 15:123-137.

Rosa I, Purves D, Souza Jr. C, Ewers RM (2013) Predictive modelling of contagious deforestation in the Brazilian Amazon. PLOS One 8: e77231. doi:10.1371/journal.pone.0077231

Sills EO, Herrera D, Kirkpatrick AJ, Brandão Jr. A, Dickson R, Hall S, Pattanayak S, Shoch D, Vedoveto M, Young L, Pfaff A (2015) Estimating the Impacts of Local Policy Innovation: The Synthetic Control Method Applied to Tropical Deforestation. PLoS ONE 10:e0132590. doi:10.1371/journal.pone.0132590

Soares-Filho B, Alencar A, Nepstad D, Cerqueira G, del Carmen VDM, Rivero S, Solórzano L, Voll E (2004). Simulating the response of land-cover changes to road paving and governance along a major highway: the Santarém-Cuiabá corridor. Global Change Biology 10:745-764.

Soares-Filho B, Rodrigues H, Follador M (2013) A hybrid analytical-heuristic method for calibrating land-use change models. Environmental Modelling and Software 43:80-87

Soares-Filho BS, Nepstad D, Curran LM, Cerquieira G, Garcia RA, Ramos CA, Voll E, McDonald A, Lefebvre P, Schlesinger P (2006) Modelling conservation in the Amazon Basin. Nature 440:540-523

Strassburg BBN, Rodrigues ASL, Gusti M, Balmford A, Fritz S, Obersteiner M, Turner RK, Brooks TM (2012) Impacts of incentives to reduce emissions from deforestation on global species extinctions. Nat. Clim. Change DOI: 10.1038/NCLIMATE1375

Sylla MB, Giorgi F, Pal JS, Gibba P, Kebe I, Nikiema M (2015) Projected Changes in the Annual Cycle of High-Intensity Precipitation events over West Africa for the Late Twenty-First Century. J. Clim. 28:6475-6488.

Taylor KE, Stouffer RJ, Meehl GA (2012) An Overview of CMIP5 and the Experiment Design. Bull. Am. Meteorol. Soc., Vol:485-498. DOI:10.1175/BAMS-D-11-00094.1

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philos. Trans. R. Soc., A: Mathematical, Physical and Engineering Sciences 365:2053-2075

van Vliet J, Bregt AK, Brown DG, van Delden H, Heckbert S, Verburg PH, (2016) A review of current calibration and validation practices in land-change modelling. Environmental Modelling & Software 82:174-182

van Vliet J, Hagen-Zanker A, Hurkensa J, van Deldena H, (2013) A fuzzy set approach to assess the predictive accuracy of land use simulations. Ecol. Modell. 261–262:32–42

Viana C, Coudel E, Barlow J, Ferreira J, Gardner T, Parry L. (2016) How does hybrid governance emerge? Role of the elite in building a Green Municipality in the Eastern Brazilian Amazon. Environmental Policy and Governance, 26:337–350. doi: 10.1002/eet.1720.

Verissimo A, Lima F, Lentini M (2002) Polos Madeireiros do Estado de Para. Belem, Para, Brazil, Imazon, pp 74

Whately M, Campanili M (2013) Green Municipalities Programme: Lessons learned and challenges for 2013/14. Programa Municípios Verdes, Goveno do Pará, Belém, Brazil. pp 98.

Yanai AM, Fearnside PM, Graça PMLA, Nogueira EM (2012) Avoided deforestation in Brazilian Amazonia: simulating the effect of the Juma sustainable development reserve. Forest Ecology and Management 282:78-91.

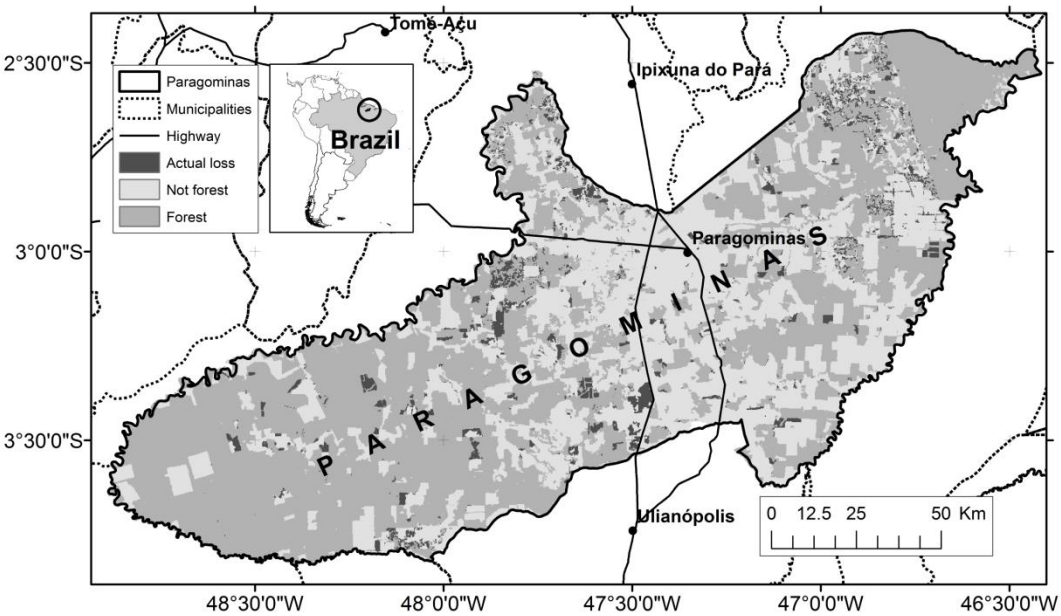
Zhang L, Liu S, Sun P, Wang T, Wang G, Zhang X, *et al.* (2015) Consensus Forecasting of Species Distributions: The Effects of Niche Model Performance and Niche Properties. PLoS ONE 10:e0120056. doi:10.1371/journal.pone.0120056

Table 1. Details for the 21 projections of the four models showing the ensemble group, batch groups (illustrated in the analysis Fig. 5), the calibration period, the projected amount of change calculated in km^2, a split bar chart with the percentage Hit (H), Miss (M), False alarm (F) and Correct persistence (C) statistics (HMFC), the top is the model projection and the bottom is the random model, and the Jaccard Similarity Index (JSI) for the projection and a random equivalent. Models are LUCR = Lulcc-R, SM = StocMod, TS-IM = TerrSet LCM Imazon, TS-CL = TerrSet LCM Clark Labs. The bottom row shows actual forest loss in 2012.

Model framework			Calibration period					LCC	Hit , miss, false alarm, correct persistence as % of potential change	JSI scores	
Model	**Projection ref.**	**Batch / ensemble grp. group**	**2000**	**2001**	**2002**	**2003**	**2004**	**km^2 (1dp)**	H M F C % 0 25 50 75 100 (Upper = projection, lower = random)	**Projection**	**random**
LUCR	1	*						1963.6		0.12	0.0
SM	2	*•						2425.0		0.16	0.0
TS-IM	3	*						1111.9		0.16	0.0
TS-CL	4	* †						1105.6		0.15	0.0
	5 [c]	* †						1105.0		0.15	0.0
	6 [d]	* †						1105.3		0.16	0.0
	7 [d,c]	* †						1105.0		0.17	0.0
	8	†						862.8		0.14	0.0
	9 [c]	†						861.4		0.14	0.0
	10 [d]	†						870.1		0.15	0.0
	11 [d,c]	†						867.0		0.16	0.0
	12	†						1862.2		0.17	0.0
	13 [c]	†						1861.8		0.17	0.0
	14 [d]	†						1862.1		0.17	0.0
	15 [d,c]	†						1861.8		0.18	0.0
SM	16	•						508.6		0.06	0.0
	17	•						2277.0		0.10	0.0
	18	•						638.7		0.09	0.0
	19	•						788.8		0.06	0.0
	20	•						2199.2		0.12	0.0
	21	•						2049.0		0.17	0.0
Diverse	cal				-			3884.0		0.14	0.0
Uniform	cal				-			5191.2		0.13	0.0
Actual	2012				-			786.5	-	-	-

Analysis batches are, •SM, †TS-CL. *Uniform cal ensemble, and Diverse cal ensemble used all projections in the table.

For the TS-CL projections the enforced calibration conditions are, [d]dynamic variables, [c]constrained (see model description for detail).



(b)

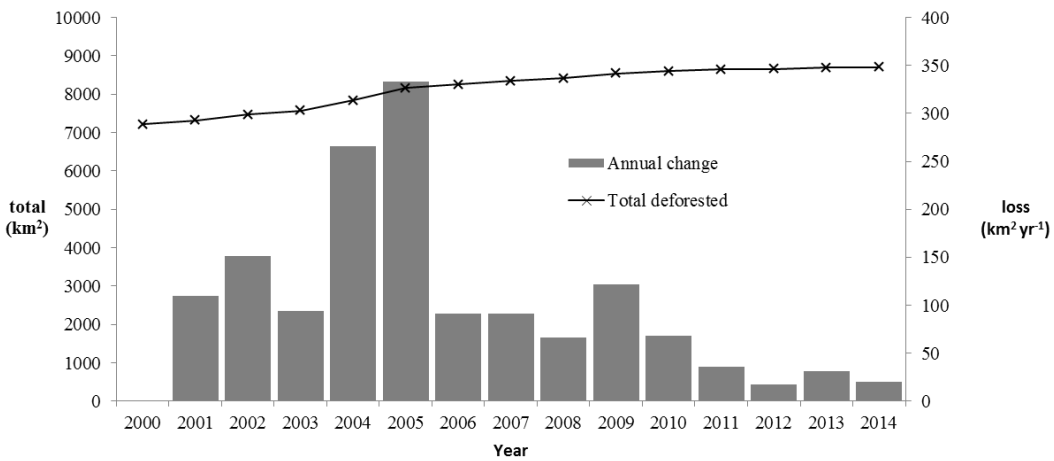
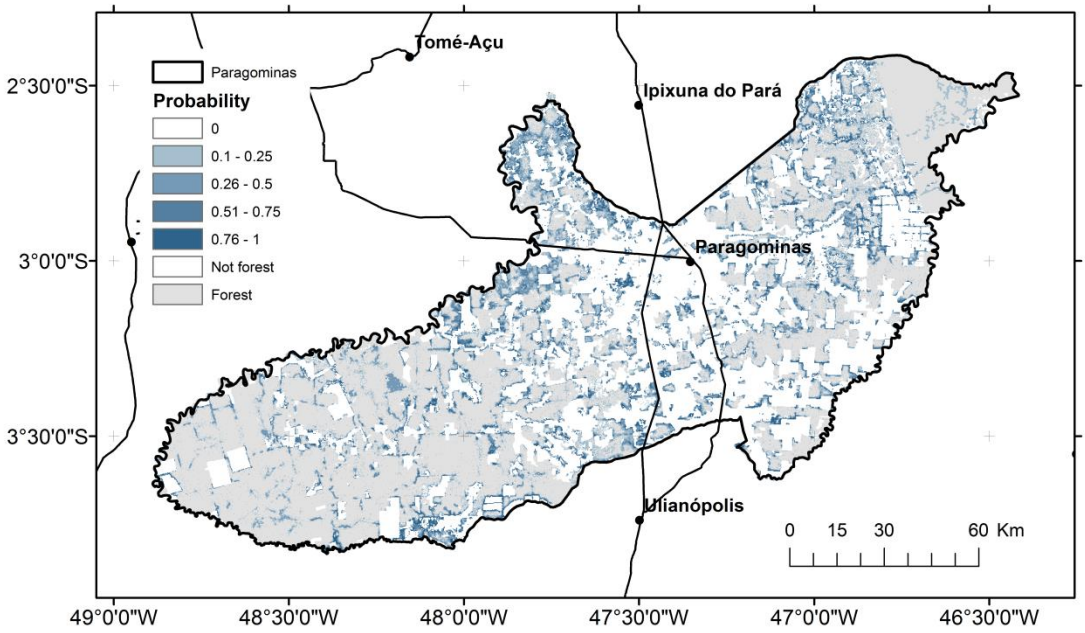


Fig. 1(a). The study area bounds outlining the municipality of Paragominas, in Brazil (location inset), with the remaining forest and the forest lost between 2004 and 2012 **(b)** Amount of deforestation and rate of change in forest for each year between 2000 and 2014 (PRODES http://www.dpi.inpe.br).

(a)



(b)

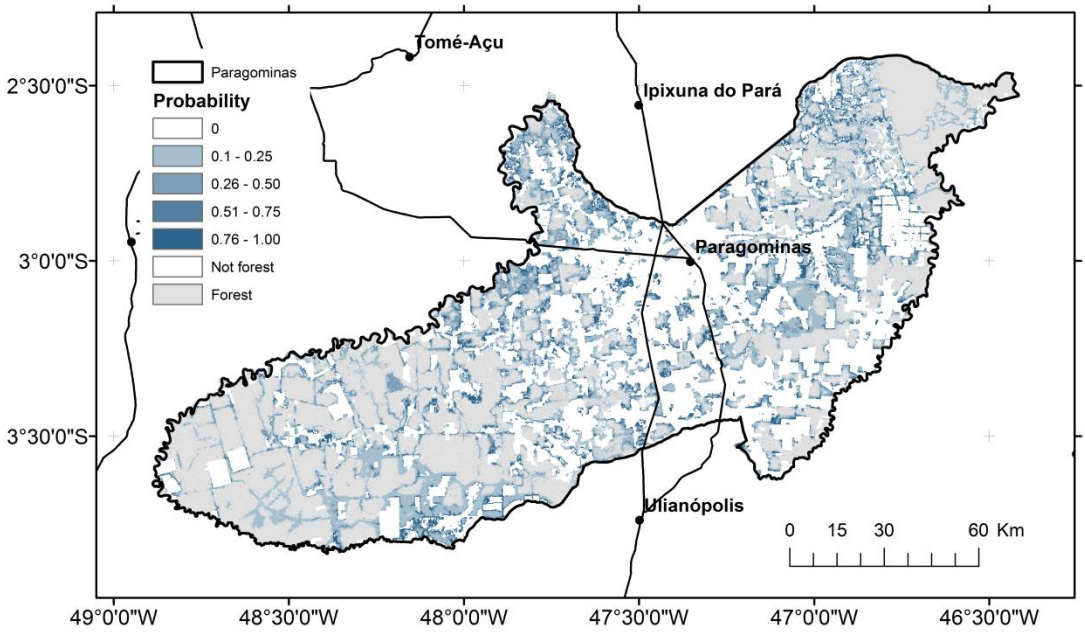


Fig. 2. The probability of deforestation using the consensus ensemble method for **(a)** Uniform cal and, **(b)** Diverse cal.

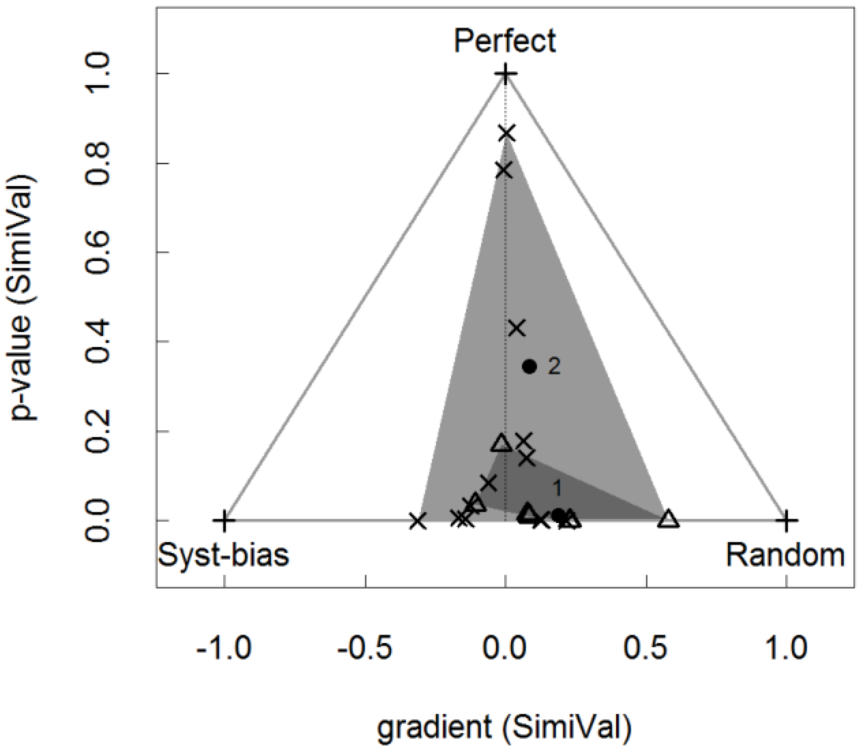


Fig. 3. The variation and dissimilarity of each model projection and the ensembles to the 'perfect' actual land-cover map 2012 (Triangles are projections 1 – 7, crosses are projections 8 - 21, and labelled dots are, Uniform cal 1, and Diverse cal 2). There are a variety of outcomes from the different model calibrations; projections that plot close to the perfect case are most similar to the actual change in terms of amount of change and landscape structure. Most of the projections plot away from the perfect and close to the vertical axis but away from corners labelled 'random' and 'systematically-biased' indicating the realism of each prediction. Uniform cal and Diverse cal fall within the cloud of their component projections (grey polygons).

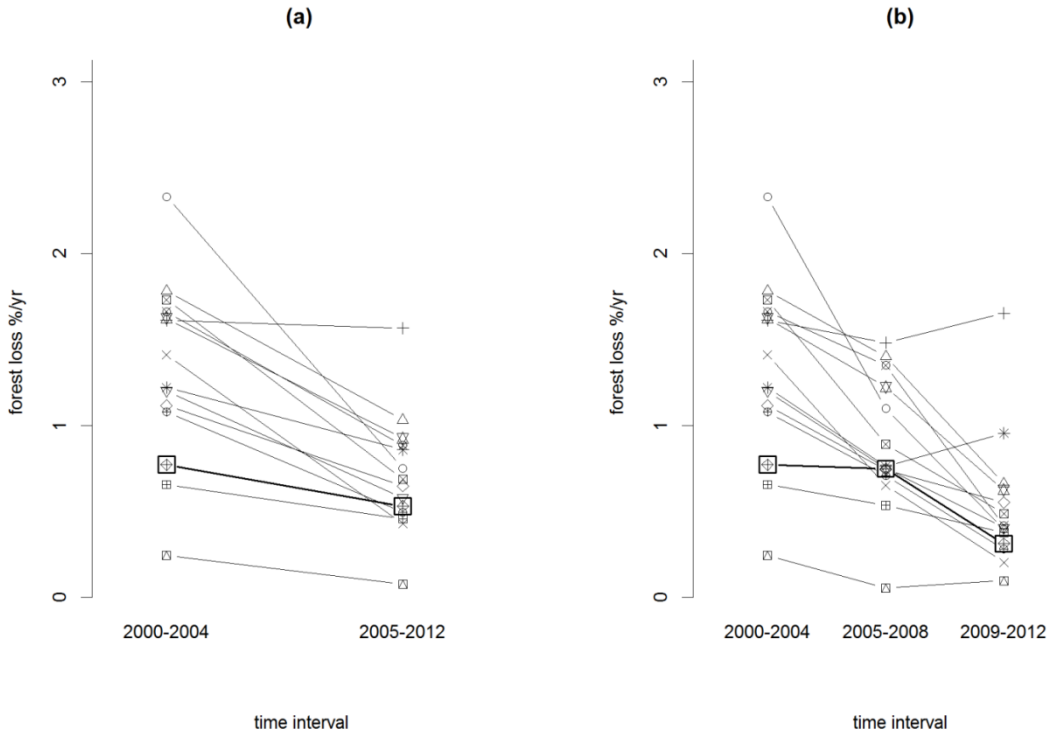


Fig. 4. Trends in aggregate forest loss rates by municipality. Paragominas is the bold square and line, all other symbols are surrounding municipalities. (a) Aggregate rates (% per year) for calibration period 2000 to 2004 against 2005 to 2012, showing a downward trend in all deforestation rates and no control condition. (b) Aggregate rates for 2000 to 2004, 2005 to 2008 and 2009 to 2012 show how control conditions may exist on shorter time scales in some municipalities (cross, star and boxed triangle).

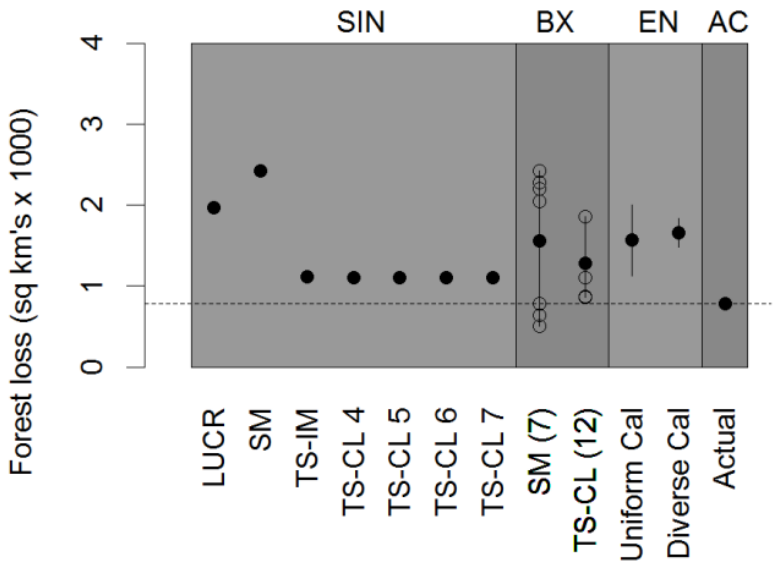


Fig. 5. Calculated forest loss estimates, and variability among models in their projections, compared to actual forest loss for 2012. **SIN;** Single projected estimates from calibrations between 2000 and 2004, **BX;** Batch projections (open circles) with the range and mean estimate for SM • and TS–CL† (symbols show projections from Table 1), **BT;** The bootstrapped ensemble *, of all seven 2000-2004 projections (Uniform cal) and the bootstrapped ensemble of all 21 calibrations (Diverse cal), and **AC;** actual forest loss. Error bars represent the range of estimates for the batch projections and 95% confidence intervals for the bootstrapped ensembles.