Global Ecology and Conservation 4 (2015) 602–613

Contents lists available at ScienceDirect

Global Ecology and Conservation

journal homepage: www.elsevier.com/locate/gecco

Original research article

A probabilistic scenario approach for developing improved Reduced Emissions from Deforestation and Degradation (REDD+) baselines

Malika Virah-Sawmy[a,b,1], Jakub Stoklosa[c,*], Johannes Ebeling[d]

[a] *Luc Hoffmann Institute, WWF International, Avenue du Mont-Blanc, 1196 Gland, Switzerland*

[b] *Institute of Environmental Studies, School of Biological, Earth and Environmental Sciences, University of New South Wales, Kensington-Sydney, NSW 2052, Australia*

[c] *School of Mathematics and Statistics and Evolution & Ecology Research Centre, University of New South Wales, Kensington-Sydney, NSW 2052, Australia*

[d] *BioCarbon Group, Friedrichstrasse 246, 10967 Berlin, Germany*

ARTICLE INFO

Article history:
Received 2 June 2015
Received in revised form 6 October 2015
Accepted 6 October 2015
Available online 11 November 2015

Keywords:
Additionality
Baselines
Carbon markets
Forecasting
Payments for ecosystem services

ABSTRACT

Performance-based payments are widely seen as a promising tool for Reduced Emissions from Deforestation and forest Degradation (REDD+) in tropical forests. Despite great advances in international REDD+ negotiations, there is a lack of consensus around the development of business-as-usual (BAU) reference scenarios or baselines to derive and quantify net carbon emission reductions. In this paper, we explore a novel approach for developing baselines (point forecasts) using exponential smoothing. Further, we introduce the concept of probabilistic BAU scenario ranges developed using this approach. We compare predictive performance with the linear trend and historical averages approaches conventionally used in policy proposals and REDD+ pilots.

We empirically test the relative performance of all three approaches by forecasting BAU baselines and scenario ranges in 36 sites (consisting of 20 countries and 8 Amazonian states with and 8 countries without REDD+ schemes). Based on two predictive performance measures (the root mean squared error and mean absolute percentage error), we find that exponential smoothing outperforms the linear trend and historical average models at predicting forest cover changes. In addition, we show how prediction intervals based on a desired confidence level generated through exponential smoothing can be used in novel ways to determine likely baseline scenario ranges. In this way it is possible to quantify the degree of variability and uncertainty in datasets. Importantly, this also provides a statistical measure of confidence to determine if REDD+ interventions have been effective.

By generating robust probabilistic baseline scenarios, exponential smoothing models can facilitate the effectiveness of REDD+ payments, support a more efficient allocation of scarce conservation resources, and improve our understanding of effective forest conservation investments, also beyond REDD+.

* Corresponding author. Tel.: +61 2 9385 7123.
E-mail addresses: malikavs@gmail.com, mvirahsawmy@wwfint.org (M. Virah-Sawmy), j.stoklosa@unsw.edu.au (J. Stoklosa), ebeling.johannes@gmail.com (J. Ebeling).

[1] Tel.: +41 22 364 9111.

http://dx.doi.org/10.1016/j.gecco.2015.10.001

1. Introduction

Mechanisms to incentivise global conservation and tackle climate change based on payment-for-performance approaches have attracted increased attention, due in large part to the perceived failure of conventional regulation in tackling cross-boundary challenges and environmental externalities (Simpson and Sedjo, 1996; Ferraro, 2001; Ferraro and Kiss, 2002; Ferraro and Simpson, 2002). To be successful, such approaches require a methodology for adequately measuring and verifying performance (Ferraro and Pattanayak, 2006), and this requires a baseline against which performance can be measured. This baseline needs to be defined in a way that is perceived as fair by relevant stakeholders, meaning it needs to be transparently, consistently, and comparably set across sites or countries to provide equitable incentives to those asked to participate (Grassi et al., 2008). Crucially, the baseline methodology needs to provide accurate prediction of what would occur under a BAU scenario (defined as the most likely scenario without implementation of an incentive system) so as to ensure that meeting or outperforming it represents additional positive outcomes (Angelsen, 2008; Olander et al., 2008).

Most prominent among the payment-for-performance systems are a variety of approaches for Reduced Emissions from Deforestation and forest Degradation (REDD+) (Ebeling and Yasué, 2008). REDD+ has become an important part of the broader international climate change debate as forest cover loss accounts for up to 15% of global greenhouse gas (GHG) emissions (Van der Werf et al., 2009). A REDD+ mechanism under the United Nations Framework Convention on Climate Change (UNFCCC), or via bilateral agreements, or though the voluntary carbon markets would reward countries, sub-national jurisdictions, or projects for reducing GHG emissions associated with forest loss (Angelsen, 2008; Olander et al., 2008). REDD+ policies require a sequence of activities including measurements of the carbon density of different types of forests, establishing a plausible BAU baseline for forest cover loss patterns and associated GHG emissions, implementing measures to reduce forest cover loss, and monitoring actual forest cover loss and GHG emissions through time. Incentive payments, for example, through the market-based sale of emission reduction units would be based on avoided GHG emissions compared to the BAU baseline (Angelsen, 2008; Ebeling and Yasué, 2008).

Determining a credible methodology for measuring and verifying performance remains one of the most significant challenges for performance-based conservation schemes such as REDD+. For example, determining a forest baseline against which actual outcomes as a result of policy interventions can be compared, is challenging due to the sheer number of interacting drivers operating across scales that affect future forest cover change. Further, new drivers or factors often create surprises in the form of abrupt and non-linear forest cover loss through time (Ghazoul et al., 2010; Angelsen et al., 2012; Sloan and Pelletier, 2012; Müller et al., 2014). A variety of options for predicting the 'most credible' baseline have been proposed and these are variably used in existing REDD+ pilot schemes. These can be roughly grouped into three types of modelling approaches for baseline setting: (i) extrapolated historical; (ii) forward looking; and (iii) adjusted historical (see below for limitations of each approach).

Currently, most project-based REDD+ mechanisms in voluntary markets, such as the Verified Carbon Standard (VCS), or negotiated multi- and bi-lateral schemes aim to set forest baselines based on extrapolated historical of BAU GHG emissions. For example, The Forest Carbon Partnership Facility Carbon Fund, the Norway–Guyana REDD+ Investment Fund, and the Norway–Amazon Fund favour the use of some sort of historical averages at either regional or global scales. The latter uses a rolling average method among states (Chagas et al., 2013) whilst the VCS uses linear regressions. Ongoing negotiations at the Conference of the Parties (COP) of the UNFCCC suggest that negotiators are leaning towards a mixture of approaches for setting REDD+ baselines (see Griscom et al., 2009 for limitations of baseline setting approaches being proposed under UNFCCC, including Compensated Reductions, Joint Research Center Proposal, Terrestrial Carbon Group, Corridor Approach, Combined Incentives, The Stock Flow approach).

All three baseline setting approaches for REDD+ have limitations. For example, the extrapolated historical methodology assumes that history is the best marker to the future (or its best approximation), and therefore simply extrapolates historical forest cover change data into the future. This applies to historical averages, or a simple linear trend, or exponential smoothing—the focus of this paper. The main issue with extrapolated historical is that it ignores dynamic non-linear interactions among drivers or cannot predict new drivers. For example, the use of linear regression models, as applied in the voluntary markets, with extrapolation to values beyond the observed time periods (also called forecasting in time series modelling) can lead to very misleading and potentially erroneous results (Wood, 2006; Moore et al., 2009). The risk in extrapolation is that the fitted linear trend may wrongly imply a linear future trend (see for example Tatem et al., 2004), and abrupt changes may occur before or within the forecasting period. This can also have severe impacts on prediction intervals associated with forecasted values (Hyndman and Athanasopoulos, 2014).

The second approach proposed for baseline setting, forward looking approach, suggests that the only way to understand future emissions is to model the drivers of land use change. Spatially explicit forward looking models have been the preferred options among those involved in REDD+ at the UNFCCC negotiations (Huettner et al., 2009). However, like extrapolated historical models, forward-looking models cannot predict new drivers nor model the complex and changing interactions among drivers (Sloan and Pelletier, 2012). Further, such models are complex and require various data inputs creating the risk that both the tool and datasets may be inconstantly used by different REDD+ countries or projects. They can also be problematic if the data requirements are too onerous or if the underlying modelling is too complicated and lacks transparency (Huettner et al., 2009). Further, empirical evidence suggest that even sophisticated spatially explicit models have limited ability to anticipate forest-cover change due to non-linear and complex interactions among drivers (Sloan and

Pelletier, 2012). The authors analysing such an approach recommended moving from sophisticated, spatially explicit to "simpler, relatively coarse scale, retrospective baselines" (Sloan and Pelletier, 2012, pp. 1).

The third approach, adjusted historical model, assumes that history is an imperfect marker for the future, and therefore adjusts historical data to improve its predictive capability. For example, UNFCCC decisions to date indicate that baselines would "take into account historic data, and adjust for national circumstances" (UNFCCC, 2009, 2013). It is unclear, however, what such adjustment would involve. One suggestion has been to set a baseline relative to a global average to accommodate countries of varying deforestation rates and reward performance relative to both national and international trends (see (Griscom et al., 2009)). However, a historical baseline would still be needed.

In this paper, we explore the use of 'exponential smoothing' as a novel approach to extrapolated historical baselines and range of scenarios. Exponential smoothing is a time-series modelling approach and is widely used in industry (*e.g.*, forecasting annual electricity consumption) and economics (*e.g.*, predicting financial returns) because of its ability to produce robust predictions (Hyndman and Athanasopoulos, 2014) with only limited amounts of data (see Section 2.2). Exponential smoothing also allows for the assignment of confidence levels around prediction intervals, which is a key advantage over reporting 'point forecasts' – defined here as a single projected value at a future point – which are often interpreted as predictions rather than projections. Such probabilistic outputs have great potential in the context of REDD+ as it gives the range of likely baseline scenarios in the absence of an incentive REDD+ mechanism under different confidence levels, using historical trends as a marker for the future. A very large range of baseline scenarios (the prediction interval for the forecast) under a specified confidence level would suggest a large amount or variability in the projected outcomes.

There has been much effort in improving the statistical foundation for the exponential smoothing model in the last 25 years (De Gooijer and Hyndman, 2006). In particular, for evaluating its performance using two techniques: (1) the use of a range of forecast accuracy measures—*e.g.*, the mean absolute percentage error (MAPE) (see review in De Gooijer and Hyndman, 2006) and; (2) the development of probabilistic forecasts, such as a forecast using a probability density function, a prediction interval, or some quantile of interest (De Gooijer and Hyndman, 2006). The development of probabilistic forecasts is a response to risk analysts, who needs to consider the many different directions in which a process may evolve whilst constraining these processes within a range of probabilities. Forestry data is often presented as a time series and data of this type can be quite non-linear due to abrupt changes from political or economic changes. Exponential smoothing methods are commonly applied to data with such characteristics; providing robust forecasts and prediction intervals (Gardner, 1985, 2006; Hyndman et al., 2008). In this paper, we (i) first assess the potential application of exponential smoothing by forecasting baselines and comparing results to more 'conventional' approaches (linear trend and average models). Secondly, we (ii) examine the role of prediction intervals in estimating the range of baseline scenarios and how this can be used as a standardised way for determining uncertainty in predictions. Lastly, (iii) using the best performing model and its prediction interval, we measure and discuss how successful new policies put into place in the Brazilian Amazon since 2004 contributed to avoided forest cover loss. To do so, we use annual time series on forest cover loss obtained from 36 sites (consisting of 28 countries: these include 20 REDD+ countries; 8 non-REDD+ countries; and 8 sub-national REDD+ regions in Brazilian Amazon). Note that we use baseline in the context of forest loss and not GHG emissions, given the scientific aims of the paper (for a review of converting forst area loss to GHG emissions, see Gibbs et al., 2007).

2. Materials and methods

2.1. Overview of historical average and simple linear trend models

Let y_t be the forest cover loss, indexed by sequential time points: $t = 1, \ldots, T$ where T is the total number observations in the data, and let x_t be the year corresponding to time point t. A typical example of these time series data is given in the following data matrix below, where we have y_t as the forest cover loss (km^2) in, say, Brazilian Amazon from 1989 to 2012:

$$\begin{pmatrix} x_t & y_t \\ 1989 & 21\,050 \\ 1990 & 17\,770 \\ \vdots & \vdots \\ 2012 & 4656 \end{pmatrix}.$$

Now, suppose we want to predict future values of forest cover loss at future time points $T+1, T+2, \ldots$. That is, we would like to predict $y_{T+1},\ y_{T+2}, \ldots$, also commonly referred to as forecasting. Then, the historical average model takes the average for each observed y_t, that is, $\bar{y} = \sum_{t=1}^{T} y_t$ is used as future prediction values for each $y_{T+1},\ y_{T+2}, \ldots$.

The simple linear trend model uses x_t as predictors in a regression model:

$$y_t = \beta_0 + x_t\beta_1 + \epsilon_t$$

where β_0 is the intercept parameter, β_1 is the slope parameter and ϵ_t are independent and normally distributed errors. The model parameters β_0 and β_1 can be estimated by using least squares, and future predicted values for $y_{T+1},\ y_{T+2}, \ldots$ are given as $\hat{y}_{T+1} = \hat{\beta}_0 + x_{T+1}\,\hat{\beta}_1, \hat{y}_{T+2} = \hat{\beta}_0 + x_{T+2}\,\hat{\beta}_1$, etc. where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the regression model estimates.

Prediction intervals for the historical average and simple linear trend model can also be easily obtained (see Section 10.1 of Moore et al., 2009).

2.2. Exponential smoothing: methodology, application and potential for REDD+

Exponential smoothing is a forecasting technique that is applied to time series data (Gardner, 1985, 2006; De Gooijer and Hyndman, 2006; Corberán-Vallet et al., 2011). In contrast to the historical average model, which is commonly used for REDD+ forecasts, where the past observations are weighted equally, exponential smoothing assigns exponentially decreasing weights to observations further in the past (see Hyndman et al., 2008; Hyndman and Athanasopoulos, 2014). As discussed below, a collection of exponential smoothing models are available. To illustrate a simple example, we use same notation as in Section 2.1 and follow Hyndman and Athanasopoulos (2014). Again, suppose we want to predict future values for each $y_{T+1},\ y_{T+2}, \ldots$. For the exponential smoothing model, we now write:

$$\hat{y}_{T+1} = \alpha y_T + \alpha\left(1-\alpha\right) y_{T-1} + \alpha\left(1-\alpha\right)^2 y_{T-2} + \cdots$$

where the smoothing parameter $0 \le \alpha \le 1$ is introduced in the model to control for the rate at which the weights decrease. For large α more weight is placed to more recently observed y_t, and for small α more weight would be placed for points in the distant past. The value for α can be set a priori to model fitting; however in most cases α is estimated in a robust and objective manner using the observed data, for example, using the sum of squared errors (see Section 7.1 of Hyndman and Athanasopoulos, 2014).

The above technique was first suggested by Robert Goodell Brown in 1956, and then expanded by Charles C. Holt in 1957 (De Gooijer and Hyndman, 2006). Brown first developed the simplest form of exponential smoothing, also known as an exponentially weighted moving average. Since then, there has been a lot of progress in developing more complex models to capture the underlying structure in the data. For example, trend and/or seasonality components can be added in the forecasting function (see Section 7.6 of Hyndman and Athanasopoulos, 2014). A trend is usually added if there is a long term trend in the data, of which there are four different types to consider: additive, damped additive, multiplicative, and damped multiplicative. A seasonal component is also usually added if the data is influenced by seasonal factors, of which there are two types: additive and multiplicative. Further details and full descriptions of the statistical formulae for the 'taxonomy' of exponential smoothing models are given in Table 7.8 of Hyndman and Athanasopoulos (2014) or Table 1 of Gardner (2006).

The analysts can potentially fit any of these models subjectively; however it is more common to perform a model selection procedure using the observed data, and thus obtain the best model. A common model selection approach that is routinely applied in practice is Akaike's Information Criterion (AIC, Burnham and Anderson, 1999); it has excellent properties when predicting or forecasting on new data (Shibata, 1980; Hyndman et al., 2008).

The statistical package *forecast* in the R statistical software (Hyndman and Athanasopoulos, 2014; R: Development Core Team, 2015) has been designed to fit exponential smoothing models and subsequently give forecasts with prediction intervals. The *forecast* R-package estimates the smoothing parameter(s) α, and uses AIC for model selection to select the exponential smoothing model that best fits the data, starting from the simplest model (*e.g.*, no trend or seasonal components), to more complex models (*e.g.*, multiplicative trend and additive)—both procedures are built-in operations in the *forecast* R-package. The historical average model is also considered as a candidate model in the model selection procedure within *forecast*. If no clear structure is evident in the data and no weighting is considered, then the historical average model may give the lowest AIC; in this case *forecast* will select the historical average model as the final model. This R-package facilitates application by non-experts; furthermore prediction intervals can be easily generated, see Section 2.7 of Hyndman and Athanasopoulos (2014).

Using the statistical package *forecast* in the R statistical software, we compared the performance of exponential smoothing with two main conventional approaches used in REDD+, which includes the linear trend and average historical model types.

2.3. Data

We used data for 28 countries (see Table 1), 20 of these countries are participating in one or more REDD+ initiatives (UN-REDD, UNFCCC-REDD, Forest Carbon Partnership Facility Carbon Fund, Guyana–Norway bilateral etc.). For all of these REDD+ countries, data was available on forest cover loss from 2001–2013 from the Global Forest Watch database (http://www.globalforestwatch.org/countries). In addition, we also included 8 non-REDD+ countries selected from the same database; there was no specific reason why these highly deforested countries were chosen other than they each have over 1.8 Mha of total tree cover loss from 2001 to 2013 (http://www.globalforestwatch.org/countries/overview). These data are based on the analysis of a collection of Landsat satellite images by Hansen et al. (2013), such that they capture natural disturbances which are inherently stochastic, and therefore can introduce large variation in measurement of performance. However, we do note that the use of the Hansen et al. (2013) data at local to regional scale is problematic due to the sheer complexity of fine scale validation issues of a global model. However, so far, this is the only dataset available with enough temporal points.

For these 28 countries, we forecast baselines for each model type for 2010–2013 (*i.e.*, 4 years of forecasting). For this forecast, the first 9 points of observed data (2001–2009) were our *training data* and we validated each model with the *test data* (2010–2013), see Section 3.2 for further details. We chose 4 years of forecasting as this was approximately 31% of the available data. The sample size of the training data considered here is small, although this is generally not an issue when using the *forecast* R-package; for example, see case studies in Section 7.1 of Hyndman and Athanasopoulos (2014). For small

Table 1
Assessment of the predictive performance for the three model types using root mean squared error (RMSE) and mean absolute percentage error (MAPE) for 28 countries. Forecasting is conducted for $h = 4$ years (from 2010–2013). Values in bold indicate the lowest RMSE/MAPE. Note that the exponential smoothing model gave better overall forecasts.

Country	RMSE			MAPE		
REDD+	Lin. trend	Hist. ave.	Exp. Smooth.	Lin. trend	Hist. ave.	Exp. Smooth.
Belize	**9 520**	9 780	9 780	**36**	38	38
Bolivia	**122 270**	136 338	136 338	42	**34**	**34**
Brazil	**532 885**	818 185	625 338	**15**	35	19
Cambodia	**80 438**	130 197	86 231	**36**	64	41
Cameroon	**18 377**	23 083	20 703	**32**	36	34
Cen. African Rep.	**10 630**	13 780	13 781	**19**	25	25
Colombia	85 425	**47 375**	**47 375**	49	**24**	**24**
Costa Rica	11 525	**5 054**	**5 054**	129	**55**	**55**
Dem. Rep. of the Con.	**197 190**	260 054	260 055	**24**	29	29
Ecuador	21 680	**15 365**	16 314	50	**19**	37
Guyana	**3 223**	3 541	3 541	**25**	26	26
Indonesia	739 086	549 980	**509 061**	51	**23**	36
Madagascar	**59 414**	88 572	88 572	30	**27**	**27**
Malaysia	156 314	142 761	**111 766**	34	24	**22**
Mexico	91 617	**17 418**	65 144	52	**8**	37
Nigeria	21 873	16 812	**16 148**	49	34	**32**
Papua New Guinea	**11 872**	13 961	13 961	19	**17**	**17**
Paraguay	**96 610**	173 890	173 890	**21**	34	34
Suriname	**5 985**	7 656	6 029	**27**	43	30
Vietnam	34 091	72 411	**27 782**	22	44	**17**
Mean	115 501	127 311	**111 843**	38	32	**31**
Median	46 752	**35 229**	37 578	33	**31**	**31**
Non-REDD+	Lin. trend	Hist. ave.	Exp. Smooth.	Lin. trend	Hist. ave.	Exp. Smooth.
Argentina	171 896	119 956	**119 955**	38	**23**	**23**
Australia	**69 124**	84 407	84 408	**41**	50	50
Canada	**318 193**	352 900	352 900	**9**	10	10
China	365 622	127 679	**122 499**	68	22	**18**
Finland	58 553	**13 821**	**13 821**	37	**7**	**7**
Russia	**1 862 002**	1 820 665	1 820 664	38	**38**	**38**
Sweden	80 905	**32 801**	**32 801**	34	**12**	**12**
United States	494 346	**431 663**	**431 663**	26	**22**	**22**
Mean (includes REDD+)	204 667	197 504	**186 271**	38	29	**28**
Median (includes REDD+)	80 671	78 409	**74 776**	35	**26**	28

sample sizes, a simpler exponential smoothing model will typically be selected. This does not mean the analysis is either unreliable or misleading; rather the resulting forecasts will be less informative (Hyndman and Athanasopoulos, 2014).

We performed the same exercise at a sub-national scale for the Brazilian Amazon and at a regional scale for 8 states within the Brazilian Amazon (see Table 2). For cover change data from INPE (Instituto Nacional de Pesquisas Espaciais) have been generated using high resolution images which can potentially give more accurate annual historical data (1988–2012) (Olander et al., 2008; INPE, 2013) than the Hansen et al. (2013) dataset. For both the Brazilian Amazon and its 8 states, we forecast baselines for 5 years from 1998 to 2002, using 1988–1997 as training data. We exclude data after 2004 since REDD+ activities (defined here as the introduction of new significant forest conservation activities) were implemented at this time.

2.4. Validation

We used two well-known statistical measures to assess the model's forecast performance, both of which require training and test data. Suppose that our data consists of $t = 1, \ldots, n$ time series observations, we first partition the data into a training dataset $(t = 1, \ldots, T)$ and a test dataset $(t = T + 1, \ldots, n)$, thus $h = n - T$ is the number of future time points that we consider for forecasting. The root mean squared error (RMSE) is given by: $\text{RMSE} = \sqrt{(1/h) \sum_{i=1}^{h} (\hat{y}_{T+i} - y_{T+i})^2}$ and the mean absolute percentage error (MAPE) is given by: $\text{MAPE} = (1/h) \sum_{i=1}^{h} |100 \times \{(\hat{y}_{T+i} - y_{T+i})/y_{T+i}\}|$ where $\|$ denotes the absolute value. Both measures quantify the difference between observed (our test data) and forecasted values—*e.g.*, the RMSE is calculated by taking the deviation (also called prediction error) from the predicted and the observed, squaring the deviation and then taking the mean (followed by a square root). In our context, the MAPE is a more appropriate measure to use over the RMSE, since it allows comparisons to be made on the same scale. For example, we can compare model performance for countries with large forest areas, such as Russia, with smaller countries like Guyana because the MAPE is a calculated as a scaled percentage.

Table 2
Assessment of the predictive performance for the three model types using root mean squared error (RMSE) and mean absolute percentage error (MAPE) at state levels within the Brazilian Amazonia and for the whole of the Brazilian Amazon. Forecasting is conducted for $h = 5$ years (from 1998 to 2002). Values in bold indicate the lowest RMSE/MAPE. Note that once again the exponential smoothing model gave better overall forecasts.

5 years from 1998–2002	RMSE			MAPE		
Brazilian Amazon state	Lin. trend	Hist. ave.	Exp. Smooth	Lin. trend	Hist. ave.	Exp. Smooth.
Acre	**164**	168	168	22	19	**19**
Amazonas	**138**	260	260	**20**	37	37
Maranhão	891	**91**	**91**	82	**7**	**7**
Mato Grosso	**944**	1488	1488	**13**	18	18
Pará	1494	**1265**	1266	17	**13**	**13**
Rondônia	889	**410**	**410**	36	**12**	**12**
Roraima	149	104	**86**	51	66	**44**
Tocantins	377	333	**146**	143	140	**33**
Brazilian Amazon (all states)	**1656**	2395	2397	**5**	9	9
Mean	745	724	**701**	43	36	**21**
Median	889	333	**260**	22	**18**	**18**

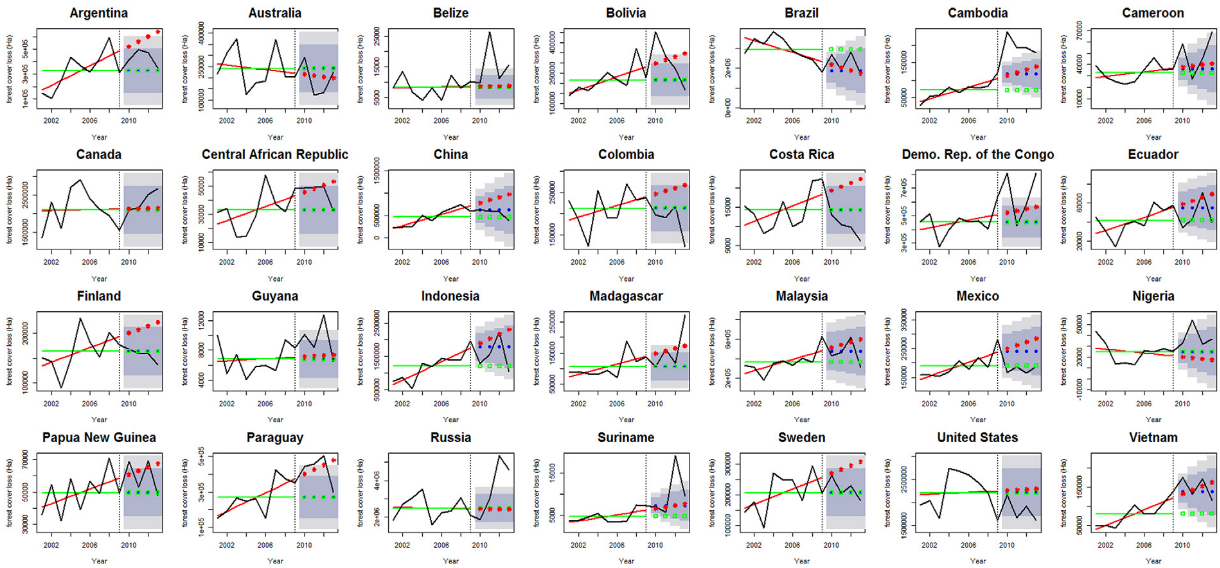

Fig. 1. Forecasting baselines for 28 countries for $h = 4$ consecutive years (from 2010 to 2013) using 2001–2009 as (training) historical data. The solid black line represents the observed data, the red line and red stars represent the fitted model and point forecasts for the linear trend model respectively, the green line and green squares represent the fitted model and point forecasts for the historical average model respectively, and the blue dots represent the point forecasts for the exponential smoothing model. The dark grey area represents the 80% prediction interval, while the light grey is the 95% prediction interval for the exponential smoothing model. Plots were generated using the forecast R-package. Note that, exponential smoothing, followed by the historical average were the best performing models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

When taking overall RMSE/MAPE measures across all countries considered, we found that for some influential countries several RMSE/MAPE values were either much smaller or much larger for some model types. These outliers can be misleading as they skew the overall measure; therefore in addition to the mean we also included the median as a more robust measure to account for such outliers in the analysis. In addition, we also constructed prediction intervals at the 95% and 80% confidence levels using the exponential smoothing model (see for example, Fig. 1); these confidence levels are often used in practice when forecasting.

2.5. Effectiveness of REDD+ policies in the Brazilian Amazon

Finally, based on the best performing model, we estimated the impact of the implementation of REDD+ policies that were put into place as from 2004 on forest conservation outcomes in Brazilian Amazon. We selected 2004 as a "simple" cut-off date to differentiate REDD and non-REDD activities, recognising that forest conversation activities in Brazil date to a much longer period, since in 2004, Brazil implemented (for the first time) a real-time system for Detection of Deforestation (DETER) to identify forest cover loss hotspots and alert authorities to intervene (Assunção et al., 2012). This was followed by institutional changes to enhance law enforcement through the use of legal instruments for the punishment of environmental crimes in the subsequent years (Assunção et al., 2012), as well as range of other policies (see later). We discuss the effectiveness of the policies at avoiding forest cover loss at sub-national and state levels based on the reduction achieved from the "most likely" BAU baseline and whether these reductions were located outside the 80% prediction interval generated as the BAU range of baseline scenarios.

3. Results

3.1. Best performing model for point forecasts

Among the 20 REDD+ countries analysed, the exponential smoothing model gave the lowest mean RMSE and mean/median MAPE values, whereas the historical trend model gave the lowest median RMSE (Table 1). Note that the historical average model was selected for several datasets in the analysis (*e.g.*, Bolivia, Colombia, Madagascar and Papua New Guinea) yielding either exact or very similar RMSE and MAPE values as given by the exponential smoothing model (Table 1, Fig. 1).

For the full set of 28 countries (REDD+ and non-REDD+ countries), once again the exponential smoothing model gave better overall predictions (lowest mean/median RMSE and mean MAPE values) compared with the other model types (Table 1, Fig. 1). The historical average model gave the lowest median MAPE, whereas the linear trend model was by far the poorest performing model type. Note that we are measuring the overall predictive performance across all countries considered rather than specific countries—*i.e.*, we consider model performance based on means and medians across all countries. Our reasoning for this is that some countries perform similarly across model type (*e.g.*, Belize, Guyana, Canada and Russia; Table 1), and so these differences are only marginally distinguishable. Also, for several countries, some model types were clearly superior compared to the others (*e.g.*, Indonesia and Mexico for the historical average model; Table 1), on the other hand some model types performed very poorly compared to others (*e.g.*, Costa Rica and China for the linear trend; Table 1). When examining results for specific countries, we note that exponential smoothing models was only considerably marginally better than the best performing model in only 5 of 28 countries—*i.e.*, the MAPE for the exponential smoothing model differed by more than 10% from the best performing model. These 5 countries were Australia, Ecuador, Indonesia, Mexico and Paraguay; see Table 1.

At the sub-national scale in Brazilian Amazon, exponential smoothing, followed by the historical average model, also gave better predictions for the $h = 5$ years of forecasting baselines (1998–2002) prior to policy change (Table 2, Fig. 2). Critically, the RMSE values for the exponential and historical average models were comparatively small, thereby indicating closer location proximity between forecasted and observed values (Table 2).

3.2. Reliability of prediction intervals generated from the best performing model (exponential smoothing)

Using the test data, we verified the reliability of the prediction interval (defined as the range of baseline scenarios) generated from the training data for the 28 countries analysed. We observed that approximately 68% of the validation points were located within the 80% prediction interval of the exponential smoothing model, approximately 15% within the 80%–95% prediction interval, with the remaining 17% outside this range (Fig. 1). At the sub-national scale for Brazilian Amazon, more reliable results were achieved: with 90% of the validation points being within the 80% prediction interval and 10% within the 80%–95% prediction interval (Figs. 2–3).

3.3. Effectiveness of policies in Brazilian Amazon according to best performing model

Our estimation of the effectiveness of REDD+ policies since 2004 within the 8 Brazilian Amazon states is based on forest loss since 2004 and beyond, and is compared with the baselines generated from the best performing model for the period of 2004 and beyond. The results indicate that most Amazonian states had achieved a positive forest conservation outcome as observed forest loss was located below the BAU baseline generated from exponential smoothing or historical averages, but that only Acre, Mato Grosso, and Para managed to reduce forest cover loss outside of the 80% prediction interval (defined as the range of baseline scenarios) given by the exponential smoothing model (see Fig. 3).

4. Discussion

Our analysis shows that among the historical extrapolated approaches, exponential smoothing models (overall) outperformed other model types in the 36 case studies considered in our empirical study (Tables 1 and 2). This is a 'comparative analysis', and thus we do not make any claim for decision-makers as to what is an acceptable level of prediction error for a best performing model. Exponential smoothing is instead evaluated comparatively from the overall mean (and median) RMSE and MAPE results against two standard methods (all based on extrapolated historical approaches) used in REDD+ schemes. We examined these model forecast performances on both national and sub-national levels, and forecasted on both short (*e.g.*, $h = 5$) to moderate (*e.g.*, $h = 8$) annual time periods. To visually compare and summarise the accuracy of our results, we plot the baseline point forecasts (*i.e.*, MAPE values) for each country and each Brazilian Amazon state in Fig. 4(a).

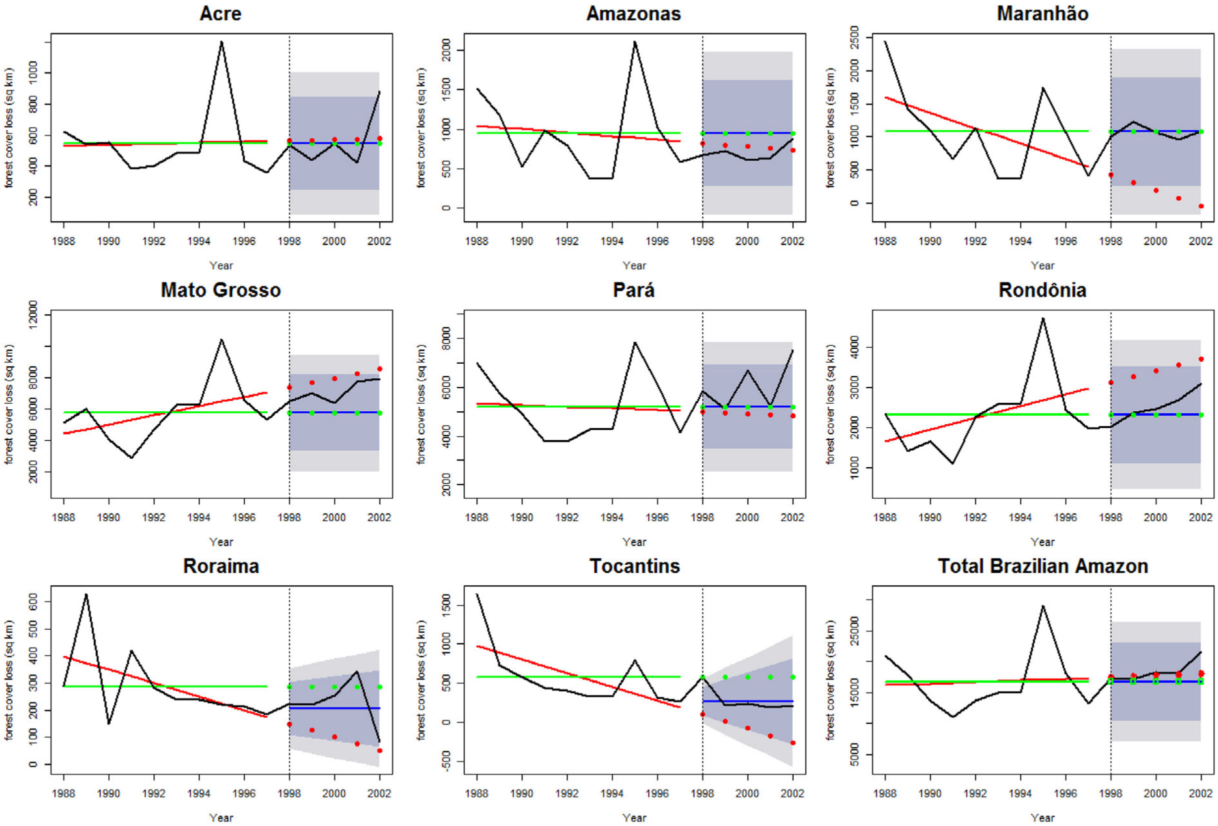

Fig. 2. Forecasting baselines on a sub-national and regional scale for 8 states within the Brazilian Amazon and for the whole of the Brazilian Amazon for $h = 5$ consecutive years (from 1998 to 2002) in the period prior to REDD+ policy implementation. We use 1988–1997 as historical (training) data. The solid black line represents the observed data, the red line and red stars represent the fitted model and point forecasts for the linear trend model respectively, the green line and green squares represent the fitted model and point forecasts for the historical average model respectively, and the blue dots represent the point forecasts for the exponential smoothing model. The dark grey area represents the 80% prediction interval for the exponential smoothing model, while the light grey is the 95% prediction interval. Note that, exponential smoothing, followed by the historical average were the best performing models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Unlike the historical average model which under the UNFCCC requires countries to potentially select the length of the sequence, exponential smoothing uses the entire historical data and assigns decreasing weights to observations further in the past (see Hyndman et al., 2008; Hyndman and Athanasopoulos, 2014). This removes any arbitrariness to cherry-picking of historical data periods, and allows for more recent, and potentially more informative events to have a greater influence on the projection. The greater accuracy of exponential smoothing thus make these models preferable for forecasting short-term baselines, or at best, a historical average model which is still a subset of the exponential smoothing model, see Section 2.2.

We note however two important challenges in our datasets that could result in large RMSE or MAPE (prediction error) values for all model types. Firstly, we assume that the datasets for all the 28 countries provided by Hansen et al. (2013) via the Global Forest Watch database were comparable and consistent. However, validation of annual allocation of change by Hansen et al. (2013) showed that this was variable among biomes. As a result, we do not know the degree to which some of the large RMSE or MAPE values could be due to data quality issues. However, the INPE dataset is probably of better quality being extracted from high resolution satellite images (Olander et al., 2008), and here we find, that the exponential smoothing still worked best overall. Secondly, model validation is based on the flawed assumption that none of these countries, except from 2004 for the Brazilian Amazon, had put into place significant forest conservation activities (REDD+ policy) in the period for which we used the training or validation data. Implementation of important forest conservation policies that were "game-changes" in forest conservation would impact our prediction errors (MAPE or RMSE) between forecast and actual outcomes. For example, low forest cover loss in the period after 2007 are apparent in four countries (Bolivia, Colombia, Costa Rica and Mexico) which may be due to important forest conservation policies in this period.

4.1. Usefulness of prediction intervals

Hyndman and Athanasopoulos (2014) lamented that "many decision making processes today cannot yet take probabilistic inputs, so the most commonly used forecasting output form is still point forecast". The use of probabilistic

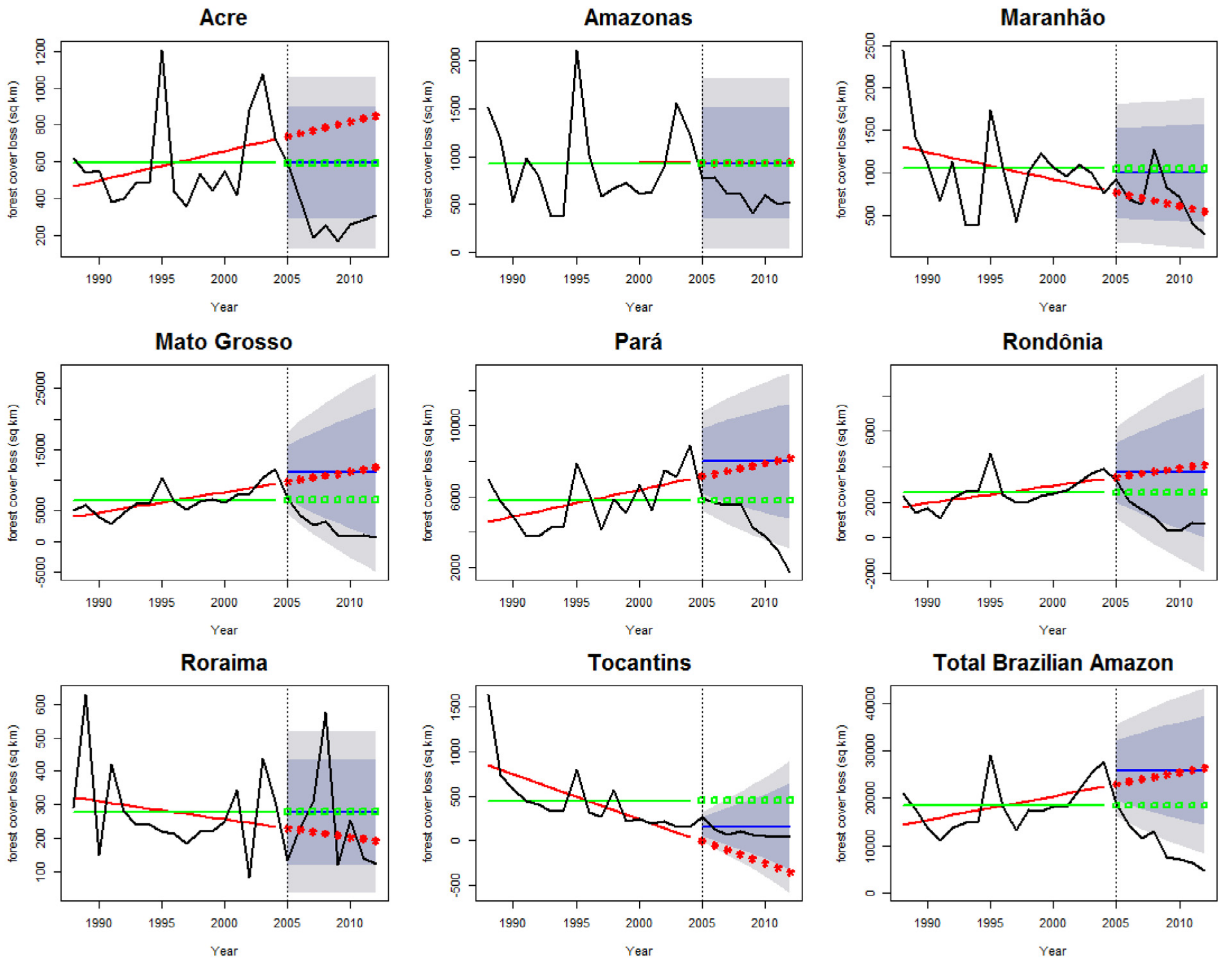

Fig. 3. Evaluating effectiveness of REDD+ policies for 8 states within Brazilian Amazon and for the whole of the Brazilian Amazon. We forecast baselines for $h = 8$ consecutive years (from 2005 to 2012) using 1988–2004 historical (training) data for the linear trend and historical average models. The solid black line represents the observed data, the red line and red dots represent the fitted model and point forecasts for the linear trend model respectively, the green line and green dots represent the fitted model and point forecasts for the historical average model respectively, and the blue dots represent the point forecasts for the exponential smoothing model. The dark grey area represents the 80% prediction interval for the exponential smoothing model, while the light grey is the 95% prediction interval. The results indicate that most states achieved below the historical averages or the exponential smoothing forecasts. In particular, the states of Acre, Mato Grosso and Para achieved below the 80% prediction interval. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

outputs (or scenarios) is critical in the context of REDD+ due to the high level of surprise in forest cover loss, being characterised by rapid rates of change, abrupt changes, complex feedbacks, and new drivers (Ghazoul et al., 2010; Angelsen et al., 2012; Sloan and Pelletier, 2012; Müller et al., 2014). Indeed, there have already been several calls for probabilistic projections in the context of REDD+ already (*cf.* Kindermann et al., 2006; Grassi et al., 2008; Sloan and Pelletier, 2012). In our study, we found the probabilistic scenario range estimation from the prediction interval to be a useful guide whether the validation data fell within this range. For example, at the country level ($n = 28$), we found that 83% of the validation data were within the 80%–95% prediction intervals whilst increasing to 100% for state levels ($n = 8$).

Prediction intervals based on a desired confidence level generated from the exponential smoothing model can be used in novel ways to determine the likely baseline scenario ranges. These inform on the degree of variability and uncertainty in datasets. Inversely, they provide a statistical measure of confidence if REDD+ interventions have been effective. For example, in Fig. 4(b) we compared the range of baseline scenarios for each individual country and Brazilian Amazon state—these were obtained by standardising each dataset (*i.e.*, subtract the mean and divide by the standard deviation), constructing (scaled) prediction intervals, and then calculating averages of the range of each interval across each future time point. The comparison indicated high variability in the Maranhão and Tocantins dataset; however we note that there is additional model uncertainty since different exponential smoothing model types can be selected for each country, see Section 2.2.

A large prediction level at any confidence level (say 95%, 80%, etc.) for the baseline generated indicates large uncertainty in anticipating forest-cover change. In such cases, it is also likely that other approaches such as spatially explicit forward looking models will also have limited ability to anticipate forest-cover change (Sloan and Pelletier, 2012). Alternatively, for

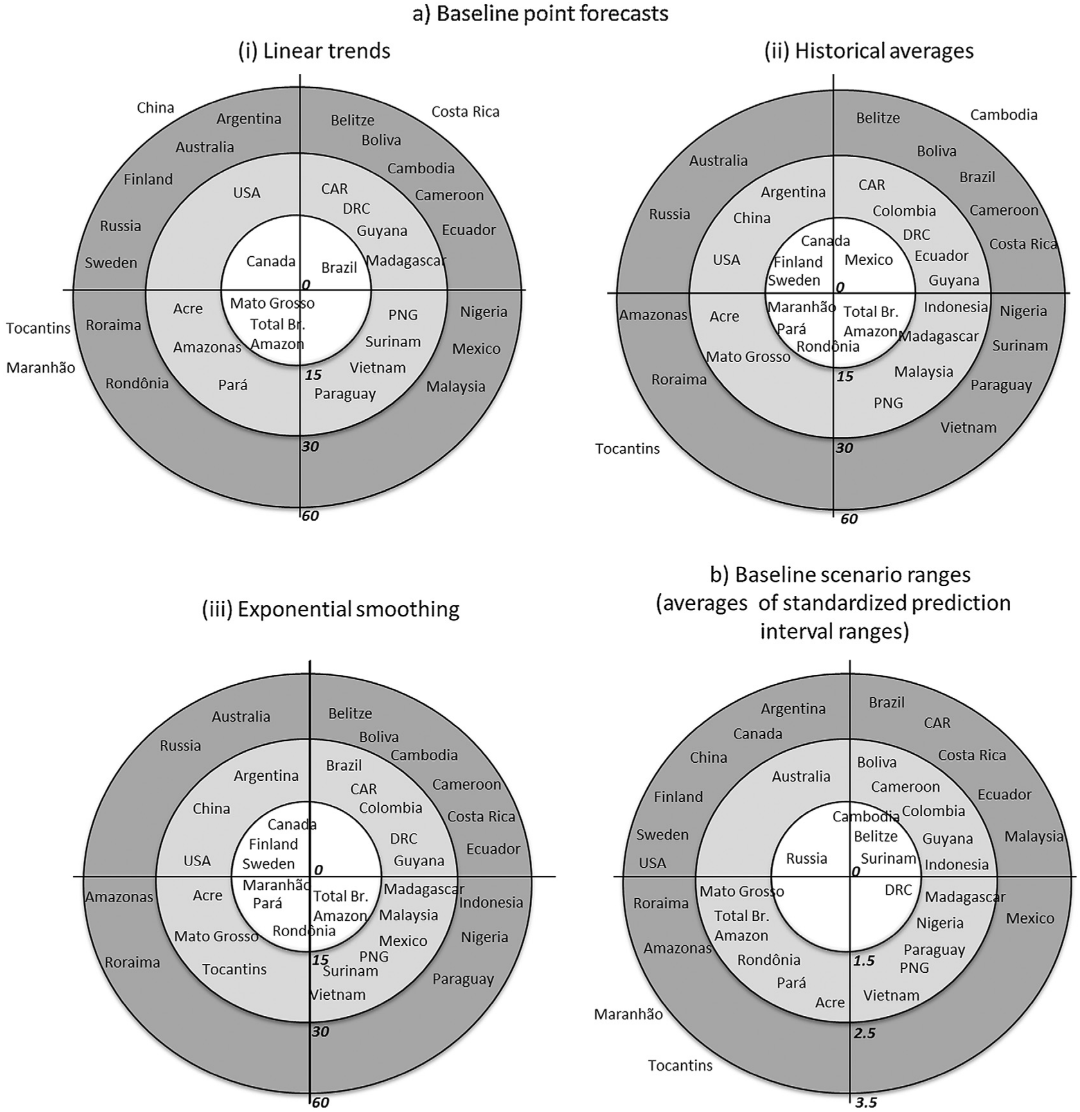

Fig. 4. A visual comparison and summary of the accuracy of our results. We plot: (a) the predicted baseline point forecasts (i.e., MAPE values); and (b) range of baseline projection scenarios using prediction intervals for each individual country and Brazilian Amazon state—these were obtained by standardising each dataset (i.e., subtract the mean and divide by the standard deviation), constructing the (scaled) prediction intervals, and then calculating averages of the range of each interval across each future annual time point. The diagram indicates higher overall accuracy for the exponential smoothing. Further, the prediction intervals indicate large variation for the Maranhão and Tocantins dataset.

datasets with smaller prediction intervals due to historically low forest cover loss—*e.g.*, Belize, achieving a REDD+ outcome outside the 80% prediction interval could not only be feasible, but if achieved it would indicate that we can be 80% confident that a (true) reduction of forest cover loss has been achieved outside the range of baseline scenarios (see Fig. 3). It would be interesting to explore whether the use of premiums achieved below the point forecast baseline and outside the range of baseline scenario (prediction interval) could be potentially a fair reward and incentive for countries or projects with historically low deforestation.

Importantly, when using either historical average or linear trend models on time series data, prediction intervals can be similarly obtained (see Section 10.1 of (Moore et al., 2009)). However, it is well-known (both theoretically and empirically) that prediction intervals from linear regression models on time series data can be biased due to temporal autocorrelation—*i.e.*, if the influence of previous forest cover change values in the current time point is strong, prediction intervals from linear regression models can be much broader than they should be (Hyndman and Athanasopoulos, 2014). This occurs because historical average and linear trend models are constructed under the assumption that observations are independent of one another; in the case of forest cover change, these values are not independent. In contrast, the exponential smoothing model (and the associated prediction interval) appropriately accounts for this dependence in the model structure, providing more reliable and correct prediction intervals.

4.2. Effectiveness of REDD+ policies in the Brazilian Amazon

In the Brazilian the Amazon, a reduction in forest cover loss was achieved below the forecasted baseline generated (Fig. 3). In addition, these reductions were located outside the 80% prediction interval of the range of baseline scenarios from the first year of the implementation of the policy in 2005 (Fig. 3). Thus we can be 80% confident that this represents additional positive outcomes compared to the range of baseline scenarios. It is indeed generally known that Brazil recognised the complexity of deforestation and put into place since 2004 a combination of incentives for enhancing forest conservation, disincentives to prevent land-use change, and enabling conditions through cross-sectoral policy reform to reshape forest use (Kissinger, 2015). For example, Brazil established the Action Plan to Prevent and Control Deforestation in the Amazon in 2004, in order to control illegal activities, and identify solutions for regulation and monitoring (Assunção et al., 2012). Later in 2006, Brazil set a ban on the commercialisation of soy grown in the Amazon, which was set to expire in 2013 but has since been renewed, (Gibbs et al., 2015)). Further, during the same time, the Bank of Brazil, as a disincentive measure, refused agricultural credit for soy farmers who planned cultivation in newly cleared forest (Kissinger, 2015). In parallel, intensification practices through the use of double cropping in soy croplands and pastures put into place could have encouraged investment in deforested lands rather than promoting new deforestation (DeFries et al., 2013).

At the Brazilian Amazon state level, reductions in forest cover loss below the baseline levels were also achieved for most states. However, these reductions were located outside the 80% prediction interval only in Acre, Mato Grosso and Pará (Fig. 3), which we assume here relates to far better forest conservation activities there. Our objective here is more of a case example to demonstrate that the prediction intervals can serve as a strong guide both in understanding uncertainty, or vice versa, confidence in forest conservation outcomes located not only below the baselines but also outside the range of baseline scenarios (prediction interval).

5. Conclusion

In this paper, we have demonstrated the value of introducing exponential smoothing for setting REDD+ baselines and developing ranges of baseline scenarios. Moreover, we argue that this method is broadly applicable to any conservation performance scheme and for assessing the effectiveness of conservation inteventions. Exponential smoothing provides a baseline methodology that can be applied transparently, consistently, and comparably across sites. In contrast to commonly used alternative approaches, this can be done irrespective of chosen reference time periods. The method also accounts for uncertainty through the use of prediction intervals and generates comparatively more robust results. Exponential smoothing can ensure that meeting or outperforming baselines represents real additional positive outcomes of a REDD+ intervention. The proposed approach allows conservation actors to acknowledge the substantial uncertainties inherent in forest cover change predictions and focus on scenario ranges rather than point trajectories. Perhaps most importantly, being able to use this improved approach for comparing actual impacts of given policies and interventions can support the design of more effective and efficient conservation policies and REDD+ mechanisms.

We propose that future scientific work should incorporate further types of spatial information (*e.g.*, generating spatial datasets constrained to counterfactual areas or areas with similar conservation approaches, both in intensity and timing) and further investigate the predictive performance of exponential smoothing among those. Although exponential smoothing is known to be quite robust (Hyndman and Athanasopoulos, 2014) it would be interesting to examine how sensitive it is to break years between model training and validation periods. Finally, we encourage researchers to use the *forecast* R-package; this software is user friendly and point forecasts along with prediction intervals are easily obtainable with minimal computer coding required. All materials presented in this manuscript are reproducible and R-code is available upon request.

Acknowledgements

We would like to thank the associate editor and three anonymous referees for their helpful comments. We would also like to thank Dr. Joshua Tewksbury for his valuable comments to an earlier draft, and Professor William T.M. Dunsmuir for a helpful discussion on forecasting with time series models. The first author would like to warmly thank Rhett Butler for the support in finding the relevant databases for our analysis and for doing an excellent job encouraging scientists to make their data publicly available. Further thanks go to Nancy Harris and James Anderson from the World Resource Institute for their help in giving us access to the Global Forest Watch database (http://www.globalforestwatch.org/countries), which is now available to the public.

References

Angelsen, A., 2008. Moving Ahead with REDD: Issues, Options and Implications. Center for International Forestry Research (CIFOR), Bogor, Indonesia.
Angelsen, A., Brockhaus, M., Sunderlin, W.D., Verchot, L., 2012. Analysing REDD+: Challenges and Choices. Center for International Forestry Research (CIFOR), Bogor, Indonesia.
Assunção, J., Gandour, C.C., Rocha, R., 2012. Deforestation slowdown in the legal Amazon: Prices or policies? Report by the Climate Policy Initiative/PUC-Rio.
Burnham, K.P., Anderson, D.R., 1999. Model Selection and Inference: A Practical Information-Theoretic Approach. Springer-Verlag, New York.
Chagas, T., Costenbader, J., Streck, C., Roe, S., 2013. Baselines: concepts, functions, and application in REDD+ and forest carbon standards. A report by Climate Focus.

Corberán-Vallet, A., Bermúdez, J.D., Vercher, E., 2011. Forecasting correlated time series with exponential smoothing models. Int. J. Forecast. 27, 252–265.
DeFries, R., Herold, M., Verchot, L., Macedo, M.N., Shimabukuro, Y., 2013. Export-oriented deforestation in Mato Grosso: harbinger or exception for other tropical forests? Philos. Trans. R. Soc. B 368, 20120173. http://dx.doi.org/10.1098/rstb.2012.01731471-2970.
De Gooijer, J.G., Hyndman, R.J., 2006. 25 years of time series forecasting. Int. J. Forecast. 22, 443–473.
Ebeling, J., Yasué, M., 2008. Generating carbon finance through avoided deforestation and its potential to create climatic, conservation and human development benefits. Philos. Trans. R. Soc. B 363, 1917–1924.
Ferraro, P.J., 2001. Global habitat protection: Limitations of development interventions and a role for conservation performance payments. Conserv. Biol. 15, 990–1000.
Ferraro, P.J., Kiss, A., 2002. Direct payments for biodiversity conservation. Science 298, 1718–1719.
Ferraro, P.J., Pattanayak, S.K., 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. PLoS Biol. 4, e105. (0482–0488).
Ferraro, P., Simpson, R., 2002. The cost-effectiveness of conservation payments. Land Econ. 78, 339–353.
Gardner Jr., E.S., 1985. Exponential smoothing: The state of the art. J. Forecast. 4, 1–28.
Gardner Jr., E.S., 2006. Exponential smoothing: The state of the art—Part II. Int. J. Forecast. 22, 637–666.
Ghazoul, J., Butler, R.A., Mateo-Vega, J., Koh, L.P., 2010. REDD: a reckoning of environment and development implications. Trends Ecol. Evol. 25, 396–402.
Gibbs, B.H.K., Brown, S., Niles, J.O., Foley, J.A., 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environ. Res. Lett. 2, 045023. http://dx.doi.org/10.1088/1748-9326/2/4/045023.
Gibbs, B.H.K., Rausch, L., Munger, J., Schelly, I., Morton, D.C., Noojipady, P., Barreto, P., Micol, L., Walker, N.F., Amazon, B., Cerrado, E., 2015. Brazil's soy moratorium. Science 347, 377–378.
Grassi, G., Monni, S., Federici, S., Achard, F., Mollicone, D., 2008. Applying the conservativeness principle to REDD to deal with the uncertainties of the estimates. Environ. Res. Lett. 3, 035005.
Griscom, B., Shoch, D., Stanley, B., Cortez, R., Virgilio, N., 2009. Sensitivity of amounts and distribution of tropical forest carbon credits depending on baseline rules. Environ. Sci. Policy 12, 897–911.
Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853.
Huettner, M., Leemans, R., Kok, K., Ebeling, J., 2009. A comparison of baseline methodologies for reducing emissions from deforestation and degradation. Carbon Balance Manage. 4, 4.
Hyndman, R.J., Athanasopoulos, G., 2014. Forecasting: principles and practice. http://otexts.org/fpp/ (accessed on 28.07.15).
Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. Forecasting with Exponential Smoothing: The State Space Approach. Springer, Berlin.
INPE, 2013. National Institute for Space Research. http://www.inpe.br/ingles/ (accessed on 28.11.13).
Kissinger, G., 2015. Fiscal incentives for agricultural commodity production: Options to forge compatibility with REDD+. UN-REDD Programme Policy Brief 7.
Moore, D.S., McCabe, G.P., Craig, B.A., 2009. Introduction to the Practice of Statistics. W.H. Freeman, New York.
Müller, D., Sun, Z., Vongvisouk, T., Pflugmacher, D., 2014. Regime shifts limit the predictability of land-system change. Global Environ. Change 28, 75–83.
Olander, L.P., Gibbs, H.K., Steininger, M., Swenson, J.J., Murray, B.C., 2008. Reference scenarios for deforestation and forest degradation in support of REDD: a review of data and methods. Environ. Res. Lett. 3, 025011.
R: Development Core Team, 2015. R: A Language and Environment for Sstatistical Computing. R Foundation for Statistical Computing, Vienna, http://www.Rproject.org.
Shibata, R., 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. Ann. Statist. 8, 147–164.
Simpson, R., Sedjo, A., 1996. Paying for the conservation of endangered ecosystems: A comparison of direct and indirect approaches. Environ. Dev. Econ. 1, 241–257.
Sloan, S., Pelletier, J., 2012. How accurately may we project tropical forest-cover change? A validation of a forward-looking baseline for REDD. Global Environ. Change 22, 440–453.
Tatem, A.J., Guerra, C.A., Atkinson, P.M., Hay, S.I., 2004. Athletics: Momentous sprint at the 2156 Olympics? Nature 431, 525. http://dx.doi.org/10.1038/431525a. (30 September 2004).
UNFCCC 2009. Decision 4/CP.15, paragraph 7. http://unfccc.int/resource/docs/2009/cop15/eng/11a01.pdf.
UNFCCC 2013. Decision 13/CP.19, paragraph 1. http://unfccc.int/resource/docs/2013/cop19/eng/10a01.pdf.
Van der Werf, G.R., Morton, D.C., DeFries, R.S., Olivier, J.G.J., Kasibhatla, P.S., Jackson, R.B., Collatz, G.J., Randerson, J.T., 2009. CO_2 emissions from forest loss. Nat. Geosci. 2, 737–738.
Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, Boca Raton, Florida.