TIMBER v0.1: a conceptual framework for emulating temperature responses to tree cover change

Shruti Nath[1,2], Lukas Gudmundsson[2], Jonas Schwaab[2], Gregory Duveiller[3], Steven J. De Hertog[4], Suqi Guo[5], Felix Havermann[5], Fei Luo[6,7], Iris Manola[6], Julia Pongratz[5,8], Sonia I. Seneviratne[2], Carl F. Schleussner[1,9], Wim Thiery[4], and Quentin Lejeune[1]

[1]Climate Analytics, Berlin, Germany
[2]Institute of Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[3]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany
[4]Vrije Universiteit Brussel, Department of Hydrology and Hydraulic Engineering, Brussels, Belgium
[5]Ludwig-Maximilians-University Munich, Department of Geography, Munich, Germany
[6]Vrije Universiteit Amsterdam, Institute for Environmental studies, Amsterdam, Netherlands
[7]Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands
[8]Max Planck Institute for Meteorology, Hamburg, Germany
[9]Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys) and Geography Department, Humboldt-Universität zu Berlin, Berlin, Germany

Correspondence: shruti.nath@climateanalytics.org

Abstract. Society is set to experience significant land cover changes in order to achieve the temperature goals agreed upon under the Paris Agreement. Such changes carry both global implications, pertaining to the biogeochemical effects of land cover change and thus the global carbon budget, and regional/local implications, pertaining to the biogeophysical effects arising within the immediate area of land cover change. Biogeophysical effects of land cover change are of high relevance to national

5   policy- and decision- makers and their accountance is essential towards effective deployment of land cover practices that optimises between global and regional impacts. To this end, ESM outputs that isolate the biogeophysical responses of climate to land cover changes are key in informing impact assessments and supporting scenario development exercises. Generating multiple such ESM outputs, in a manner that allows comprehensive exploration of all plausible land cover scenarios however, is computationally untenable. This study proposes a framework to agilely explore the local biogeophysical responses of climate

10   under different land cover scenarios by means of a computationally inexpensive emulator, TIMBER v0.1. The emulator is novel in that it solely represents the land cover forced, biogeophysical responses of climate, and can be used as either a standalone device or supplementary to existing climate model emulators that represent greenhouse gas (GHG)- or Global Mean Temperature (GMT)- forced climate responses. We start off by modelling local minimum, mean and maximum surface temperature responses to tree cover changes by means of a month- and Earth System Model (ESM)- specific Generalised

15   Additive Model (GAM) trained over the whole globe. 2-m air temperature responses are then diagnosed from the modelled minimum and maximum surface temperature responses using observationally derived relationships. Such a two-step procedure accounts for the different physical representations of surface temperature responses to tree cover changes under different ESMs, whilst respecting a definition of 2-m air temperature that is more consistent across ESMs and with observational datasets. In exploring new tree cover change scenarios, we employ a parametric bootstrap sampling method to generate multiple possible

20   temperature responses, such that the parametric uncertainty within the GAM is also quantified. The output of the final emulator
is demonstrated for the SSP 1-2.6 and 3-7.0 scenarios. Relevant temperature responses are identified as those displaying a clear
signal in relation to their surrounding parametric uncertainty, calculated as the "signal-to-noise" ratio between the sample set
mean and sample set variability. The emulator framework developed in this study thus provides a first step towards bridging
the information-gap surrounding biogeophysical implications of land cover changes, allowing for smarter land-use decision

25   making.

1   Introduction

Following the Paris Agreement in 2015, 42% of Nationally Determined Contributions (NDCs) submitted by countries included
afforestation/reforestation based actions and targets (Seddon et al., 2020). The recent COP26 in Glasgow furthermore saw a
pledge to halt and reverse deforestation by 2030 (COP, 2021). Considering this, society is set to experience notable land cover

30   changes in hopes to achieve global warming levels well below +2 °C and pursue efforts in limiting them to +1.5 °C above pre-
industrial levels. In anticipation of this, the Earth System Model (ESM) community has put great effort into understanding and
quantifying the biogeochemical and biogeophysical effects of land cover changes (De Noblet-Ducoudré et al., 2012; Lawrence
et al., 2016; Davin et al., 2020).

Biogeochemical effects of land cover changes largely affect the global carbon budget, while biogeophysical effects are essen-

35   tial towards understanding regional climate impacts as well as extremes (De Noblet-Ducoudré et al., 2012; Pitman et al., 2012;
Lejeune et al., 2018). Recent studies by Windisch et al. (2021) and Lawrence et al. (2022), highlighted the need to consider the
biogeophysical effects of land cover changes in order to effectively identify and prioritise areas for re/afforestation and con-
servation. Such underscores the regional importance of the biogeophysical effects of land cover changes under future climate
scenarios (Seneviratne et al., 2018; Hirsch et al., 2018), and evidences the need to consider them within impact assessments

40   (Popp et al., 2017) and scenario development exercises (Van Vuuren et al., 2012; Calvin and Bond-Lamberty, 2018). Exploring
the biogeophysical effects of land cover changes under all possible future land cover scenarios solely through ESMs however,
quickly becomes untenable due to computational costs, and it is worth pursuing computationally inexpensive alternatives such
as climate model emulators.

Climate model emulators are computationally inexpensive tools, trained on available climate model runs to then render

45   probability distributions of key climate variables for runs that have not been generated yet. By statistically representing select
climate variables, emulators are able to reduce the dimensionality of climate model outputs, allowing for agile exploration of the
uncertainty phase space surrounding climate projections. Climate model emulators designed to reproduce regional/grid point
level, annual to monthly temperature projections usually operate as ESM-specific and start by deterministically representing
the regional/grid point level mean response of temperatures to a certain forcing, after which the residual variability – treated as

50   the uncertainty due to natural climate variability – is sampled or stochastically generated (Alexeeff et al., 2018; McKinnon and
Deser, 2018; Link et al., 2019; Castruccio et al., 2019; Beusch et al., 2020; Nath et al., 2022b). Outputs of such emulators act
as approximations of multi-model initial-condition ensembles, providing distributions of temperature responses to the forcing

of choice for impact assessments. To date however, such climate model emulators mainly represent the greenhouse gas (GHG)-
or Global Mean Temperature (GMT)- forcing within their mean response, neglecting the biogeophysical effects of land cover
55   changes.

In this study, we set up a conceptual framework for emulating the biogeophysical responses of climate variables to land cover
changes, hereon referred to simply as "responses". As a first step, we focus on emulating the surface and 2-m air temperature
responses to land cover changes between forest and cropland, simply denoted as "tree cover changes". The resulting emulator
constitutes a prototype version of the Tree cover change clIMate Biobphyscial responses EmulatoR, i.e. TIMBER v0.1. Since
60   representation of natural climate variability is well-explored in other emulators, TIMBER v0.1 purely focusses on representing
the mean response of temperatures to tree cover change. In doing so, we recognise that the ESM data available for training
(described under Section 2) is under-representative of the full range of possible tree cover changes across the globe. Conse-
quently, we pursue a more probabilistic representation, such that parametric uncertainties given the training data population are
accounted for. TIMBER v0.1 can thus be used as a standalone device or as supplementary to other emulators. The structure of
65   this paper is as follows: Section 3 introduces the emulator framework and its calibration and evaluation procedure; Section 4
presents the calibration and evaluation results, and illustrates some emulator outputs; Section 5 demonstrates the application of
the emulator to different Shared Socio-economic Pathway (SSPs) scenarios; and Section 6 wraps up with the conclusion and
outlook.

2   Training datasets

2.1   ESM experiments

Idealised Earth System Model (ESM) experiments that isolate the effects of tree cover change on the climate were run as
part of the LAnd MAnagement for CLImate Mitigation and Adaptation (LAMACLIMA) project. The experimental setup
was designed to capture the maximal potential climate response due to af/re/deforestation as compared to present day land
cover conditions. Accordingly, extreme afforestation (AFF) and deforestation (DEF) scenarios were run alongside a reference
75   scenario (REF). The REF scenario comprises of simulations spanning 150 years with land cover conditions and other forcings
(GHG emissions etc.) kept constant at 2015 levels. The AFF (DEF) scenario then consists of full expansion of forest (crop)
cover relative to that of 2015 levels with all other forcings again kept constant at 2015 levels. The difference between AFF
(DEF) run and REF run outputs provides the climate response to re/afforestation (deforestation).

Temperature responses derived from the ESM simulations are distinguished into local and non-local responses following
80   the checkerboard approach developed by Winckler et al. (2017). Local responses represent the expected climate responses to
land cover change within the immediate area of change and can be applied in any global tree cover change scenario, whereas
non-local responses represent remote effects of land cover change and depend on the global extent and patterns of land cover
change. Given that local responses are independent of land cover change scenario, we focus only on them for the rest of this
study and the term "response" exclusively refers to the local response hereon. Participating ESMs running simulations within

85 the LAMACLIMA project are the Community ESM (CESM2), the Max Planck Institute ESM (MPI-ESM) and the European
Consortium ESM (EC-EARTH).

2.1.1 Observational dataset

To demonstrate the applicability of this study's approach on observational data we use the Duveiller et al. (2018b) dataset
which provides an observationally derived DEF scenario. The dataset maps potential local responses of daytime, mean and
90 nighttime surface temperatures to land cover transitions using observational data and a "space-for-time" substitution approach
as described by Duveiller et al. (2018a). The dataset considers transitions from forest to several other land cover types (e.g.
shrubland, grassland etc.). For comaparability with ESM runs, we choose to only focus on forest transitions to cropland. It
should be noted that the dataset contains some information gaps in space and is thus spatially sparse as compared to the spatially
complete ESM output fields. Additionally, since no dataset providing an observationally derived extreme afforestation scenario
95 is available, only on one end of extreme tree cover change is represented.

3 Statistical emulation of temperature response to changes in tree cover

3.1 Overview on modelling approach

The emulation of temperature responses to tree cover changes is split into two parts. The first part seeks to statistically repre-
sent monthly minimum ($\Delta TS_{m,s}^{min}$), mean ($\Delta TS_{m,s}^{mean}$) and maximum ($\Delta TS_{m,s}^{max}$) surface temperature responses, generically
100 referred to as $\Delta TS_{m,s}$, to tree cover change (Sect. 3.2). The second part then seeks to diagnose 2-m air temperature responses
($\Delta T_{m,s}^{2m}$) from the statistically represented surface temperature responses using observationally derived relationships (Sect.
3.3).

    $T_{m,s}^{2m}$ is an important variable for impact assessments, however is diagnosed differently across ESMs (leading to inter-ESM
discrepancies in their modelled response to tree cover change) and is also defined differently between ESMs and observations.
105 Splitting the emulation exercise into its two parts maintains a response of surface temperature to tree cover change that is
specific to the ESM/observational data trained on, from which $\Delta T_{m,s}^{2m}$ is then diagnosed using observationally derived relation-
ships independent of training data. In such, we account for the different physical representations of temperature responses to
tree cover change for each ESM, whilst also ensuring a consistent definition of $\Delta T_{m,s}^{2m}$ across ESMs and with observational
datasets, ergo the possibility to compare them.

110 3.2 Modelling surface temperature responses to tree cover change

In the following subsections, we introduce the model used for representing $\Delta TS_{m,s}$ to tree cover changes, followed by our
approach in calibrating and evaluating it. In choosing and calibrating the model, we are especially mindful of the training
datasets being solely representative of grid points which undergo extreme tree cover changes relative to the 2015 period, as
performed within the training simulations. Consequently, we require a model that can train over the whole globe (as otherwise

115 there are at most two samples per grid point to train on) and need to account for the resulting spatially-structured training data during model calibration. Moreover, we recognise that evaluation can only be done on the training datasets as no other ESM simulations performing the checkerboard approach of Winckler et al. (2017) exist, and thus settle for synthesising the closest possible representation of the model's out-of-sample performance during evaluation.

3.2.1 Model

120 We model $\Delta TS_{m,s}$ conditional on tree cover change and geographical attributes using a month-specific Generalized Additive Model (GAM), trained over the whole globe. The GAM, hereon referred to as $\Gamma_m^{min/mean/max}$ or more generically as Γ_m, is provided by the python pyGAM package. Γ_m can easily ingest multidimensional data and has the advantage that it does not prescribe any functional form, allowing flexibility in representing linear to more complex response types. The input predictor matrix ($\mathbf{X}$) given to Γ_m is composed of tree cover changes relative to the 2015 ($\Delta_{2015}treeFrac$) and geographical attributes

125 of longitude (lon), latitude (lat) and orography ($orog$) and the conditional distribution of $\Delta TS_{m,s}^{mean}$ is assumed as normal,

$$\Gamma_m = \mathbb{E}[\Delta TS_{m,s}|\mathbf{X}] = te_m(\Delta_{2015}treeFrac_{m,s}, lon_s, lat_s, by = orog_s) \qquad \text{where} \qquad [\Delta TS_{m,s}|\mathbf{X}] \sim \mathcal{N} \qquad (1)$$

te_m represents a tensor spline term built across the $\Delta_{2015}treeFrac$, lon, lat space with coefficient terms stratified according to $orog$ using the by operator For further details on tensor splines and the by operator, see Wood (2017). Γ_m^{mean} can be calibrated for its lambda parameter (λ), which controls the complexity in shape of te_m (where a smaller λ value allows for a more

130 complex shape) and its number of basis functions (where more basis functions means more degrees of freedom).

3.2.2 Calibration

We calibrate Γ_m by means of blocked cross validation which entails systematic train-test splitting (Roberts et al., 2017), as shown in Figure 1. We choose to do so as Γ_m ingests spatially-structured training data and consequently, traditional cross validation with random train-test splitting, cannot be done as samples are not independent of one another.

135 Systematic train-test splitting is done by creating blocks of samples which share climate and geographical characteristics. For this, grid points are first clustered according to their background climate in the REF simulation (and for Observations, historical climatological data as provided by WorldClim v2, https://www.worldclim.org/data/worldclim21.html) and then further split into groups according to continuous geographical regions: Africa, North America, South America, Australia, Eurasia, Tibetan Plateau and the South-East Asian Islands. K-means clustering is used to cluster grid points according to background climate,

140 with temperature and relative humidity considered as background climate indicators. To select the optimal number of clusters we calculate the improvement in performance of the K-means clustering algorithm (measured as the average distance of all points from the centre of their respective cluster groups, a smaller distance indicating better performance) with increasing number of clusters and select the number of clusters after which no further improvement in performance is observed.

By cross validating across composite blocks dimensioned along climate and geographical space, we hope to nudge the λ

145 parameter and number of basis functions to values that most flexibly apply across all possible response types to tree cover changes (as modulated by the background climate) whilst ensuring independence between training and test sets by accounting

for spatial auto-correlations (by considering geographical attributes). Cross validation is carried out across all train-test splits, such that each block appears in the test set once, and the λ parameter and number of basis functions yielding the best Γ_m performance, as measured by Root Mean Squared Error (RMSE), across all test sets are selected.
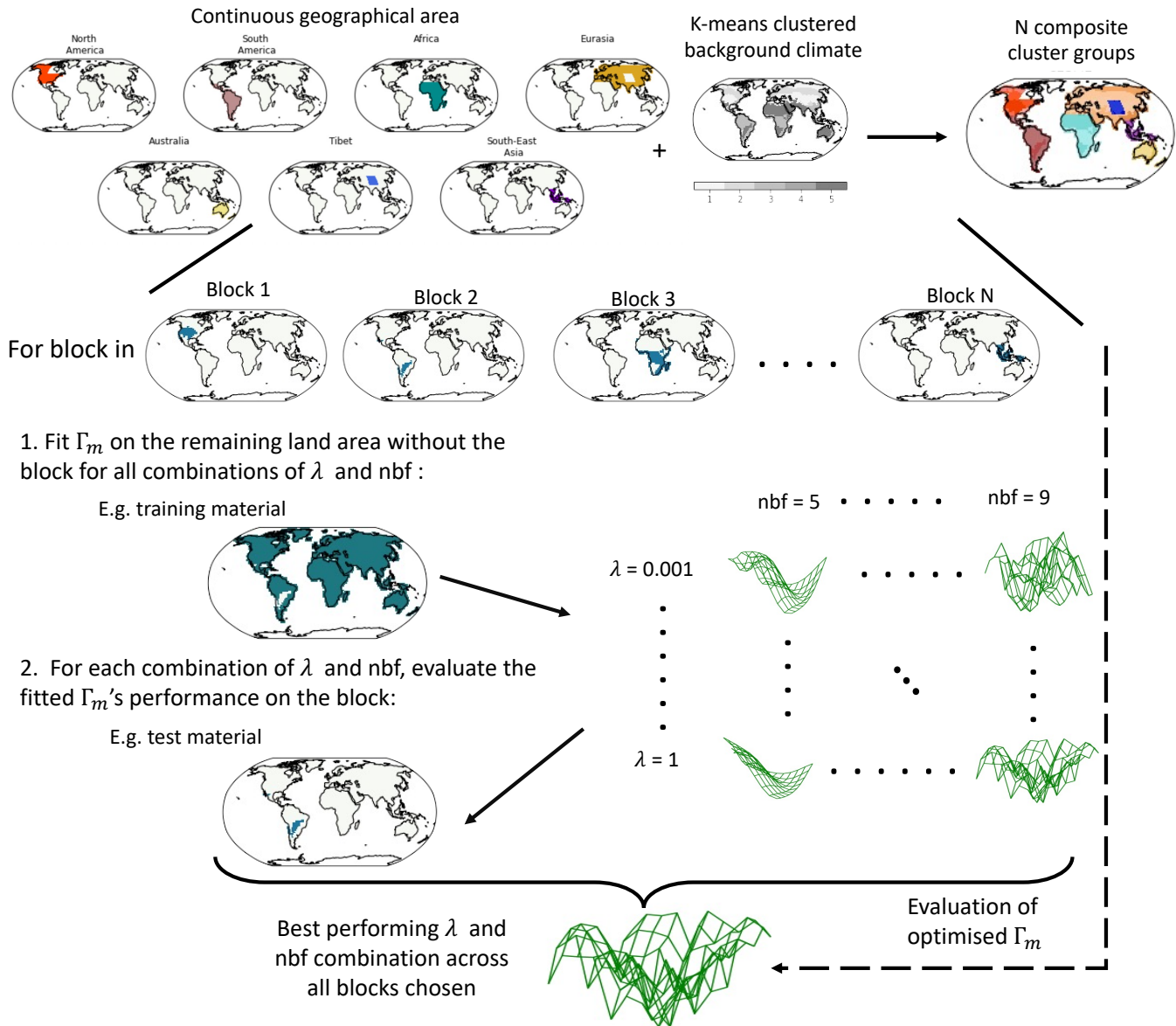
3.2.3 Evaluation

Blocked cross validation is further employed to evaluate the calibrated Γ_m's performance into "no-analogue" conditions where the model has the least information (Roberts et al., 2017), thus providing a representative idea of the model's ability to predict into new, unseen tree cover change scenarios. When predicting into unseen tree cover change scenarios, we mainly require the model to predict well across different background climates as well as different ranges of tree cover change. We hence divide the "no-analogue" conditions evaluated for into those pertaining to background climate conditions and those pertaining to tree cover change conditions, and consider Γ_m's performance for each continuous geographical region separately. Since Γ_m was originally calibrated by creating blocks that considered both climate and geographical space, the performance into "no-analogue" background climates is simply taken as the final chosen Γ_m's test set RMSEs calculated during calibration. Γ_m's prediction ability into "no-analogue" tree cover change conditions is assessed by performing an additional blocked cross validation with blocks composed of continuous geographical region and $\Delta_{2015} treeFrac$ bin and once again taking the test set RMSEs. In binning $\Delta_{2015} treeFrac$ we consider the sign of change and the magnitude of change (partitioned into [0.01-0.15), [0.15-0.3), [0.3-0.5),[0.5-0.8) and [0.8-1.0]).

3.3 Diagnosing 2m air temperatures from surface temperatures

Hooker et al. (2018) were able to derive month-specific relationships between observational night and day surface temperatures ($TS_{m,s}^{night/day}$) and observational $T_{m,s}^{2m}$. They did so by performing both Geographical and Climate Space Weighted Regression (GWR and CSWR) between observational $TS_{m,s}^{night/day}$ and observational $T_{m,s}^{2m}$ values, so as to obtain grid point level coefficients specific to geographical/background climate conditions. By taking a stacked generalisation of the GWR and CSWR outputs, Hooker et al. (2018) were able to reconstruct global $T_{m,s}^{2m}$ maps over the period 2003 to 2016 in a geographically and climatically consistent manner.

In this study, we use the Hooker et al. (2018) model to diagnose $T_{m,s}^{2m}$ from surface temperatures. Ideally, the Hooker et al. (2018) model would be refitted to derive ESM-specific coefficients between ESM surface temperatures and observed $T_{m,s}^{2m}$ data. Given that this study primarily focusses on setting up a conceptual framework however, we choose to directly apply the original coefficients derived by Hooker et al. (2018) as an initial proof-of-concept. Before applying the Hooker et al. (2018) model, we first make some modifications to it so as to enable a smooth translation between observed and ESM spaces. In the following subsections, we introduce the modifications made to the Hooker et al. (2018) model and furthermore outline some tests performed to check that the modified version of it applied to ESMs still yields results comparable to those expected from observations.

Figure 1. Framework for block cross validation used in calibrating and evaluating Γ_m.

3.3.1 Modifications of the Hooker et al. (2018) model

$T_{m,s}^{2m}$ values are diagnosed using a modified version of the Hooker et al. (2018) model which uses $TS_{m,s}^{min/max}$ values instead

180 of $TS_{m,s}^{night/day}$ and only considers the GWR coefficient terms,

$$T_{m,s}^{2m} = \beta_{0,m,s}^{GWR} + \beta_{1,m,s}^{GWR} \cdot TS_{m,s}^{min} + \beta_{2,m,s}^{GWR} \cdot TS_{m,s}^{max} \tag{2}$$

assuming that the effects of land cover type are minimal on $\beta_{0,m,s}^{GWR}$, we then get,

$$\Delta T_{m,s}^{2m} = \beta_{1,m,s}^{GWR} \cdot \Delta TS_{m,s}^{min} + \beta_{2,m,s}^{GWR} \cdot \Delta TS_{m,s}^{max} \tag{3}$$

Where $\beta_{0,m,s}^{GWR}$, $\beta_{1,m,s}^{GWR}$ and $\beta_{0,m,s}^{GWR}$ are coefficient terms obtained from GWR. We choose not to use the CSWR coefficient

185 terms as background climates between observations and ESMs are not consistent and there is the additional uncertainty sur-
rounding the evolution of CSWR coefficient terms under changing background climates. Additionally, we use $TS_{m,s}^{min/max}$
values instead as they are the only available DEF and AFF scenario ESM outputs which are most similar to $TS_{m,s}^{night/day}$.

3.3.2 Tests on the modified Hooker et al. (2018) model applied to the ESM space

Since we look at relative changes in $T_{m,s}^{2m}$, the modifications made to the Hooker et al. (2018) model are expected to have

190 minimal impact as long as the biases in $T_{m,s}^{2m}$ values calculated using ESM $TS_{m,s}^{min/max}$ values have the same spread as those
arising from natural variability within observational $TS_{m,s}^{night/day}$ values, and are thus "acceptable". To determine this, we
compare the spread of biases obtained when calculating $T_{m,s}^{2m}$ values from observational $TS_{m,s}^{night/day}$ values to those obtained
from $TS_{m,s}^{min/max}$ ESM outputs for the REF scenario. $TS_{m,s}^{min/max}$ outputs from the REF scenario are used, as we consider
them representative of the natural variability surrounding $TS_{m,s}^{min/max}$ values. We approximate the spread of biases by taking

195 into account the natural variability surrounding the surface temperature values and compare them through the following steps:

1. Construct a multivariate Gaussian process across all observational $TS_{m,s}^{night/day}$ values to generate spatially corre-
   lated pairs of $TS_{m,s}^{night/day}$ which also take into account cross-correlations between $TS_{m,s}^{night}$ and $TS_{m,s}^{day}$. Generated
   $TS_{m,s}^{night/day}$ pairs will act as "psuedo-samples" that represent the underlying uncertainty due to natural variability within
   observational data.

200 2. For each timestep of $T_{m,s}^{2m}$ predictions available from the original Hooker et al. (2018) model (going from 2003 to 2016):

   (a) Generate 100 synthetic pairs of $TS_{m,s}^{night/day}$ values using the Gaussian process constructed in Step 1.

   (b) Calculate the biases between the $T_{m,s}^{2m}$ prediction available from the original Hooker et al. (2018) model and those
       obtained by applying Equation 2 to the synthetically generated pairs of $TS_{m,s}^{night/day}$.

3. Take the 95% Interquartile Range (IQR) of the biases calculated in Step (2b) as a measure of their spread.

205 4. Repeat steps 1-3 for $TS_{m,s}^{min/max}$

5. Check the difference between the IQR calculated in step 3 using ESM $TS_{m,s}^{min/max}$ values and that calculated using
   observational $TS_{m,s}^{night/day}$ values. A positive difference indicates more spread within the biases for $TS_{m,s}^{min/max}$ derived
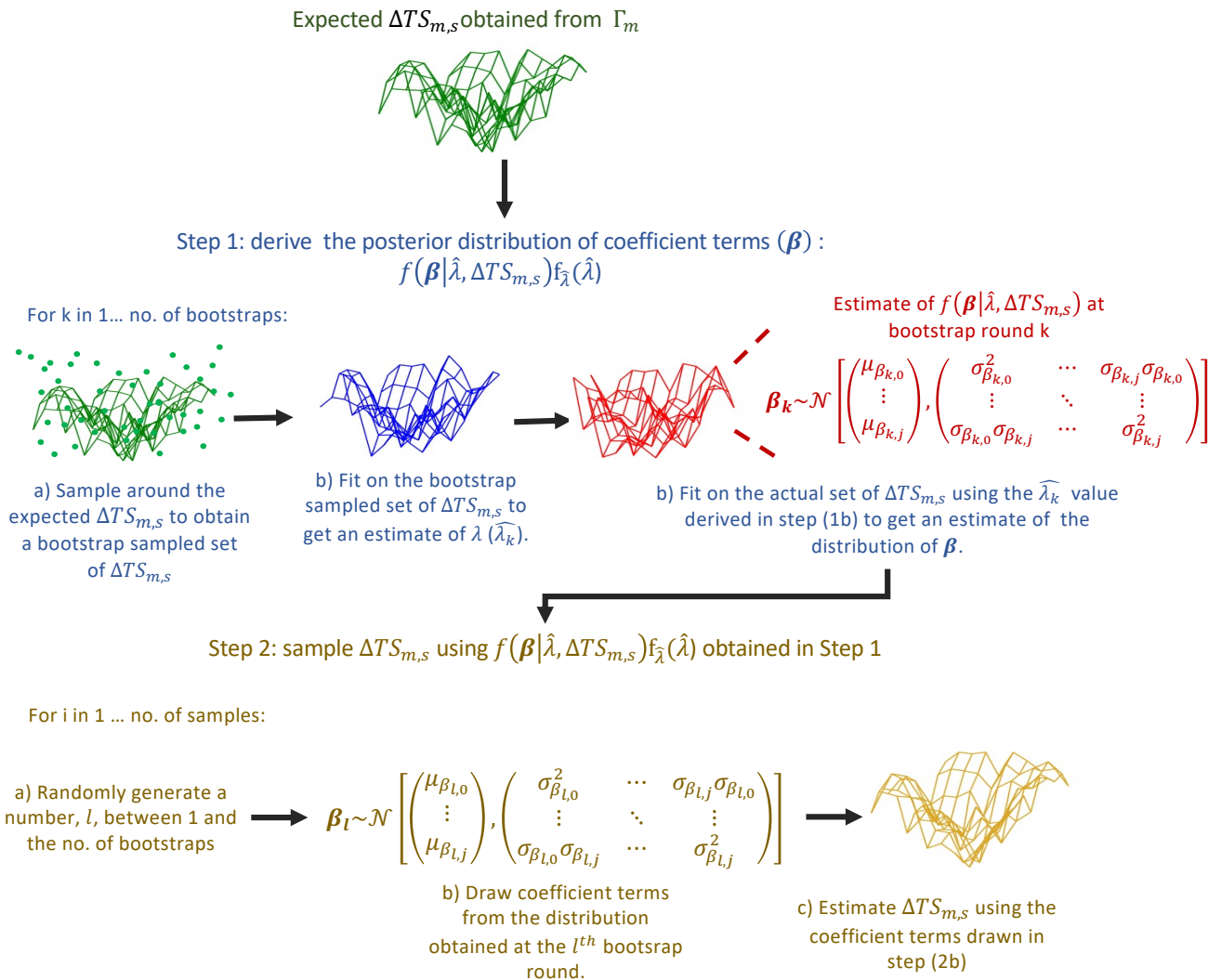
$T_{m,s}^{2m}$ values, in which case the biases are not acceptable considering those arising from natural variability within the observational data.

3.4 Emulating 2m air temperature changes under different land cover change scenarios

Γ_m is able to generate multiple $\Delta TS_{m,s}$ values, at any given point in the predictor space, in a manner that accounts for the parametric uncertainty within the GAM given the training data provided. It does so following a parametric bootstrap method as outlined in Figure 2 (Wood, 2017; Efron and Tibshirani, 1993). Parametric bootstrapping constitutes of first approximating the joint distribution of the coefficients ($\boldsymbol{\beta}$) and λ parameter used within Γ_m, conditional on the training data available i.e. $f(\boldsymbol{\beta}, \lambda | \mathbf{X})$ (Step 1, Figure 2), from which $\boldsymbol{\beta}$ values are then sampled to estimate surface temperature responses (Step 2, Figure 2). To avoid high computational costs, the joint distribution is approximated by first bootstrap sampling the distribution of λ conditional on the training material, i.e. $f_\lambda(\lambda)$ (Steps 1a-1b, Figure 2), from which the distribution of $\boldsymbol{\beta}$ conditional on both λ and the training material is constructed over the whole $f_\lambda(\lambda)$ space (Step 1c, Figure 2), such that $f(\boldsymbol{\beta}, \lambda | \mathbf{X}) \approx f(\boldsymbol{\beta} | \lambda, \mathbf{X}) \cdot f_\lambda(\lambda)$. Surface temperature response values are then sampled by drawing $\boldsymbol{\beta}$ distributions from random parts of the $f_\lambda(\lambda)$ space (Step 2a, Figure 2) and sampling coefficient values from them (Step 2b, Figure 2), which are then used to estimate $\Delta TS_{m,s}$ values (Step 2c, Figure 2).

By estimating $\Delta TS_{m,s}$ values in a manner where the $\boldsymbol{\beta}$ and λ uncertainty space is explored, parametric bootstrap allows for evaluation of the uncertainty in the derived shape of surface temperature responses to tree cover change. Such is especially useful when the training information only includes grid points which have experienced extreme tree cover change in the training simulations. To this end, sampling from Γ_m provides all possible shapes of responses in between the extreme ends of tree cover change available in the training data, and is suited towards robustly interpolating values which are structured along multiple dimensions (spatial and tree cover change dimensions in our case).

In representing temperature responses under new tree cover change scenarios, we thus sample $\Delta TS_{m,s}^{min/max}$ values from $\Gamma_m^{min/max}$ and diagnose $\Delta T_{m,s}^{2m}$ for each sample. The $\boldsymbol{\beta}$ and λ parameter uncertainty spaces are constructed using 10 bootstraps from which 200 samples are then drawn. We take the mean across all samples as the expected $\Delta T_{m,s}^{2m}$ value and the standard deviation across all samples as the underlying parametric uncertainty within the GAM. Given the computational expenses of running ESMs, such gives Γ_m the benefit of mainly requiring extreme tree cover change scenarios as training material, from which it can further explore all possible outcomes of in-between scenarios itself. It should be noted however that this does not remove the benefit of having more training material ontop of the extreme scenarios, but simply minimises the training data requirements of Γ_m.

Expected $\Delta TS_{m,s}$ obtained from Γ_m

Step 1: derive the posterior distribution of coefficient terms $(\boldsymbol{\beta})$:
$$f(\boldsymbol{\beta}|\hat{\lambda}, \Delta TS_{m,s})f_{\hat{\lambda}}(\hat{\lambda})$$

For k in 1... no. of bootstraps:

Estimate of $f(\boldsymbol{\beta}|\hat{\lambda}, \Delta TS_{m,s})$ at bootstrap round k

$$\boldsymbol{\beta_k} \sim \mathcal{N}\left[\begin{pmatrix}\mu_{\beta_{k,0}} \\ \vdots \\ \mu_{\beta_{k,j}}\end{pmatrix}, \begin{pmatrix}\sigma^2_{\beta_{k,0}} & \cdots & \sigma_{\beta_{k,j}}\sigma_{\beta_{k,0}} \\ \vdots & \ddots & \vdots \\ \sigma_{\beta_{k,0}}\sigma_{\beta_{k,j}} & \cdots & \sigma^2_{\beta_{k,j}}\end{pmatrix}\right]$$

a) Sample around the expected $\Delta TS_{m,s}$ to obtain a bootstrap sampled set of $\Delta TS_{m,s}$

b) Fit on the bootstrap sampled set of $\Delta TS_{m,s}$ to get an estimate of λ $(\widehat{\lambda_k})$.

b) Fit on the actual set of $\Delta TS_{m,s}$ using the $\widehat{\lambda_k}$ value derived in step (1b) to get an estimate of the distribution of $\boldsymbol{\beta}$.

Step 2: sample $\Delta TS_{m,s}$ using $f(\boldsymbol{\beta}|\hat{\lambda}, \Delta TS_{m,s})f_{\hat{\lambda}}(\hat{\lambda})$ obtained in Step 1

For i in 1 ... no. of samples:

a) Randomly generate a number, l, between 1 and the no. of bootstraps

$$\boldsymbol{\beta_l} \sim \mathcal{N}\left[\begin{pmatrix}\mu_{\beta_{l,0}} \\ \vdots \\ \mu_{\beta_{l,j}}\end{pmatrix}, \begin{pmatrix}\sigma^2_{\beta_{l,0}} & \cdots & \sigma_{\beta_{l,j}}\sigma_{\beta_{l,0}} \\ \vdots & \ddots & \vdots \\ \sigma_{\beta_{l,0}}\sigma_{\beta_{l,j}} & \cdots & \sigma^2_{\beta_{l,j}}\end{pmatrix}\right]$$

b) Draw coefficient terms from the distribution obtained at the lth bootsrap round.

c) Estimate $\Delta TS_{m,s}$ using the coefficient terms drawn in step (2b)

Figure 2. Sampling routine of the Generalized Additive Model. First an approximation of the coefficients' $(\boldsymbol{\beta})$ and λ parameter's joint distribution given the available training data is constructed (Step 1), from which coefficient terms are sampled to calculate $\Delta TS_{m,s}^{mean}$ values with (Step 2). Steps 1a-1b construct the sampling distribution of the λ parameter $(f_\lambda(\lambda))$ given the known variability in the training data, and Step 1c then constructs the distribution of $\boldsymbol{\beta}$ conditional on the training data and λ parameter at each point of the $f_\lambda(\lambda)$ space. In such the $\Delta TS_{m,s}^{mean}$ values calculated in Step 2 account for the uncertainty in the shape of $\Delta TS_{m,s}^{mean}$ responses, as modulated by $\boldsymbol{\beta}$ and λ values.

4 Results

4.1 Blocked cross validation results

Results from the blocked cross validation are two fold. Firstly, blocked cross validation is used to calibrate Γ_m for its optimal λ parameter and number of basis functions. Secondly, blocked cross validation is employed to evaluate the calibrated Γ_m's performance into "no-analogue" conditions. "No-analogue" conditions of background climate and those of tree cover change are considered specifically with a separate block cross validation performed for each. The following subsections show the blocked cross validation results for Γ_m^{mean} only as this gives a representative idea of the validity of this study's framework. Blocked cross validation results for $\Gamma_m^{min/max}$ are provided in the Appendix B

4.1.1 Calibration

Figure 3 provides the chosen λ parameter values and number of basis functions (nbf), alongside maps of the calibrated Γ_m^{mean}'s prediction RMSEs for the DEF and AFF scenarios (only DEF for observations). Maps of $\Delta treeFrac_{2015}$ implemented under the AFF and DEF scenarios are available for reference in Figure C1, Appendix C.

The Γ_m^{mean} trained on observational data has a λ parameter value of 0.001 for both January and July, which is significantly lower than that of 1 otherwise chosen for all ESMs. This could be as the observationally trained Γ_m^{mean} only receives training data for the DEF scenario, which implements large magnitudes of tree cover change localised to specific regions (e.g. Amazon, Congo Basin and South-East Asia). Thus, lower λ parameter values are favoured to allow for complex representation with higher spatial variability. CESM2 and EC-EARTH show higher RMSEs for the DEF scenario than the AFF scenario, which is related to difficulty in representing the complex response types with high spatial variabilities within the DEF scenario. This highlights a design consequence of Γ_m^{mean}, where te_m is fitted smoothly over lon, lat and $\Delta_{2015}treeFrac$, thus falling short in representing high spatial variabilities as brought about by large magnitudes of localised tree cover change.

MPI-ESM is the only ESM that shows similar performance across both DEF and AFF scenarios with no concentration of higher RMSEs around areas of significant deforestation. Additionally, MPI-ESM's Γ_m^{mean} favours the simplest representation across all ESMs with the lowest number of basis functions chosen for both January and July. Such indicates a smoother response type outputted by MPI-ESM, with deforestation in the tropics not necessarily leading to significant temperature jumps within space.

Overall, Γ_m^{mean} mostly displays RMSEs less than or equal to 0.25, with higher RMSEs localised to regions of extreme magnitudes of deforestation. In such, Γ_m^{mean} proves itself as a flexible framework to represent expected temperature responses to more realistic magnitudes of tree cover change.

11

RMSE of the fully calibrated Γ_m^{mean}
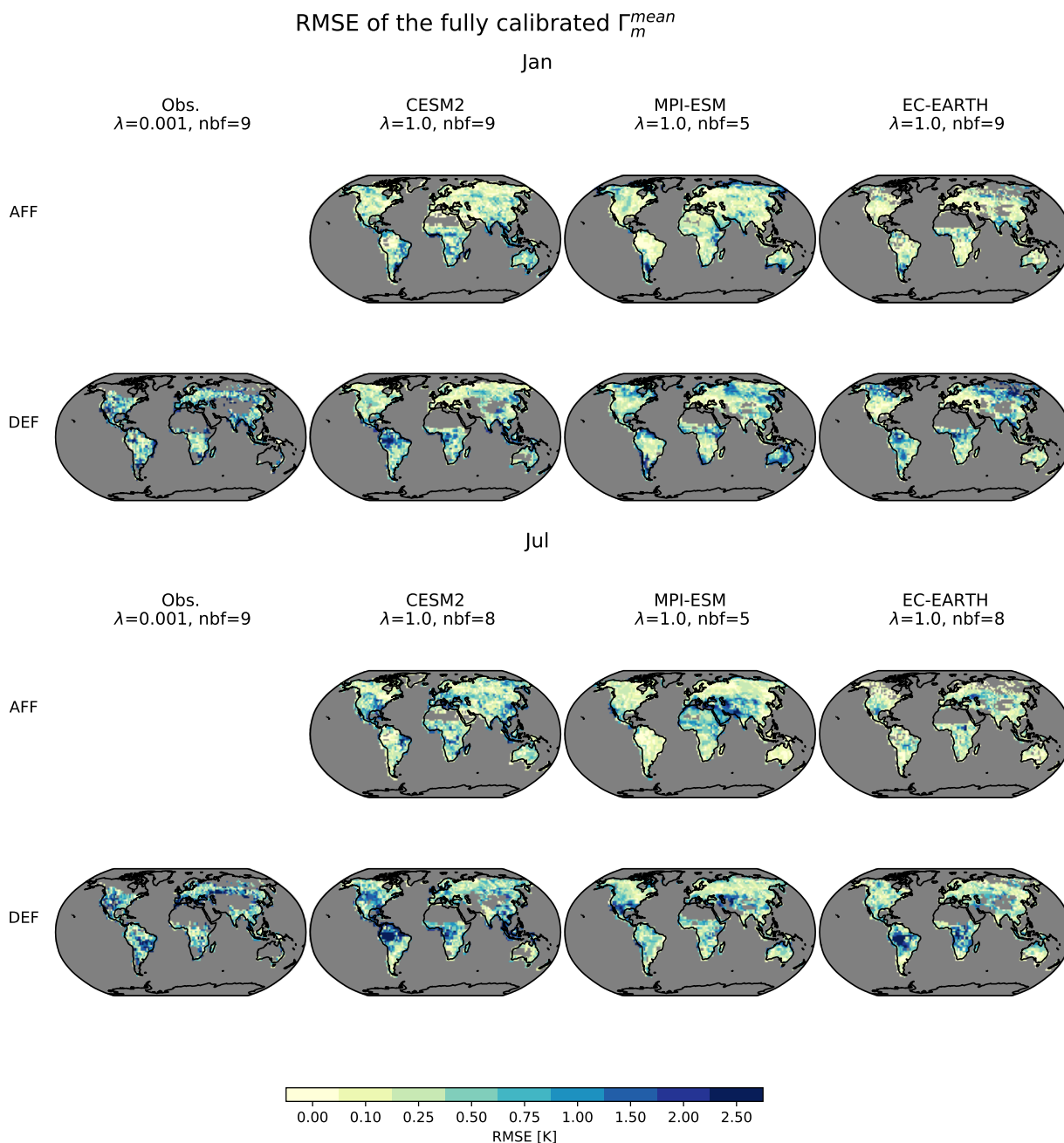
Figure 3. Performance of the final Γ_m^{mean} trained on each full set of observational/ESM data for example months of January (upper panel) and July (lower panel) shown as RMSE maps (rows) for AFF (first row) and DEF (second row) scenarios. Columns headers indicate the training dataset used and the respective λ parameter and number of basis functions (nbf) chosen during blocked cross validation.

4.1.2 Evaluation of Γ_m^{mean} under "no-analogue" background climate conditions

265 Figure 4 shows RMSEs obtained for Γ_m^{mean}'s predictions into "no-analogue" background climates aggregated to latitudinal bands for example months of January and July. RMSEs are differentiated into those obtained under the AFF scenario and the DEF scenario, except for observations where RMSEs are only available for the DEF scenario. For observations and ESMs, the spread in RMSEs displays a month dependency across all latitudinal bands, evidencing the seasonality in $\Delta TS_{m,s}^{mean}$ responses to tree cover change as well as the need for prior background climate information being more important for certain months

270 than others.

Despite the spread in RMSEs being large, median values are mostly in line with those seen in Figure 3, indicating overall good prediction skill for Γ_m^{mean} into unseen background climate conditions. Observation RMSEs for DEF in -45 °N to -15 °N however show significantly higher median values than those in Figure 3, although this is more likely due to data sparsity in the region.

275 Across ESMs, DEF in the tropics (-15 °N to 10 °N) shows the largest spreads in RMSEs with slightly higher median values than those of Figure 3. Given that Γ_m^{mean} is known to struggle within these areas due to the localised, large magnitudes of deforestation, exploration of its performance into "no-analogue" tree cover changes is first required before concluding lower prediction skill for unseen background climate conditions within these areas.

4.1.3 Evaluation of Γ_m^{mean} under "no-analogue" tree cover change conditions

280 Figure 5 shows the median RMSEs (with error bars indicating 50% confidence intervals) obtained for Γ_m^{mean}'s predictions into "no-analogue" tree cover changes aggregated to latitudinal bands for example months of January and July. For observations and ESMs, magnitudes and patterns of RMSEs are similar between January and July across all latitudinal bands, which is expected as background climate information is intuitively more important for representing seasonality in $\Delta TS_{m,s}^{mean}$ values.

Median RMSEs for $\Delta_{2015}treeFrac \leq$ -0.5 in the tropics are higher than those seen for DEF in Figure 4, indicating that the 285 prediction skill for Γ_m^{mean} is more dependent on the availability of training information for similar tree cover changes than for similar background climate. MPI-ESM is an exception to this, displaying much larger RMSEs for DEF in Figure 4. Such could result from MPI-ESM outputting a smoother response to tree cover change in the tropics as previously suggested in Section 4.1.1, making availability of prior background climate information the main factor influencing Γ_m^{mean}'s prediction skill.

Observations, CESM2 and EC-EARTH show an increase in RMSEs across all latitudinal bands as $\Delta_{2015}treeFrac$ values 290 move towards the more extreme ends (-1 for observations and +/-1 for CESM2 and EC-EARTH), sometimes even reaching RMSEs higher than those seen in Figure 3. This indicates lower prediction skill for Γ_m^{mean} into unseen, extreme tree cover change conditions for observations, CESM2 and EC-EARTH. Nevertheless, the resolved skill seen in Figure 3 verifies the need to have a training dataset representative of the extreme ends of tree cover change, as $TS_{m,s}^{mean}$ responses may systematically become more non-linear with increasing magnitudes of tree cover change.
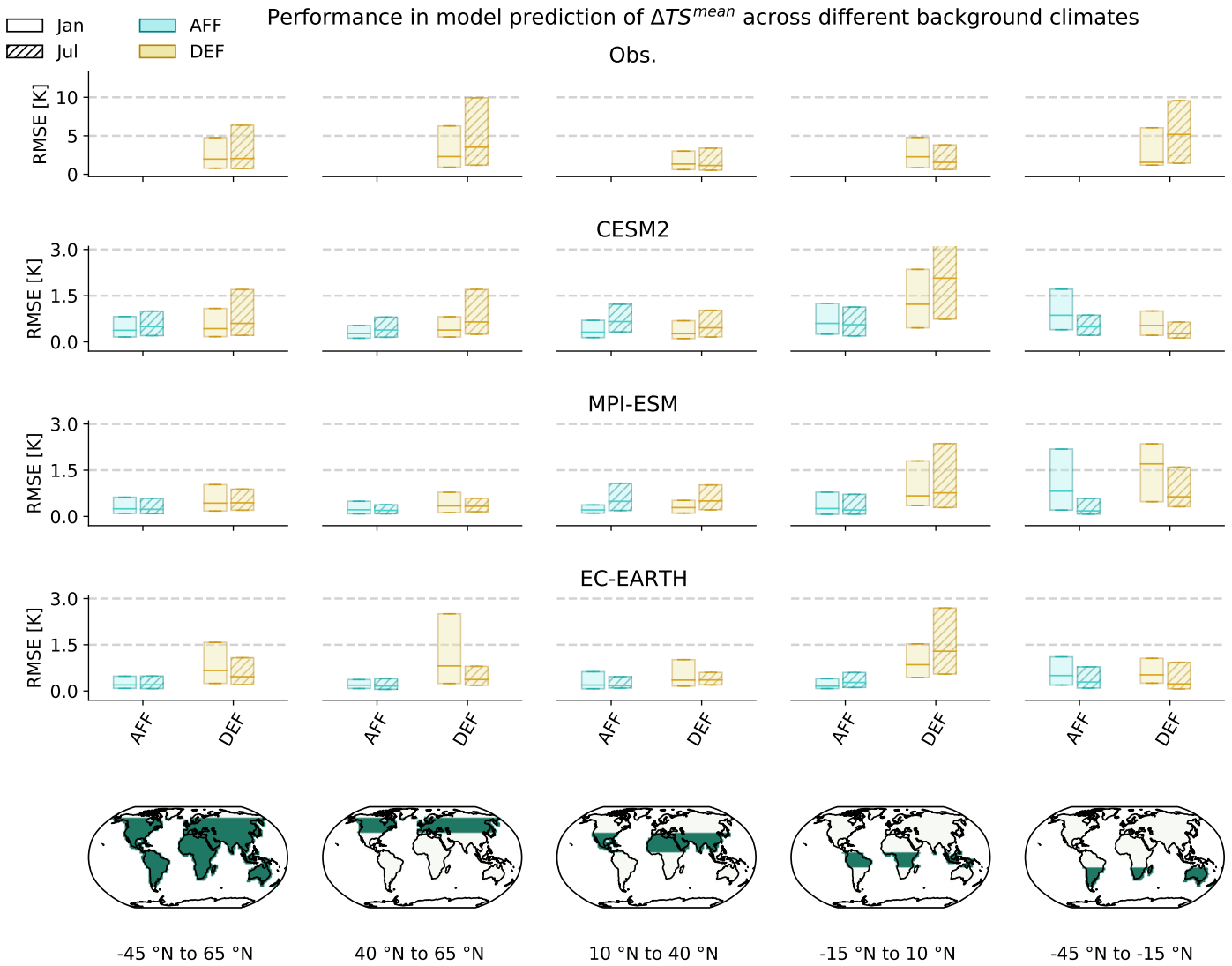
Figure 4. Evaluation of Γ_m^{mean}'s ability to predict across background climates within each continuous geographical region. Test set RMSEs, obtained during blocked cross validation with blocks clustered according to background climate and continuous geographical region are considered (as shown in Figure A1). RMSEs are shown for the months of January (unhatched) and July (hatched) and are aggregated to latitudinal band and direction of tree cover change, yellow indicating a negative change (DEF) and blue indicating a positive change (AFF). 50% confidence intervals are shown.
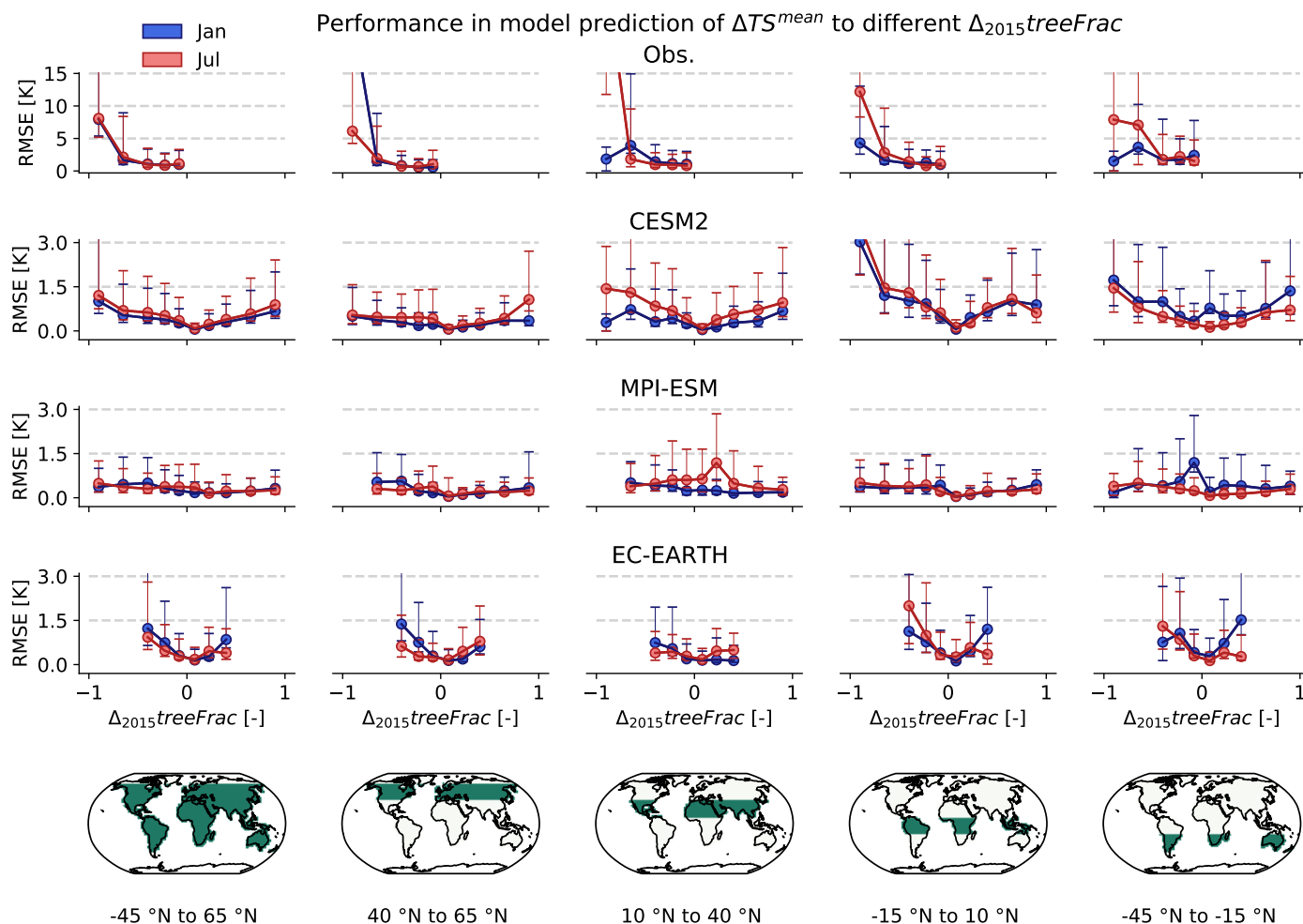
Figure 5. Evaluation of Γ_m^{mean}'s ability to predict across $\Delta_{2015}treeFrac$. Test set RMSEs, obtained during blocked cross validation with blocks clustered according to continuous geographical region and $\Delta_{2015}treeFrac$ (grouped according to sign of change and absolute value as binned into [0.01,0.15), [0.15-0.3),[0.3-0.5),[0.5-0.8) and [0.8-1.0]), are considered. RMSEs are shown for January (blue) and July (red) and are aggregated to latitudinal band and plotted against the centre of each $\Delta_{2015}treeFrac$ bin. Error bars indicate the 50% confidence interval.

4.2 Illustration of Γ_m^{mean} outputs

In this section, we demonstrate the ability of Γ_m^{mean} to predict $\Delta TS_{m,s}^{mean}$ across all ranges of tree cover changes. Figure 6 illustrates the mean $\Delta TS_{m,s}^{mean}$ predictions as well as their 95% interval calculated across all grid points within a given latitudinal band. We choose latitudinal bands which have similar properties in $TS_{m,s}^{mean}$ responses to tree cover changes as seen under De Hertog et al. (2022), namely: northern-hemispheric, temperate (40°N to 65°N); subtropical, temperate (10°N to 40°N); tropical (-15°N to 10°N); and southern-hemispheric (-45°N to -15°N). Southern-hemispheric results are not differentiated into subtropical and temperate as the sample size of predictions would become too small otherwise.

As a preliminary check, the predictions can be roughly compared to the ESM outputs for the idealised AFF and DEF simulations as analysed by De Hertog et al. (2022). Only a rough comparison is possible however, as we generate predictions for tree cover change maps of constant values across grid points, whereas the tree cover change maps applied within the AFF/DEF scenarios vary in values across grid points since they represent full expansion of forest/cropland relative to the 2015 period. To this extent, $\Delta TS_{m,s}^{mean}$ predictions shown in Figure 6 correspond well in terms of direction and magnitude to the results shown in De Hertog et al. (2022). Moreover, Γ_m^{mean} is notably able to capture the inter-ESM spread in $\Delta TS_{m,s}^{mean}$ values. For example, in the latitudinal band 40°N to 65°N, EC-EARTH based predictions show a cooling trend after +25% tree cover change, in contrast to the warming trend seen in other ESMs. Such a difference was also noted in (De Hertog et al., 2022) and attributed to lower amounts of boreal afforestation implemented.

Over all latitudinal bands and months shown, largest 95% intervals occur towards the extreme ends of tree cover change for both observations and ESM based predictions. This is especially the case for deforestation, where the 95% intervals are in general larger than those of afforestation. Higher 95% intervals at extreme ends of tree cover change results from less grid points which undergo more extreme tree cover changes, ergo less training material.

Mean observation based predictions remain close to 0 across all latitudinal bands for both January and July, owing to the high data sparsity which makes it difficult to extract significant $TS_{m,s}^{mean}$ responses during training. Nonetheless, observation based 95% intervals are in general agreement with those of ESMs across all latitudinal bands and months shown.

4.3 Surface to 2m air temperature diagnosis

In this section, we apply the modified Hooker et al. (2018) model (Equation 3) to the outputs of Γ_m^{min} and Γ_m^{max} so as to derive the expected $T_{m,s}^{2m}$ responses to tree cover change. Results are shown for the select ESM, CESM2, as it covers the whole range of possible $\Delta_{2015}treeFrac$ (unlike observations and EC-EARTH) and provides local $TS_{m,s}^{min/max}$ values (not available from MPI-ESM otherwise). We first ascertain that applying Equation 3 in the ESM space does not introduce additional biases to $T_{m,s}^{2m}$ predictions ontop of those arising from the natural variability in observed values, after which we proceed with predicting $T_{m,s}^{2m}$ responses based off $\Gamma_m^{min/max}$ outputs.
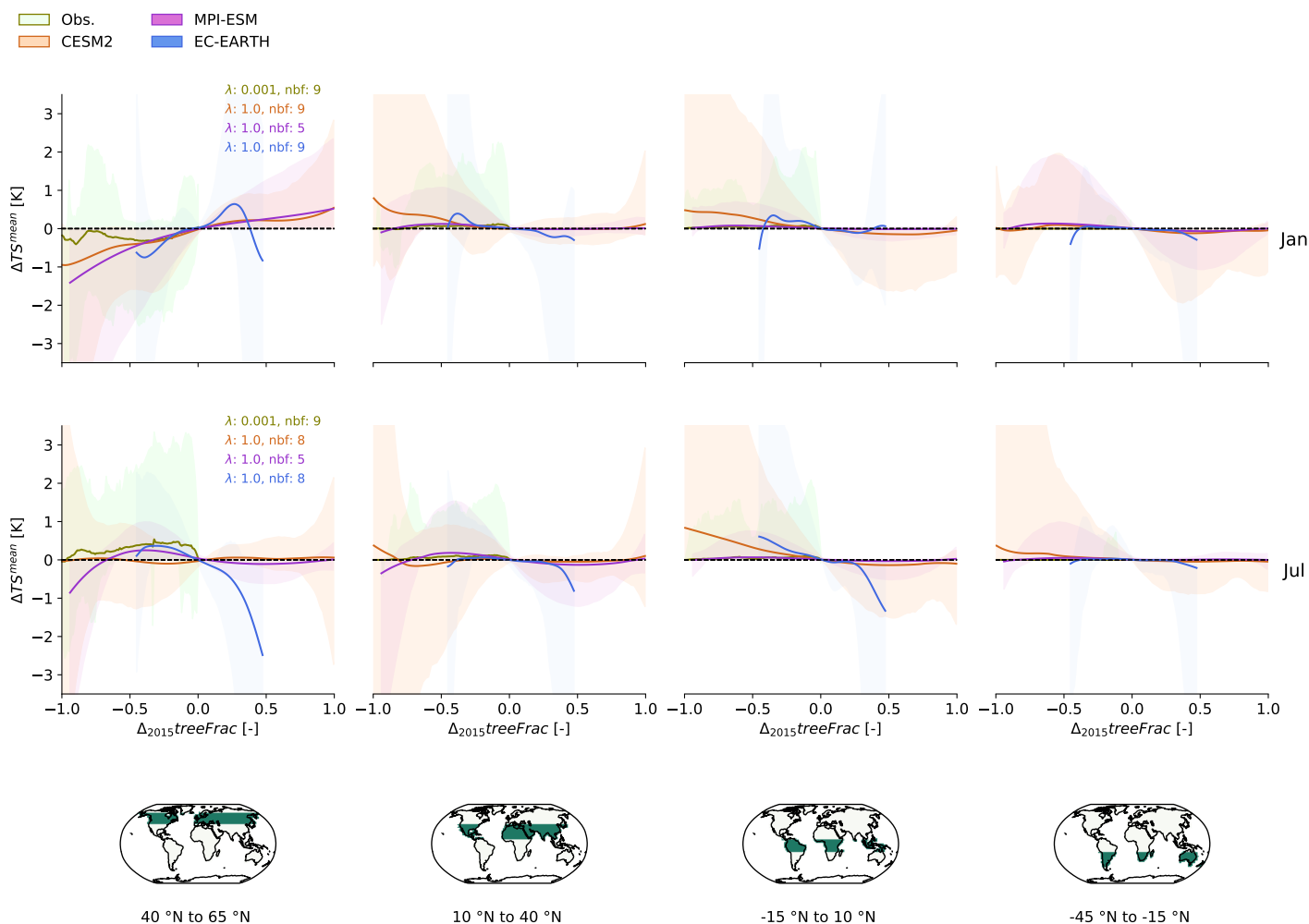
Figure 6. Γ_m^{mean}'s depiction of $\Delta TS_{m,s}^{mean}$ shown for observations and ESMs (colours) at months of January (first row) and July (second row) across the whole range of $\Delta_{2015}treeFrac$ and aggregated to latitudinal bands (columns). λ parameters and number of basis functions (nbf) chosen through blocked cross validation are given in the first column of their respective month and colour coded according to their respective training data (observations or ESMs). Solid lines represent the mean $\Delta TS_{m,s}^{mean}$ predictions and the surrounding band represents the 95% interval calculated over $\Delta TS_{m,s}^{mean}$ predictions for all grid-points within the respective latitudinal band.

4.3.1   Tests on the modified Hooker et al. (2018) model applied to the ESM space

Figure 7 compares the spread of biases in $T^{2m}_{m,s}$ calculated using ESM values to that obtained when using observational values. Positive values indicate more spread within the biases of ESM derived $T^{2m}_{m,s}$ values, suggesting that biases outside the range of those arising from natural, observational variability may occur when calculating $\Delta T^{2m}_{m,s}$.

Across most months, less than 40% of grid points have positive values, and these mostly occur in the Northern Hemisphere for the months between and including January and June. Such may result from the change in length of day during these months such that $TS^{min/max}_{m,s}$ values do not necessarily correspond to the $TS^{night/day}_{m,s}$ values. To be specific, the time of overpass for measuring $TS^{night}_{m,s}$ and $TS^{night}_{m,s}$ are fixed at 0100 and 1300 respectively, however given the longer nights in Northern-hemispheric winters, $TS^{min}_{m,s}$ are likely to occur later and $TS^{max}_{m,s}$ earlier than these times.

4.3.2   2m air temperature diagnoses

Since less than half of grid points have positive values and such values are isolated to certain months and geographical areas, we proceed with diagnosing $\Delta T^{2m}_{m,s}$ from $\Delta TS^{min/max}_{m,s}$ values outputted by $\Gamma^{min/max}$. The calibration and evaluation results for $\Gamma^{min/max}$ are available in Appendix B and show similar results as those seen in Section 4.1, with minimal additional RMSEs when predicting into "no-analogue" conditions sampled out of the training dataset as compared to when predicting after having seen the full training dataset (i.e. comparing RMSE values from Figures B3 and B4 to those of Figure B2).

Figure 8 shows the $\Delta T^{2m}_{m,s}$ values obtained at different tree cover change values, alongside the Γ^{min}_m and Γ^{max}_m predictions for example months of January and July. Patterns of Γ^{min}_m and Γ^{max}_m predictions correspond well to one another and generally well to $\Delta TS^{min/max}_{m,s}$ values as derived in other studies (Meier et al., 2018). An exception here are Northern Hemispheric, July $\Delta TS^{max}_{m,s}$ values for which a cooling was observed in Meier et al. (2018) in contrast to the warming seen in the training material used within this study (see Appendix B, Figure B1). Such discrepancy could arise from too large albedo responses shown by CESM2 and highlights the caveats of diagnosing $\Delta T^{2m}_{m,s}$ from $\Delta TS^{min/max}_{m,s}$, where physical inconsistencies in the surface temperature responses as represented within ESMs can be transferred to $T^{2m}_{m,s}$ during diagnosis. Nevertheless, the task of Γ_m is to mimic ESM outputs irrespective of their 'realism' and to this end, the statistically derived relationships for $\Delta TS^{min/max}_{m,s}$ to tree cover changes match those of the ESM outputs trained on.
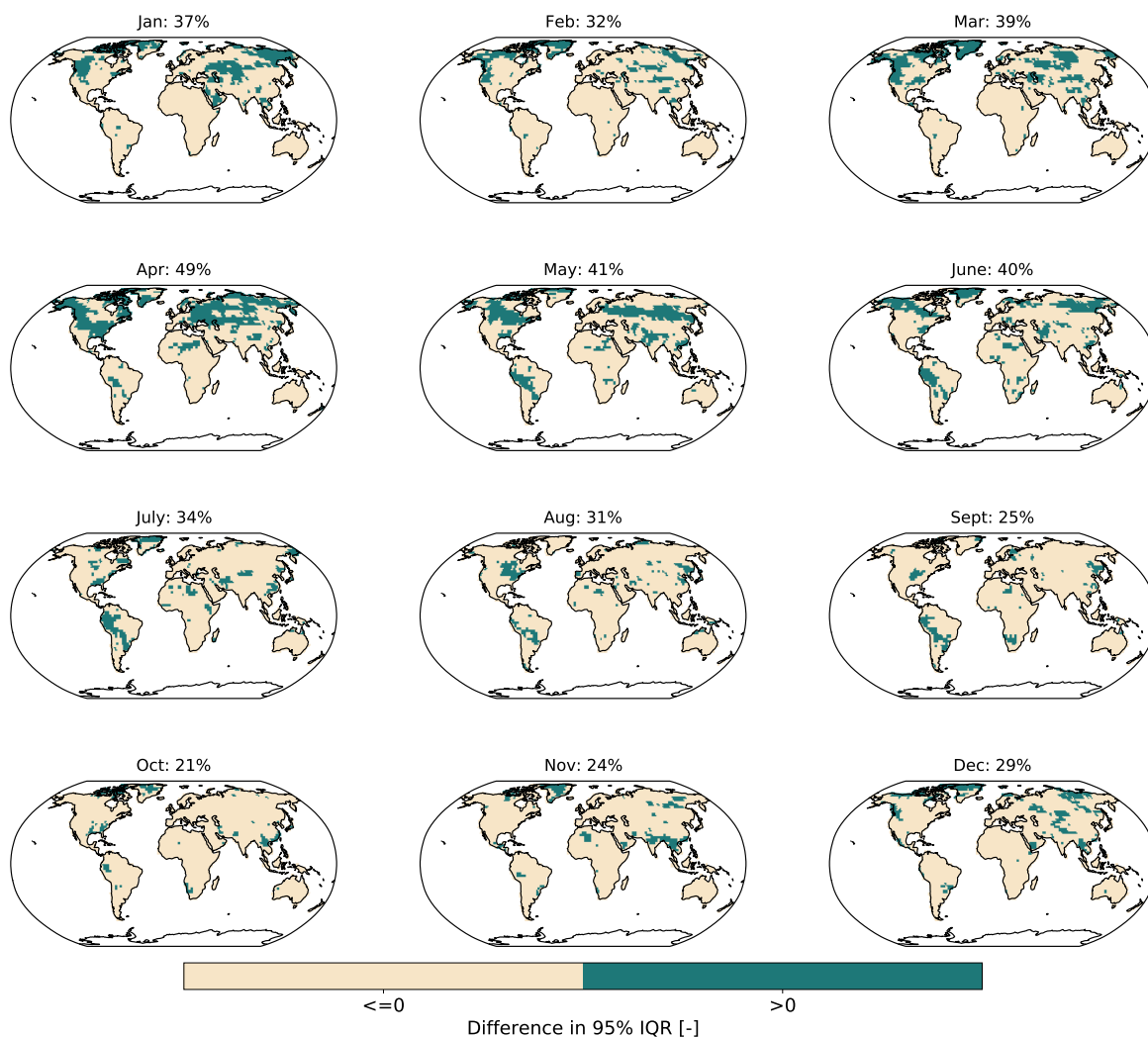
Figure 7. Differences between the spread of biases for ESM vs observationally derived $T_{m,s}^{2m}$ values, obtained as described in Section 3.3. The 95% Inter-quartile range (IQR) is considered as a measure of spread and results are shown for CESM2 across all months. Percentage values indicate the proportion of land grid points where ESM-derived $T_{m,s}^{2m}$ values have a larger spread in bias as compared to observationally derived $T_{m,s}^{2m}$ values.
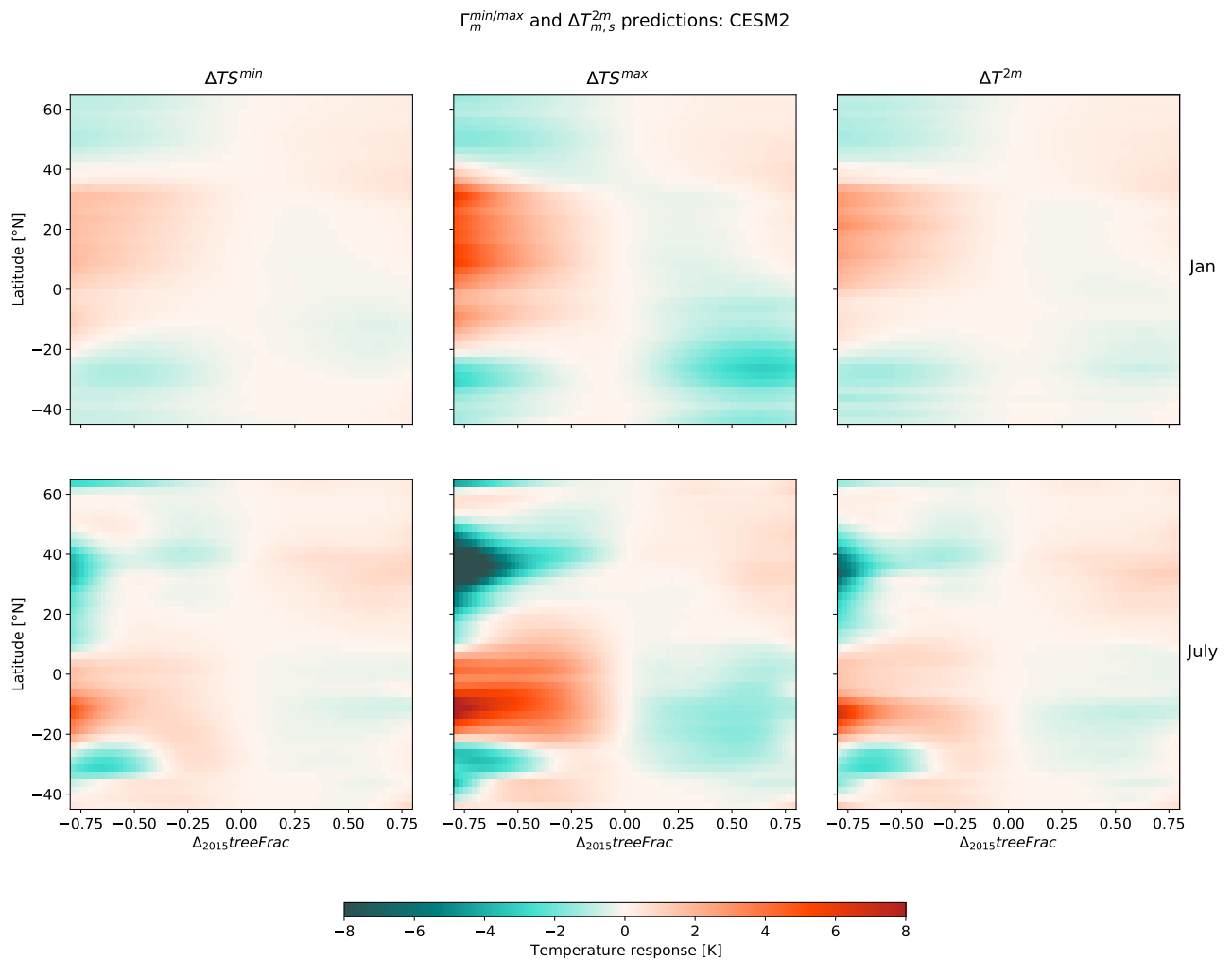
Figure 8. Latitudinally aggregated $\Delta TS_{m,s}^{min/max}$ given by $\Gamma_m^{min/max}$ (first two columns) shown for CESM2 at months of January (first row) and July (second row) across the full range of $\Delta_{2015} treeFrac$. The resulting $\Delta T_{m,s}^{2m}$ values obtained using the modified Hooker et al. (2018) model are shown in the third column.

5   Exploration of tree cover change effects within SSP scenarios

350   In this section, we showcase TIMBER v0.1 applied to tree cover changes implemented under SSP scenarios 1-2.6 and 3-7.0. Results are shown for the select ESM, CESM2, and we employ the sampling method as described in Section 3.4 such that parametric uncertainties within the GAM are also represented. This provides a first step towards statistically predicting $T_{m,s}^{2m}$ responses to tree cover change, in a manner that not only provides the expected response, but also gives an idea of the signal-to-noise ratio within predictions.

355   Figure 9 shows maps of end-of-century tree cover changes (shown in the first column) under SSP 1-2.6 and SSP 3-7.0 and their associated mean $T_{m,s}^{2m}$ responses (second column), obtained by sampling $\Delta TS_{m,s}^{min/max}$ values from $\Gamma_{m}^{min/max}$, applying Equation 3 to get $\Delta T_{m,s}^{2m}$ and taking its sample average. The signal-to-noise ratio is furthermore given by taking the ratio between the absolute mean $\Delta T_{m,s}^{2m}$ value and its standard deviation calculated across sample results for $\Delta T_{m,s}^{2m}$ (third column). We consider areas with a signal-to-noise ratio lower than 0.5 as having an insignificant temperature response, as their

360   surrounding parametric uncertainty is double that of the magnitude of response.

SSP 1-2.6 shows significant cooling from afforestation in Southern Africa and Brazil for both January and July. A significant July warming due to deforestation can also be seen in the Tibetian region due to deforestation. SSP 3-7.0 shows a significant January and July warming due to deforestation in Central Africa, Tibetian region and South America. West North America shows a significant cooling from deforestation especially in July, while parts of East Asia show significant cooling from

365   afforestation for both January and July.

In general, areas with a tree cover change lower than 0.1 in magnitude tend to have a signal-to-noise ratio lower than 0.5 and thus an insignificant temperature response. Such systematically lower signal-to-noise ratios indicates that Γ_{m} is not only aware of the lack of information it has for smaller changes in tree cover, but can also infer that temperature responses to such tree cover changes are likely to be trivial.
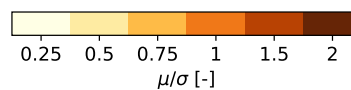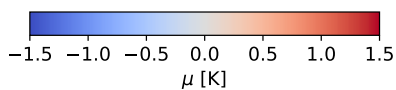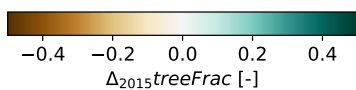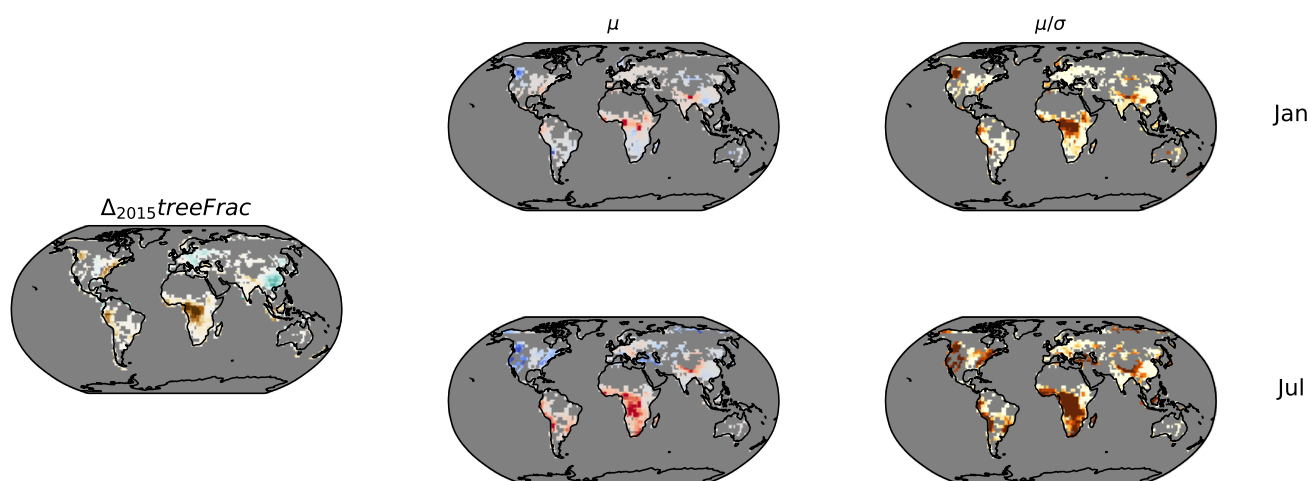
Responses of T_m^{2m} to tree cover change under SSP scenarios: CESM2
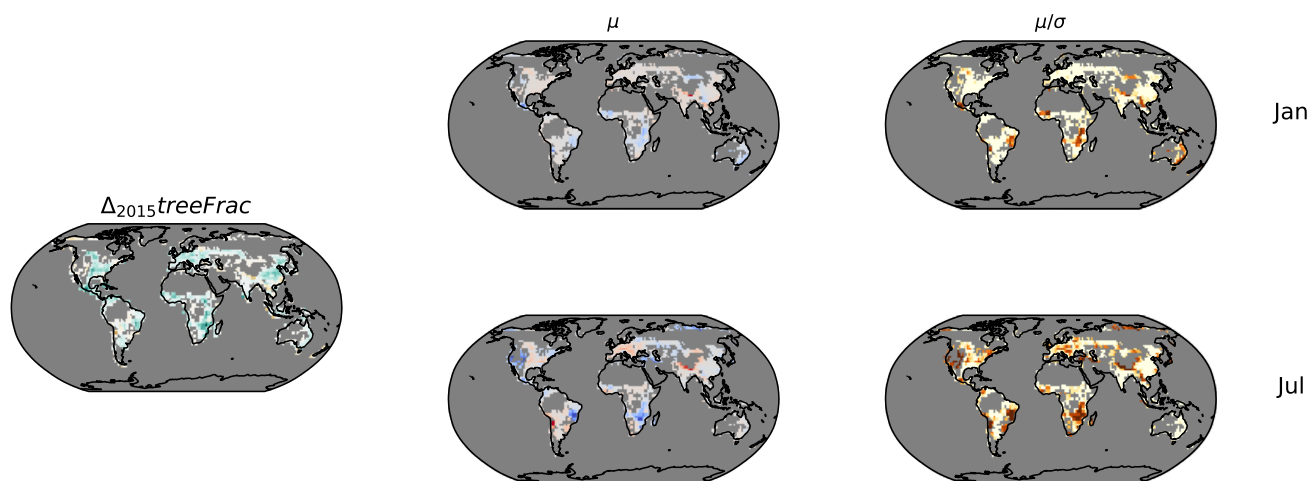
Figure 9. $\Delta T_{m,s}^{2m}$ values resulting from end-of-century changes in tree cover for SSP 1-2.6 (upper panel) and SSP 3-7.0 (lower panel) scenarios at the months of January (top rows) and July (bottom rows). Mean $\Delta T_{m,s}^{2m}$ values (second column) as well as their signal-to-noise ratios (third column) calculated over the sampling distributions are shown. $\Delta_{2015}treeFrac$ maps are given in the first column, grid points with $|\Delta_{2015}treeFrac| < 0.01$ are not considered.

6 Conclusion and Outlook

This study presents TIMBER v0.1, a conceptual framework for representing monthly temperature responses to changes in tree cover. TIMBER v0.1 starts by modelling minimum, mean and maximum surface temperature responses to tree cover change with a month-specific GAM which is trained over the whole globe. 2-m air temperature responses are then diagnosed from the modelled minimum and maximum surface temperatures using observational relationships derived by Hooker et al. (2018). Such an approach maintains the ESM-specific temperature response to tree cover change, whilst ensuring a constant diagnosis and observationally consistent definition of 2-m air temperature.

The GAM is evaluated for its ability to predict into unseen, i.e. "no-analogue", background climate as well as tree cover change conditions. Due to lack of additional ESM experiments which separate local and non-local temperature responses to tree cover change (Winckler et al., 2017), evaluation is conducted by means of blocked cross validation on the available training data. Overall, the GAM shows good skill in predicting into "no-analogue" conditions, with minimal additional RMSEs to those occurring when predicting after having seen the full training dataset and thus all available background climate and tree cover change information. Such provides confidence in the GAM's ability to derive meaningful relationships from the training data provided by the ESMs. Nevertheless, poorer representation for extreme, localised tree cover changes – such as deforestation in the tropics – was identified, most likely due to difficulty in adequately representing high spatial variability.

When predicting into new tree cover change scenarios, we are especially mindful of the training data only including grid points which experience extreme tree cover change in the training simulations. To this extent, surface temperature responses are sampled from the GAM, in a manner that explores all possible shapes of responses in between the two extreme ends of tree cover change as provided by the training data. 2-m air temperature responses are then diagnosed from the sampled surface temperature responses and relevant responses are identified as those having a high signal-to-noise ratio (>0.5).

The final outputs of TIMBER v0.1 are demonstrated for SSP 1-2.6 and SSP 3-7.0. Generally, areas with less than ±10% of tree cover change render a low signal-to-noise ratio, which is intuitive as responses to such low changes in tree cover are likely to be minimal. Employing TIMBER v0.1 thus provides avenue to explore impacts of tree cover change and their underlying uncertainty due to availability of training data and model calibration. In the following subsections, we further highlight areas of potential improvement, elaborate upon the suitable modes of application for TIMBER v0.1 and detail possible further developments.

6.1 Areas of potential improvement

One area of potential improvement pertains to the model calibration procedure. When inspecting the calibrated λ parameter values and number of basis functions, the limits of values cross validated for (0.001 and 1 for the λ parameter and 5 and 9 for the number of bases functions) seem to be favoured. Reasons behind this could be: (1) the blocked cross validation sometimes removes too large chunks of data, leading to an overestimation of RMSEs chosen, and/or (2) the range of λ parameter/number of basis functions values calibrated for is too narrow. The first reason could be tackled by further splitting the blocks such that each block has a predefined number of samples. Alternatively, the GAM could be fitted over specific climate regions, and

blocked cross validation conducted with uniformly sized blocks composed along latitude and longitude dimensions; although here it is likely that the complete spectrum of tree cover change information will be lost for some regions. The second reason is easily solved by cross validating over a larger range of values.

Another area of improvement could be to derive ESM-specific coefficients for the Hooker et al. (2018) model. Such would entail fitting for the relationships between ESM surface temperatures and the observational 2-m air temperatures as used by Hooker et al. (2018). Since the additional biases introduced by using the original Hooker et al. (2018) coefficients on the ESM surface temperatures were ascertained as minimal (Section 4.3.1), such an exercise would mostly target deriving the complete Hooker et al. (2018) model for each ESM. The resultant ESM-specific Hooker et al. (2018) models obtained would allow for more consistent 2-m air temperature diagnoses facilitating better comparison.

6.2 Modes of application

TIMBER v0.1 provides a framework to agilely explore local-level, temperature implications of tree cover changes under different tree cover change scenarios. TIMBER v0.1 can be used as both a standalone device as well as supplementary to other emulators. Here, we summarise some key take-aways pertaining to the use of TIMBER v0.1 for generating new tree cover change scenarios.

Upon inspection of the $TS_{m,s}^{mean}$ response patterns across all tree cover changes (Figure 6), inter-ESM differences become quite apparent. Such differences are continuously studied and mainly arise from differences in model physical representation (Boisier et al., 2012; Lawrence et al., 2016; Lejeune et al., 2018; Davin et al., 2020; Boysen et al., 2020; De Hertog et al., 2022). Being able to train the GAM across all ESMs presents the opportunity to capture these uncertainties due to model physical representation, which may sometimes be higher than the parametric uncertainty within the GAM given the training data. When exploring new tree cover change scenarios, the need to have as many ESMs represented should therefore be emphasised. Moreover, the outputs of TIMBER v0.1 should always be interpreted as representative of the ESM-simulated world which does not necessarily translate to observed reality.

In applying TIMBER v0.1 to different tree cover change and climate scenarios, it should furthermore be acknowledged that the effects of initial starting conditions and those of background global warming levels have not been accounted for. As we focus on modelling the local surface temperature responses from which 2-m air temperature responses are diagnosed, such omissions are expected to have little effect. Furthermore, if the Hooker et al. (2018) coefficients are recalibrated for the ESM space, impacts of changing climate on 2-m air temperatures could well be represented through the CSWR coefficients. Nonetheless, outputs of TIMBER v0.1 should more so be treated as hypothetical sensitivities and not definite responses.

6.3 Future Developments

It would be possible to extend tTIMBER v0.1 to represent other impact-relevant climate variables. A variable to start with could be relative humidity, from which metrics such as Wet Bulb Globe Temperature (WBGT) and labour productivity could be derived. In doing so, variable cross-correlations between temperature and relative humidity should be conserved, such that

435 compound events – which largely affect WBGTs – are sufficiently captured. To this extent, a Vectorised Generalised Additive Model (VGAM) (Yee and Stephenson, 2007) could be employed, which retains variable cross-correlations by constructing a multivariate conditional probability distribution e.g. by using a bi-normal distribution as opposed to the normal distribution used within this study. Another idea could be to couple TIMBER v0.1 to other emulators built to ingest temperature fields in order to generate additional climate variables. A suitable emulator for example could be PREMU (Liu et al., 2022), which

440 is able to derive the principle modes of spatial variability from temperature fields, in order to generate monthly precipitation fields.

Looking into other land management practices such as irrigation and wood harvest could also be of interest, particularly as their effects on surface temperatures are expected to be similar in magnitude as those due to land cover changes (Luyssaert et al., 2014). In doing so, customisation of TIMBER v0.1's framework to the LCLM practice of choice could be necessary.

445 For example, when looking at irrigation, implementation of irrigation can be extremely localised and seasonal and it would be preferable to train the GAM as region-specific and across all months, instead of month-specific and across all grid-points. To this extent, the GAM has the advantage of not prescribing any functional form, giving it flexibility in deriving climate responses to different types of LCLM forcings regardless of the format of the training data.

In order to jointly explore future tree cover and GHG scenarios, coupling TIMBER v0.1 with other temperature emulators

450 such as MESMER-M or -X (Beusch et al., 2020; Nath et al., 2022b; Quilcaille et al., 2022) also proves worthwhile. In doing so, care would have to be taken to not "double-count" the tree cover change signal as MESMER-M and -X are trained on SSP runs, which contain both GHG and tree cover change signals. Accordingly, it is advisable to first model the expected tree cover change signals within the SSP runs using TIMBER v0.1, following which MESMER-M or -X can be trained on the SSP runs with the modelled tree cover change signals removed.

Author contributions. QL, CFS and SIS identified the need for an emulator framework applied towards representing land cover forced climate responses. SN developed the emulator set up and calibration/evaluation procedure with help from QL, LG and JS. QL and GD furthermore

460 assisted SN in integrating the observationally based, 2-m air temperature diagnosis scheme within the emulator. All authors contributed to interpreting results and streamlining the text.
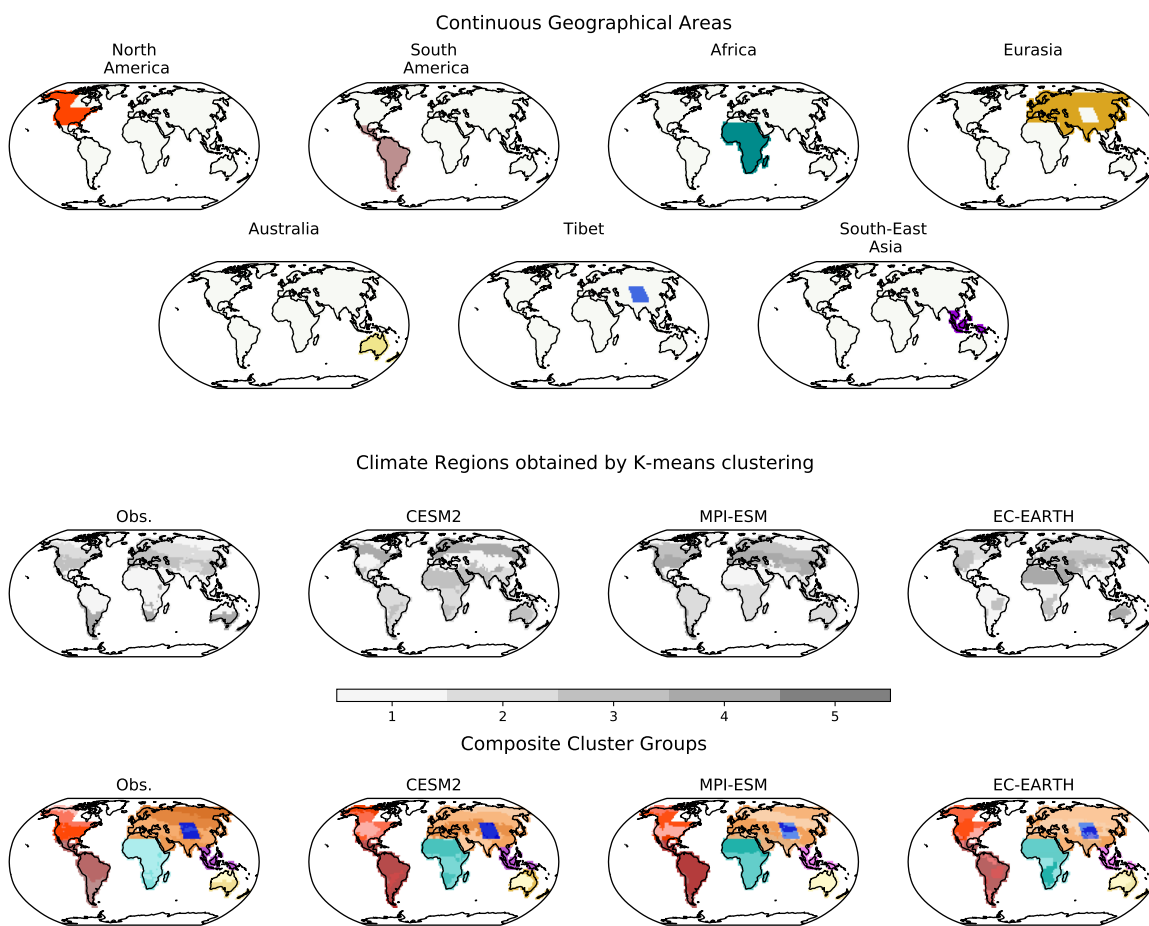
Competing interests. The authors declare that they have no conflict of interest

470   # Appendix A

Figure A1. Composite cluster blocks obtained by combining clusters of grid points with similar background climate and continuous geographical area. Grid points are clustered into groups with similar background climate using K-means clustering with temperature and relative humidity as indicator variables.
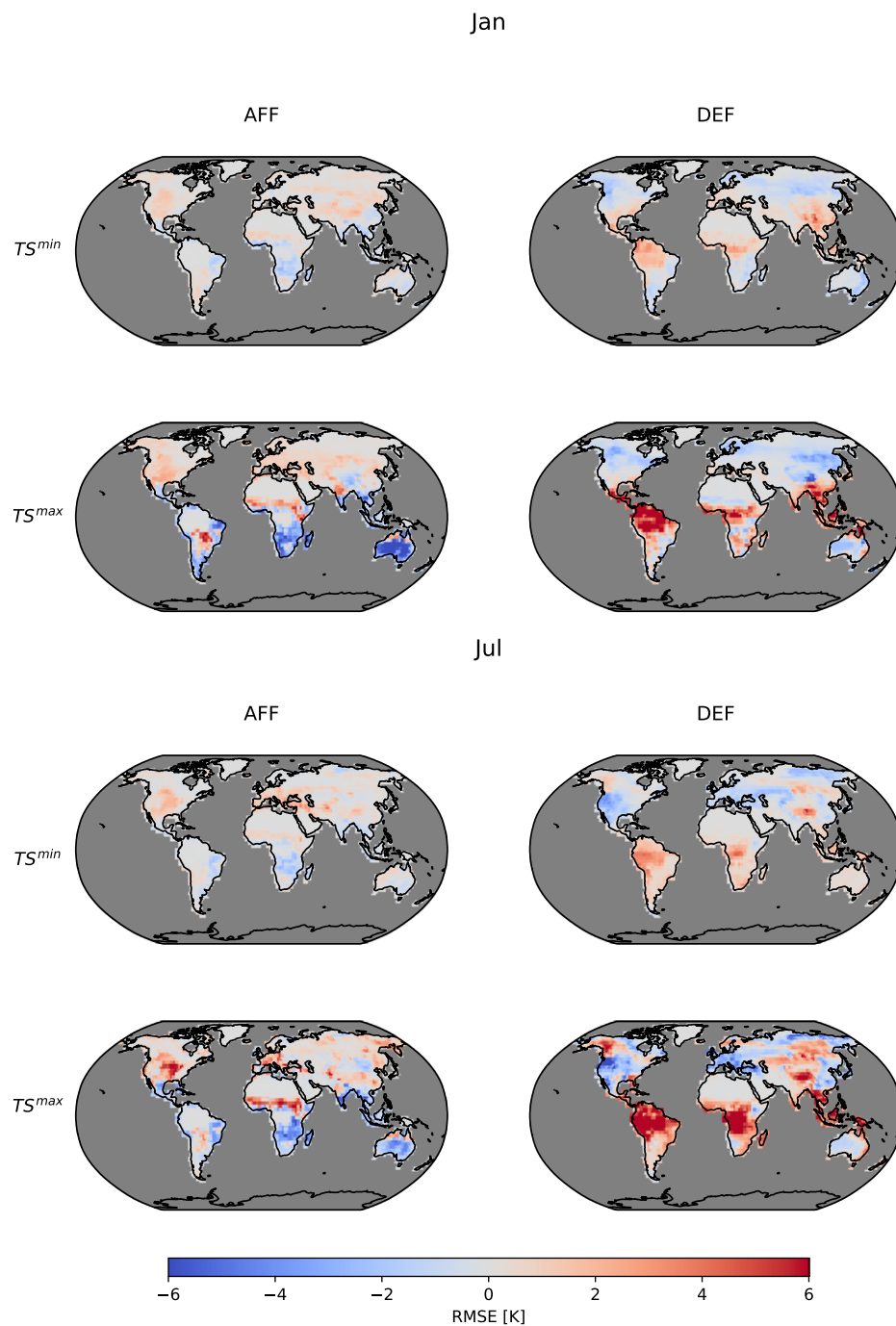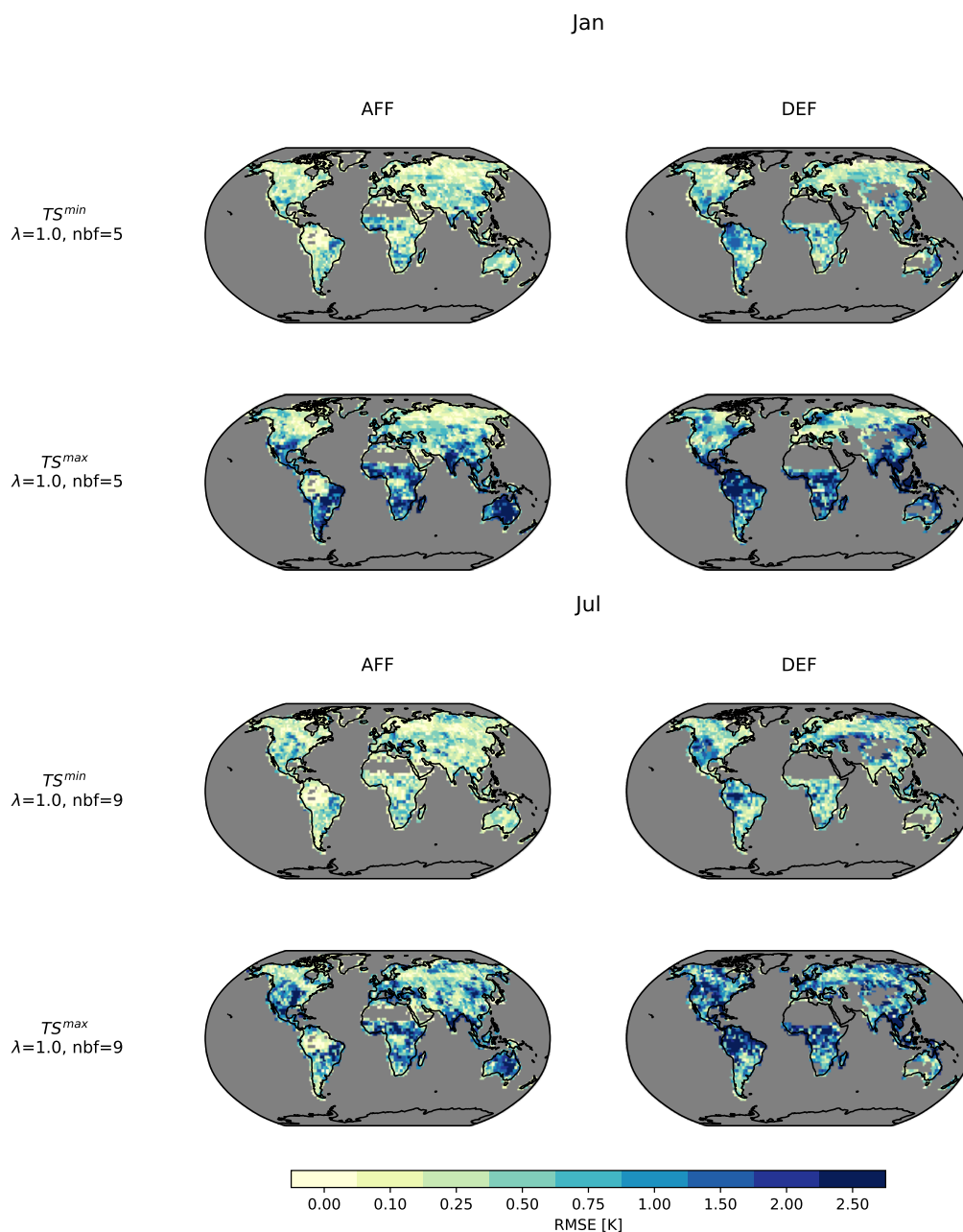
Appendix B
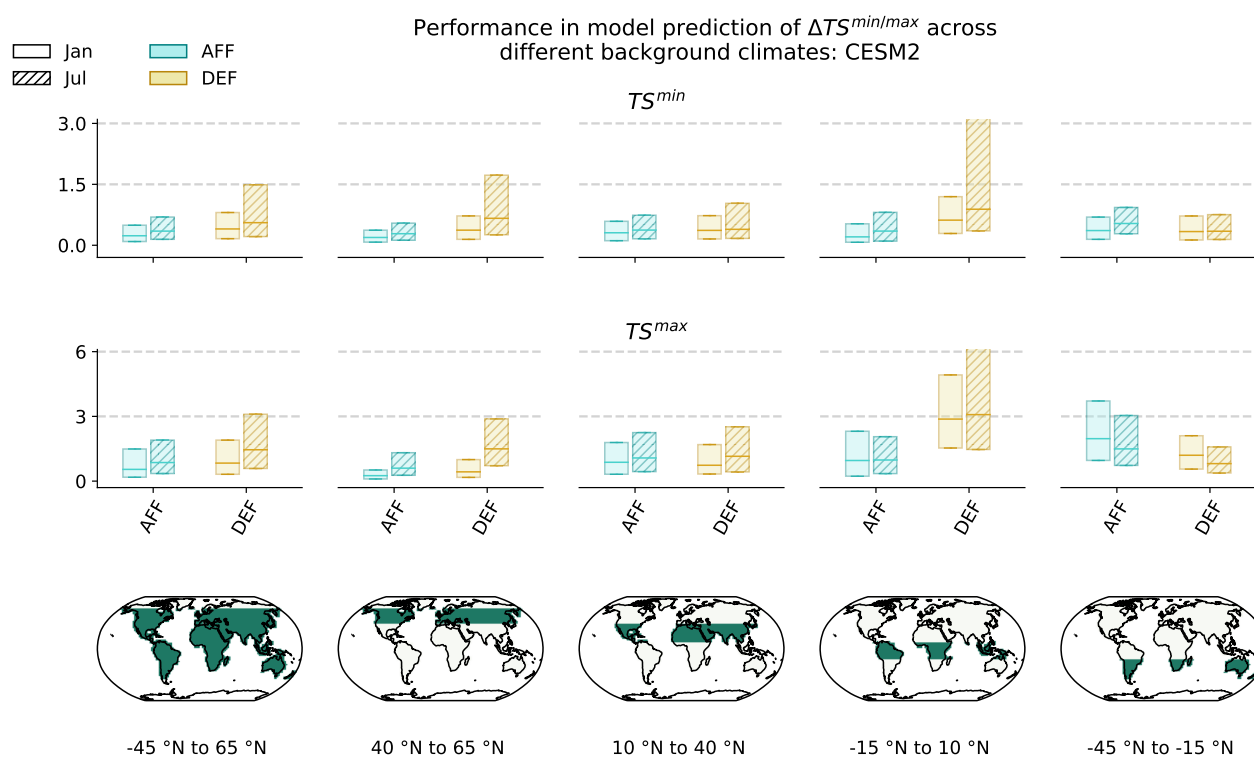
TS^{min} and TS^{max} training outputs: CESM2

Figure B1. $TS_{m,s}^{min/max}$ responses (rows) from the LAMACLIMA AFF and DEF experiments (columns) for the months of January (upper panel) and July (lower panel)
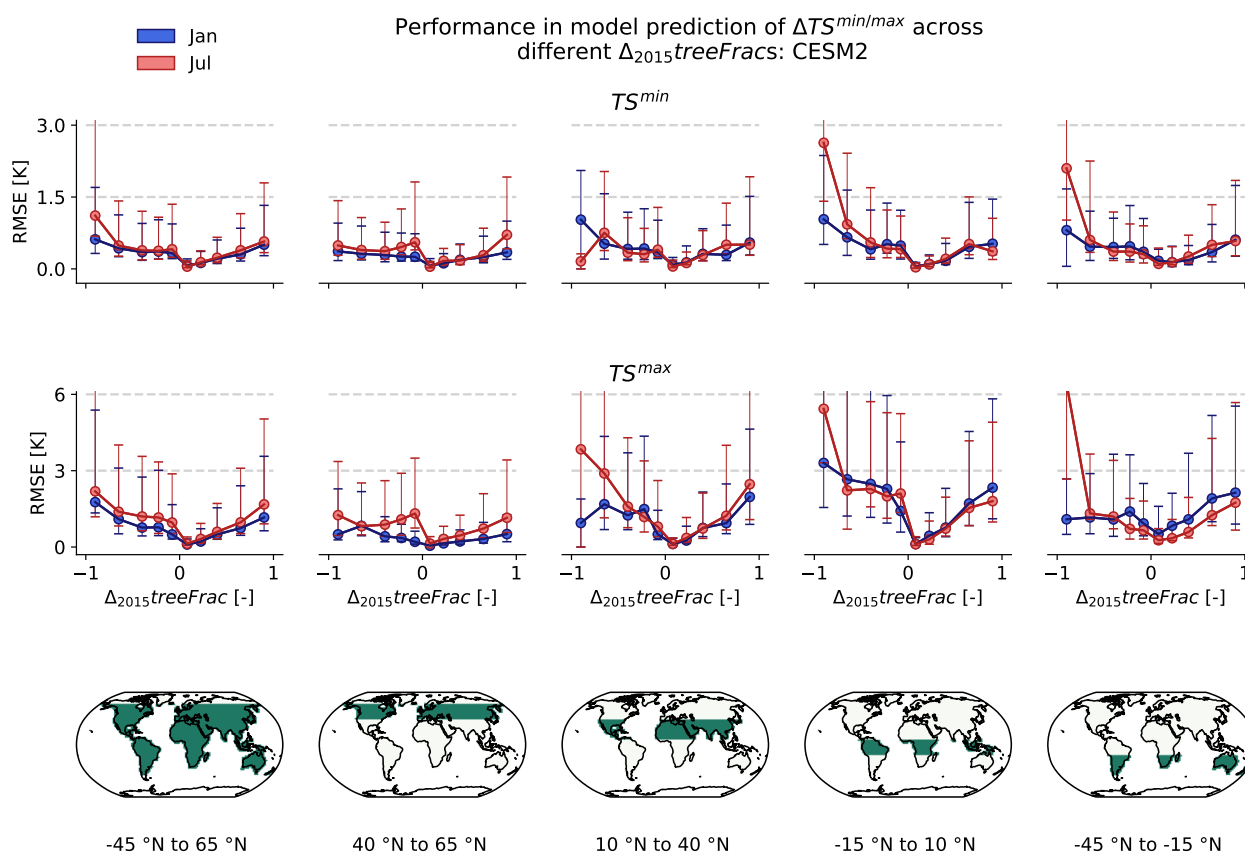
29

Figure B2. Same as Figure 3 but for CESM2, TS^{min} and TS^{max}

Figure B3. Same as Figure 4 but for CESM2, TS^{min} and TS^{max}
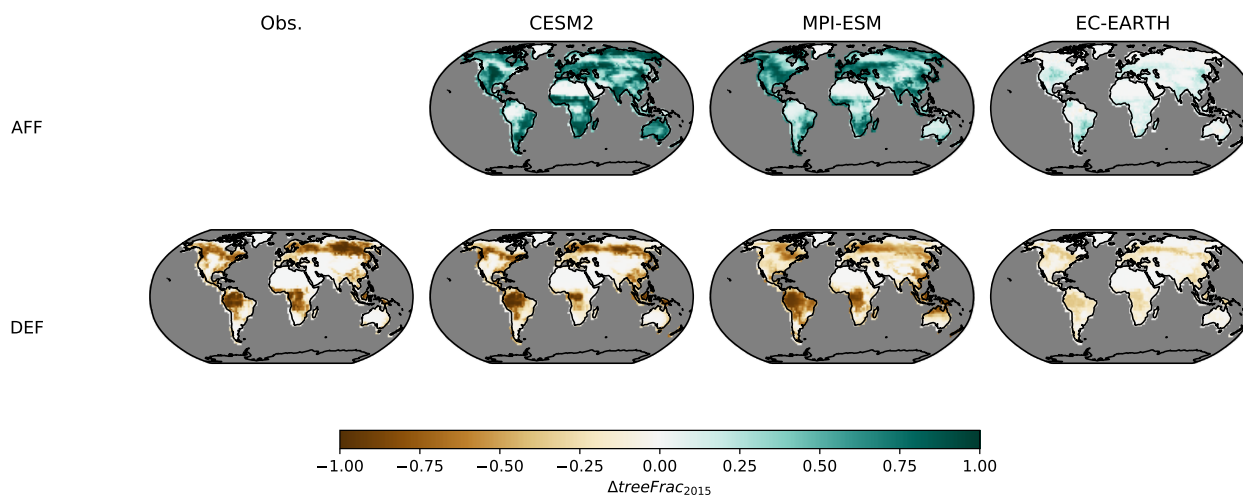
Figure B4. Same as Figure 5 but for CESM2, TS^{min} and TS^{max}

Appendix C

Figure C1. Tree cover changes relative to the 2015 period implemented in the LAMACLIMA AFF (top row) and DEF (bottom row) experiments

References

Glasgow leaders' declaration on forest and land-use, https://ukcop26.org/glasgow-leaders-declaration-on-forests-and-land-use/, 2021.

475 Alexeeff, S. E., Nychka, D., Sain, S. R., and Tebaldi, C.: Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments, Climatic Change, 146, 319–333, https://doi.org/10.1007/s10584-016-1809-8, 2018.

Beusch, L., Gudmundsson, L., and Seneviratne, S. I.: Emulating Earth system model temperatures with MESMER: from global mean temperature trajectories to grid-point-level realizations on land, Earth System Dynamics, 11, 139–159, https://doi.org/10.5194/esd-11-139-2020, 2020.

480 Boisier, J. P., De Noblet-Ducoudré, N., Pitman, A. J., Cruz, F. T., Delire, C., Van Den Hurk, B. J., Van Der Molen, M. K., Mller, C., and Voldoire, A.: Attributing the impacts of land-cover changes in temperate regions on surface temperature and heat fluxes to specific causes: Results from the first LUCID set of simulations, Journal of Geophysical Research Atmospheres, 117, 1–16, https://doi.org/10.1029/2011JD017106, 2012.

Boysen, L. R., Brovkin, V., Pongratz, J., Lawrence, D. M., Lawrence, P., Vuichard, N., Peylin, P., Liddicoat, S., Hajima, T., Zhang, Y., Rocher,
485 M., Delire, C., Séférian, R., Arora, V. K., Nieradzik, L., Anthoni, P., Thiery, W., Laguë, M. M., Lawrence, D., and Lo, M. H.: Global climate response to idealized deforestation in CMIP6 models, Biogeosciences, 17, 5615–5638, https://doi.org/10.5194/bg-17-5615-2020, 2020.

Calvin, K. and Bond-Lamberty, B.: Integrated human-earth system modeling - State of the science and future directions, Environmental Research Letters, 13, https://doi.org/10.1088/1748-9326/aac642, 2018.

Castruccio, S., Hu, Z., Sanderson, B., Karspeck, A., and Hammerling, D.: Reproducing internal variability with few ensemble runs, Journal
490 of Climate, 32, 8511–8522, https://doi.org/10.1175/JCLI-D-19-0280.1, 2019.

Davin, E. L., Rechid, D., Breil, M., Cardoso, R. M., Coppola, E., Hoffmann, P., Jach, L. L., Katragkou, E., de Noblet-Ducoudré, N., Radtke, K., Raffa, M., Soares, P. M. M., Sofiadis, G., Strada, S., Strandberg, G., Tölle, M. H., Warrach-Sagi, K., and Wulfmeyer, V.: Biogeophysical impacts of forestation in Europe: first results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison, Earth System Dynamics, 11, 183–200, https://doi.org/10.5194/esd-11-183-2020, 2020.

495 De Hertog, S. J., Havermann, F., Vanderkelen, I., Guo, S., Luo, F., Manola, I., Coumou, D., Davin, E. L., Duveiller, G., Lejeune, Q., Pongratz, J., Schleussner, C.-F., Seneviratne, S. I., and Thiery, W.: The biogeophysical effects of idealized land cover and land management changes in Earth System Models, Earth System Dynamics Discussions, pp. 1–53, https://doi.org/10.5194/esd-2022-5, 2022.

De Noblet-Ducoudré, N., Boisier, J. P., Pitman, A., Bonan, G. B., Brovkin, V., Cruz, F., Delire, C., Gayler, V., Van Den Hurk, B. J., Lawrence, P. J., Van Der Molen, M. K., Müller, C., Reick, C. H., Strengers, B. J., and Voldoire, A.: Determining robust impacts of land-use-induced
500 land cover changes on surface climate over North America and Eurasia: Results from the first set of LUCID experiments, Journal of Climate, 25, 3261–3281, https://doi.org/10.1175/JCLI-D-11-00338.1, 2012.

Duveiller, G., Hooker, J., and Cescatti, A.: The mark of vegetation change on Earth's surface energy balance, Nature Communications, 9, https://doi.org/10.1038/s41467-017-02810-8, 2018a.

Duveiller, G., Hooker, J., and Cescatti, A.: A dataset mapping the potential biophysical effects of vegetation cover change, Scientific Data,
505 5, 1–15, https://doi.org/10.1038/sdata.2018.14, 2018b.

Duveiller, G., Hooker, J., and Cescatti, A.: A dataset mapping the potential biophysical effects of vegetation cover change. figshare, https://doi.org/https://doi.org/10.6084/m9.figshare.c.3829333.v1, 2018c.

Efron, B. and Tibshirani, R. J.: An Introduction to the Bootstrap, https://doi.org/10.1007/978-1-4899-4541-9, 1993.

Hirsch, A. L., Guillod, B. P., Seneviratne, S. I., Beyerle, U., Boysen, L. R., Brovkin, V., Davin, E. L., Doelman, J. C., Kim, H.,
510   Mitchell, D. M., Nitta, T., Shiogama, H., Sparrow, S., Stehfest, E., van Vuuren, D. P., and Wilson, S.: Biogeophysical Impacts
       of Land-Use Change on Climate Extremes in Low-Emission Scenarios: Results From HAPPI-Land, Earth's Future, 6, 396–409,
       https://doi.org/10.1002/2017EF000744, 2018.

Hooker, J., Duveiller, G., and Cescatti, A.: Data descriptor: A global dataset of air temperature derived from satellite remote sensing and
       weather stations, Scientific Data, 5, 1–11, https://doi.org/10.1038/sdata.2018.246, 2018.

515   Lawrence, D., Coe, M., Walker, W., Verchot, L., and Vandecar, K.: The Unseen Effects of Deforestation: Biophysical Effects on Climate,
       Frontiers in Forests and Global Change, 5, 1–13, https://doi.org/10.3389/ffgc.2022.756115, 2022.

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., Noblet-Ducoudré, N. D.,
       Pongratz, J., Seneviratne, S. I., and Shevliakova, E.: The Land Use Model Intercomparison Project (LUMIP) contribution to CMIP6:
       Rationale and experimental design, Geoscientific Model Development, 9, 2973–2998, https://doi.org/10.5194/gmd-9-2973-2016, 2016.

520   Lejeune, Q., Davin, E. L., Gudmundsson, L., Winckler, J., and Seneviratne, S. I.: Historical deforestation locally increased the intensity of
       hot days in northern mid-latitudes, Nature Climate Change, 8, 386–390, https://doi.org/10.1038/s41558-018-0131-z, 2018.

Link, R., Snyder, A., Lynch, C., Hartin, C., Kravitz, B., and Bond-Lamberty, B.: Fldgen v1.0: An emulator with internal variability and space-
       Time correlation for Earth system models, Geoscientific Model Development, 12, 1477–1489, https://doi.org/10.5194/gmd-12-1477-2019,
       2019.

525   Liu, G., Peng, S., Huntingford, C., and Xi, Y.: A new precipitation emulator (PREMU v1 . 0) for lower complexity models, 2022.

Luyssaert, S., Jammet, M., Stoy, P. C., Estel, S., Pongratz, J., Ceschia, E., Churkina, G., Don, A., Erb, K., Ferlicoq, M., Gielen, B., Grünwald,
       T., Houghton, R. A., Klumpp, K., Knohl, A., Kolb, T., Kuemmerle, T., Laurila, T., Lohila, A., Loustau, D., McGrath, M. J., Meyfroidt, P.,
       Moors, E. J., Naudts, K., Novick, K., Otto, J., Pilegaard, K., Pio, C. A., Rambal, S., Rebmann, C., Ryder, J., Suyker, A. E., Varlagin, A.,
       Wattenbach, M., and Dolman, A. J.: Land management and land-cover change have impacts of similar magnitude on surface temperature,
530   Nature Climate Change, 4, 389–393, https://doi.org/10.1038/nclimate2196, 2014.

McKinnon, K. A. and Deser, C.: Internal variability and regional climate trends in an observational large ensemble, Journal of Climate, 31,
       6783–6802, https://doi.org/10.1175/JCLI-D-17-0901.1, 2018.

Meier, R., Davin, E. L., Lejeune, Q., Hauser, M., Li, Y., Martens, B., Schultz, N. M., Sterling, S., and Thiery, W.: Evaluating and improving
       the Community Land Model's sensitivity to land cover, Biogeosciences, 15, 4731–4757, https://doi.org/10.5194/bg-15-4731-2018, 2018.

535   Nath, S.: snath-xoc/TIMBER-v0.1_Nath_et_al_2022: TIMBER, https://doi.org/10.5281/zenodo.7261281, 2022.

Nath, S., Hertog, S. J. D., Guo, S., Havermann, F., Luo, F., Manola, I., Pongratz, J., and Thiery, W.:
       LAMACLIMA_experiments_for_training_TIMBERv0.1, https://doi.org/10.5281/zenodo.7261374, 2022a.

Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., and Schleussner, C. F.: MESMER-M: an Earth system model emulator for spatially
       resolved monthly temperature, Earth System Dynamics, 13, 851–877, https://doi.org/10.5194/esd-13-851-2022, 2022b.

540   Pitman, A., De Noblet-Ducoudré, N., Avila, F., Alexander, L., Boisier, J.-P., Brovkin, V., Delire, C., Cruz, F., Donat, M., Gayler, V., Hurk,
       B. v. d., Reick, C., and Voldoire, A.: Effects of land cover change on temperature and rainfall extremes in multi-model ensemble 3
       simulations, Earth System Dynamics, p. 213–231, 2012.

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., Bodirsky, B. L., Dietrich, J. P., Doelmann, J. C., Gusti, M.,
       Hasegawa, T., Kyle, P., Obersteiner, M., Tabeau, A., Takahashi, K., Valin, H., Waldhoff, S., Weindl, I., Wise, M., Kriegler, E., Lotze-
545   Campen, H., Fricko, O., Riahi, K., and Vuuren, D. P.: Land-use futures in the shared socio-economic pathways, Global Environmental
       Change, 42, 331–345, https://doi.org/10.1016/j.gloenvcha.2016.10.002, 2017.

Quilcaille, Y., Gudmundsson, L., Beusch, L., Hauser, M., and Seneviratne, S. I.: Showcasing MESMER-X: Spatially resolved emulation of annual maximum temperatures of Earth System Models, pp. 1–11, https://doi.org/10.1029/2022GL099012, 2022.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller,
550   W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2017.

Seddon, N., Sengupta, S., Hauler, I., and Rizvi, A. R.: Nature-based solutions in nationally determined contributions | IUCN Library System, https://portals.iucn.org/library/node/48525, 2020.

Seneviratne, S. I., Wartenburger, R., Guillod, B. P., Hirsch, A. L., Vogel, M. M., Brovkin, V., Van Vuuren, D. P., Schaller, N., Boysen, L.,
555   Calvin, K. V., Doelman, J., Greve, P., Havlik, P., Humpenöder, F., Krisztin, T., Mitchell, D., Popp, A., Riahi, K., Rogelj, J., Schleussner, C. F., Sillmann, J., and Stehfest, E.: Climate extremes, land-climate feedbacks and land-use forcing at 1.5C, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376, https://doi.org/10.1098/rsta.2016.0450, 2018.

Van Vuuren, D. P., Batlle Bayer, L., Chuwah, C., Ganzeveld, L., Hazeleger, W., Van Den Hurk, B., Van Noije, T., Oneill, B., and Strengers, B. J.: A comprehensive view on climate change: Coupling of earth system and integrated assessment models, Environmental Research
560   Letters, 7, https://doi.org/10.1088/1748-9326/7/2/024012, 2012.

Winckler, J., Reick, C. H., and Pongratz, J.: Robust identification of local biogeophysical effects of land-cover change in a global climate model, Journal of Climate, 30, 1159–1176, https://doi.org/10.1175/JCLI-D-16-0067.1, 2017.

Windisch, M. G., Davin, E. L., and Seneviratne, S. I.: Prioritizing forestation based on biogeochemical and local biogeophysical impacts, Nature Climate Change, 11, 867–871, https://doi.org/10.1038/s41558-021-01161-z, 2021.

565 Wood, S. N.: Generalized Additive Models, Chapman and Hall/CRC, https://doi.org/10.1201/9781315370279, 2017.

Yee, T. W. and Stephenson, A. G.: Vector generalized linear and additive extreme value models, Extremes, 10, 1–19, https://doi.org/10.1007/s10687-007-0032-4, 2007.