RESEARCH ARTICLE

Predicting plant species distributions using climate-based model ensembles with corresponding measures of congruence and uncertainty

Stephen B. Stewart¹ | Melissa Fedrigo²,³ | Sabine Kasel³ | Stephen H. Roxburgh⁴ | Kunzang Choden⁵ | Karma Tenzin³ | Kathryn Allen⁶,³,⁷ | Craig R. Nitschke³

¹Land and Water Business Unit, CSIRO, Sandy Bay, Tasmania, Australia
²GeoRubix Solutions, Hobart, Tasmania, Australia
³School of Ecosystem and Forest Sciences, University of Melbourne, Burnley, Victoria, Australia
⁴Land and Water Business Unit, CSIRO, ACT, Australia
⁵Bhutan for Life Fund Secretariat, Royal Textile Academy Complex, Thimphu, Bhutan
⁶Geography, Planning, Spatial Sciences, University of Tasmania, Sandy Bay, Tasmania, Australia
⁷Centre of Excellence for Australian Biodiversity and Heritage, University of New South Wales, New South Wales, Australia

Correspondence
Stephen B. Stewart, Land and Water Business Unit, CSIRO, Sandy Bay, TAS, Australia.
Email: stephen.stewart@csiro.au

Funding Information
The research was supported by the Melbourne Research Scholarship (University of Melbourne) with additional funding provided by the State of Victoria Department of Environment, Land, Water and Planning (DELWP) through the Integrated Forest Ecosystem Research (IFER) program

Abstract

Aim: The increasing availability of regional and global climate data presents an opportunity to build better ecological models; however, it is not always clear which climate dataset is most appropriate. The aim of this study was to better understand the impacts that alternative climate datasets have on the modelled distribution of plant species, and to develop systematic approaches to enhancing their use in species distribution models (SDMs).

Location: Victoria, southeast Australia and the Himalayan Kingdom of Bhutan.

Methods: We compared the statistical performance of SDMs for 38 plant species in Victoria and 12 plant species in Bhutan with multiple algorithms using globally and regionally calibrated climate datasets. Individual models were compared against each other and as SDM ensembles to explore the potential for alternative predictions to improve statistical performance. We develop two new spatially continuous metrics that support the interpretation of ensemble predictions by characterizing the per-pixel congruence and variability of contributing models.

Results: There was no clear consensus on which climate dataset performed best across all species in either study region. On average, multi-model ensembles (across the same species with different climate data) increased AUC/TSS/Kappa/OA by up to 0.02/0.03/0.03/0.02 in Victoria and 0.06/0.11/0.11/0.05 in Bhutan. Ensembles performed better than most single models in both Victoria (AUC = 85%; TSS = 68%) and Bhutan (AUC = 86%; TSS = 69%). SDM ensembles using models fitted with alternative algorithms and/or climate datasets each provided a significant improvement over single model runs.

Main conclusions: Our results demonstrate that SDM ensembles, built using alternative models of the same climate variables, can quantify model congruence and identify regions of the highest uncertainty while mitigating the risk of erroneous predictions. Algorithm selection is known to be a large source of error for SDMs, and our results...
demonstrate that climate dataset selection can be a comparably significant source of uncertainty.

KEYWORDS
Cantell's inequality, climate, congruence, ensemble, species distribution modelling, uncertainty

1 | INTRODUCTION

Species distribution models (SDMs) are among the most widely used tools in ecology. These models link the presence, abundance or vital rates (e.g., growth rate and fecundity) of species to environmental conditions. The conceptual and practical issues surrounding SDMs have been explored from a wide variety of perspectives; however, the underlying uncertainty in climatic predictors is often overlooked (Baker et al., 2016; Baker et al., 2017; Bedia et al., 2013; Jiménez-Valverde et al., 2021; Stoklosa et al., 2015; Suggit et al., 2017). Recent decades have seen a rapid increase in the availability of high-resolution spatial datasets, often derived from remotely sensed data that characterize environmental conditions at broad scales. While many of these datasets provide new opportunities to directly estimate environmental gradients such as vegetation cover and vertical structure (e.g., Fedrigo et al., 2018; Fedrigo et al., 2019) there have also been ongoing efforts to enhance climate surfaces (e.g., Fick & Hijmans, 2017; Karger et al., 2017; Stewart et al., 2017; Stewart & Nitschke, 2017e) which are critical inputs for a wide range of ecological models.

The development of new methods for representing climatic conditions and weather at fine scales has often been driven by the need to better understand vegetation distributions and ecological systems (Ashcroft & Gollan, 2012; Booth et al., 2014; Hijmans et al., 2005; Karger et al., 2017; Kearney et al., 2014). Advances in computing and spline interpolation techniques (Hutchinson & Bischof, 1983; Wahba & Wendelberger, 1980) were crucial to the development of BIOCLIM in the late 1980s and early 1990s (Booth et al., 2014) and continue to provide effective methods for generating climate surfaces using weather station observations (Fick & Hijmans, 2017; Hijmans et al., 2005; Hutchinson et al., 2009; Jeffrey et al., 2001; Stewart et al., 2017; Stewart & Nitschke, 2017e). The WorldClim datasets (Fick & Hijmans, 2017; Hijmans et al., 2005) represent global climate and bioclimatic indices (see Booth et al., 2014) between 1950 and 2000 at high resolution (approx. 1 km) and have recently been updated to utilize additional covariates and remote sensing variables. WorldClim is one of the most widely used sources of climate data for ecological studies; however, many alternative methods of generating such datasets have been demonstrated at ecologically meaningful scales. These include downsampling and empirical adjustment of pre-existing climate surfaces (Abatzoglou et al., 2018; Karger et al., 2017; Kearney et al., 2014, 2020; Kriticos et al., 2012; Poggio et al., 2018), correlative interpolations calibrated using micro-loggers in forested ecosystems (Ashcroft & Gollan, 2012; Haesen et al., 2021; Lembrechts et al., 2021; Meineri & Hylander, 2017) and techniques that apply geostatistical and spline-based methods in conjunction with earth observation data (Funk et al., 2015; Hengl et al., 2012; Oyler et al., 2015; Stewart et al., 2017; Stewart & Nitschke, 2017e).

Several studies have investigated alternative climate datasets representing current conditions and their impact on SDMs. These analyses have been conducted at a range of spatial resolutions (50 m to 50 km), with various taxonomic groups, study regions, modelling algorithms and datasets, with and without consideration of climate change scenarios. Climate datasets that better characterize regional temperature and precipitation have been shown to improve predictions relative to WorldClim when modelling the distributions of European beech (Fagus sylvatica) in Northern Iberia (Bedia et al., 2013), plant species in Switzerland (Karger et al., 2017) and Sweden (Meineri & Hylander, 2017), and several bird and fern species in the humid montane forests of Bolivia (Soria-Auza et al., 2010). Bobrowski et al. (2021) found that CHELSA was more effective than WorldClim in predicting the distribution of Himalayan Birch (Betula utilis) in Nepal. In their study of 48 woody plant species in Iberia, Jiménez-Valverde et al. (2021) found that WorldClim and the regional Iberian Climate Atlas produced differences in modelled predictions, statistical performance and variable importance. Distribution models for 451 plant species calibrated using six combinations of global climate data across central Africa, western India and the Amazon basin were found to perform best when land surface temperature or precipitation derived from earth observation data were independently used as model covariates (Deblauwe et al., 2016). Baker et al. (2016) compared six alternative global climate datasets in their study of over 800 African bird species and found that the choice of climate data used in the baseline was at least as important, but could be much more important, as the choice of General Circulation Model (GCM) used for evaluating potential climate change. Lembrechts et al. (2019) showed that distribution models for forbs, graminoids and dwarf shrubs in Scandinavia performed significantly better using in-situ soil temperature in comparison to free-air or land-surface temperature. Each of these studies highlight the sensitivity of ecological models to climate dataset selection and demonstrate the need to develop robust approaches to quantifying uncertainty.

Part of the challenge is that there are many potential compound sources of uncertainty (e.g., input data, algorithm selection and parameterization) that can affect both SDMs and climate surfaces, and few SDM studies have directly considered how uncertainty in climate variables influence prediction. Stoklosa et al. (2015) showed how the statistical performance of SDMs can improve when
accounting for the underlying uncertainty in climate variables. While effective, their approach may not be practical in many cases as it requires that uncertainty in climate variables is both well characterized and made available. Furthermore, the ways in which uncertainty in climate surfaces can be represented is dependent upon factors such as algorithm selection, model calibration and observation density. Morales-Barbero and Vega-Álvarez (2019) proposed a method for characterizing uncertainty in the spatial predictions of species distribution models by fitting multiple global climate datasets in conjunction with a statistical measure of their congruence. Their approach highlights differences between the alternative climate datasets; however, it does not consider the relative importance of climate response variables.

A more general approach to constraining prediction error which drives the success of some machine learning algorithms (e.g. random forests; Breiman, 2001) and which has been commonly applied in climatology (Knutti et al., 2010; Krishnamurti et al., 1999) and SDMs (Kindt, 2018; Marmion et al., 2009; Thuiller et al., 2009), is to calculate the ensemble mean of competing models. Ensemble predictions will perform best when individual models have low bias and covariance, but high variance (Dormann et al., 2018); however, the potential for signal loss from specific predictions will in some cases favour the use of single well-tuned model over an ensemble (Crimmins et al., 2013; Hao et al., 2020; Knutti et al., 2010). SDM ensembles, commonly developed using alternative algorithms, can be evaluated in several ways (e.g. mean, variability, consensus above or below a threshold, as implemented in the R package biomod2; Thuiller et al., 2020); however, results from studies using these methods are not always presented with sufficient supporting Information to evaluate the benefits of ensemble techniques over individual models (Hao et al., 2019). Describing ensemble model performance using non-parametric methods such as cross-validation can provide a reliable mechanism for evaluating uncertainty (Dormann et al., 2018), but spatially continuous measures of ensemble behaviour are likely to be particularly useful in the case of climate dataset evaluation, as embedded biases in model covariates may not always be detected in statistical performance metrics (Bedía et al., 2013).

The objectives of this research were to assess the impact of alternative climate datasets on modelled species distributions by comparing the statistical performance and spatial distribution of predictions both individually and as ensembles. It is important to understand how alternative climate baselines influence predictions for two primary reasons. Firstly, there are an increasing number of datasets becoming available and it may not always be clear which dataset is most suitable for a given application. Secondly, ecological models rely heavily on climate variables, and when they are projected into new space such as under climate change scenarios, they adopt the biases present in the selected baseline conditions. Several datasets developed specifically for southeast Australia (Stewart & Nitschke, 2017e) and Bhutan (Stewart et al., 2017: Appendix S1) provide an opportunity to evaluate the influence that alternative climate interpolation techniques have on SDMs and how this compares to global products. These datasets were generated using identical station observations and widely adopted spline-based approaches (full dependence on latitude, longitude and elevation), with additional models having partial dependence on topographic indices to characterize cold air pooling and drainage regimes, and remotely sensed land surface temperature (LST) data that can capture mesoscale temperature gradients. We also included both WorldClim datasets (including the updated version that uses coastal proximity and LST to support the interpolation procedure) and CHELSA, which have been widely used and are well suited to ecological modelling. We first directly compare the statistical performance of models fitted using alternative climate datasets. Our analyses then consider how ensembles compare against individual models. We propose and demonstrate two non-parametric, spatially continuous measures of model congruence and uncertainty which complement and aid the interpretation of ensemble predictions.

2 | METHODS

2.1 | Study area and species data

Plant species distributions were modelled across two study regions; (1) Victoria (140.9°-150.0°E, 34.0°-39.2°S), southeast Australia and (2) Bhutan (88.7°-92.1°E, 26.7°-28.2°N). The state of Victoria has a temperate climate and is located on the southeastern coast of the Australian mainland. The heavily forested alpine regions in the east of Victoria reach elevations of up to 2000 m above sea level (asl). Bhutan is a land-locked country located on the southern slopes of the eastern Himalayas. Strong gradients in elevation mean that climatic conditions in Bhutan are highly variable, with elevation varying by over 7 km (100–7500 m asl) across a distance of only 170 km.

All species data used in this study were presence and absence records obtained from systematic field studies within forested and woodland ecosystems. Species records were collated from multiple field campaigns to maximize the pool of available data points, including 1804 sites in Victoria (see Stewart, Elith, et al., 2021) and 349 sites in Bhutan (see Choden et al., 2021). The most prevalent canopy and understory plant species in each study region (38 in Victoria and 12 in Bhutan) were selected for modelling. We focused on longer-lived, woody species, as climate plays a more important role in their turnover than edaphic properties, which vary over finer scales and play a larger role in the distribution of non-woody species with shallower root systems (Kasel et al., 2017). For Victoria, we also included two prevalent tree fern species (Dicksonia antarctica, Cyathea australis) that are associated with rainforest ecosystems and two prevalent grass-tree species (Xanthorrhoea australis and Xanthorrhoea minor) with distinctive climatic niches. Forbs, grasses and small shrubs not capable of reaching 3 m in height were not considered due to the importance of microclimate and small-scale variability in near-surface temperature for these growth forms (Lembrechts et al., 2019). The sampling domain represented a broad range of climatic and topographic gradients across each study region.
2.2 | Climate variables

Spatial climate variables were generated using maximum and minimum air temperature (1.2–2 m above ground), and precipitation data from each of the datasets included in the analysis (Table 1). We focused on temperature and precipitation as they were the most commonly available variables across data sources. Both monthly mean (MM) and time-series (TS) data were considered where available. Summaries of temporal variability (e.g. standard deviation, the observed frequency of extremes) add another dimension to the climatic niche and can improve the predictive performance of SDMs (Karger et al., 2021; Perez-Navarro et al., 2021; Stewart, Elith, et al., 2021; Zimmermann et al., 2009), therefore the TS data were used to provide an additional contrast against the MM products. Three alternative temperature datasets and one precipitation dataset developed specifically for ecological modelling in Victoria (see Fedrigo et al., 2021; Stewart, Elith, et al., 2021; Stewart & Nitschke, 2017e) were included. These were developed at high spatial (approx. 250 m) and temporal (daily, Jan 1981 to Dec 2019) resolution to capture variability in topographically complex landscapes, permitting the calculation of derived variables summarizing the time-series in ecologically relevant ways (e.g. inter-annual distribution of extremes). An analogous set of climate data products were developed for Bhutan to use in this study (see Appendix S1 for methods, statistical performance and comparison against other contemporary products), building on previous interpolations for the region using the local weather station network (Stewart et al., 2017). All interpolations were performed using ANSPLIN 4.4 (Hutchinson & Xu, 2013). We included the two WorldClim datasets (Fick & Hijmans, 2017; Hijmans et al., 2005) given their widespread use in species distribution modelling and global coverage, noting that they only represent long-term mean conditions. CHELSA data representing the period between 1980 and 2018, developed by statistical downscaling the ERA5 reanalysis (Hersbach et al., 2020), was included as an alternative global product capable of representing variability over time. We initially considered the Australian national SILO (Jeffrey et al., 2001) and AWAP (Jones et al., 2009; Raupach et al., 2012) datasets for Victoria; however, they were both excluded because they are comparatively coarse in spatial grain (0.05°) and therefore unlikely to capture fine-scale temperature gradients that occur across the topographically variable regions where many of our study species are found. No comparable fine-grain analogues were available for Bhutan (but see Appendix S1 for a comparison of the regional products with CHELSA and CRU TS 4.05; Harris et al., 2020). Except for CHELSA, each of the datasets selected use thin plate smoothing splines to interpolate climate records, with various combinations of latitude, longitude, elevation, topographic indices, coastal proximity and thermal remote sensing data as either spline variables or independent linear covariates (Table 1).

Each dataset was used to generate a series of climate predictor variables for species distribution modelling (Figure 1). Monthly precipitation was converted to a quarterly (3 months) rolling sum to characterize cumulative seasonal moisture availability. Annual mean minimum temperature, maximum temperature and quarterly precipitation were calculated for each MM and TS dataset. Maximum temperature of the hottest month (BIO5), minimum temperature of the coldest month (BIO6) and precipitation of the driest quarter (BIO17) were calculated as corresponding measures of seasonality to characterize common climatic stressors that plant communities are exposed to in a typical year. Seasonality variables were calculated annually for the TS datasets. The median was selected as an analogue to the MM seasonality variables, and 1-in-15-year values (i.e. 93.3rd or 6.7th percentile) were also included to represent rarer, but more extreme stressors that may influence plant demography. This recurrence period was selected as it had performed best overall in a previous study of climate extremes in Victoria (Stewart, Elith, et al., 2021). Each of the seasonality and extreme variables were expressed relative to corresponding annual mean and seasonality variables, respectively, to minimize the effects of collinearity. The distribution of climate variables across each presence and absence site and each source dataset are illustrated in Appendix S2.

2.3 | Species distribution modelling and performance

Species distribution models were fitted in R with the package biomod2 (Thuiller et al., 2020) using each of the climate variables described above and seven different algorithms: artificial neural networks (ANN), classification tree analysis (CTA), flexible discriminant analysis (FDA), generalized additive models (GAM), gradient boosting models (GBM), multivariate adaptive regression splines (MARS) and random forests (RF). We applied the default parameters for all models, in line with typical usage of biomod2 (Hao et al., 2019). Model performance was evaluated using the AUC. True Skill Statistic, Kappa, and overall accuracy. For brevity, we have focused on discrimination ability with AUC and provided additional calibration and classification statistics in the supplementary materials (Appendix S3). Each were multiplied by 100 (valid range 0–100) to improve readability of the results. A spatially blocked k-fold (approx. 50 x 50 km blocks, 6-10 folds) cross-validation design was used to support conservative estimates of error (Roberts et al., 2017). Spatial blocks were selected using the blockCV R package (Valavi et al., 2019), and each set of folds were held constant across all models for each species to ensure robust comparisons of performance.

Model ensembles were computed as the mean probability of SDMs fitted using alternative combinations of algorithms and climate datasets. Each selected model was assigned an equal weight in the ensemble to avoid the introduction of unknown biases (Claeskens et al., 2016). Three separate sets of SDM ensembles were first calculated: (1) ensembles averaging over each available algorithm, (2) ensembles averaging over each available MM climate dataset and (3) ensembles averaging over each available algorithm and MM climate dataset. These were compared to evaluate the changes in statistical performance that could be achieved using climate-based ensembles as opposed to multiple algorithms, the latter which has been acknowledged as one of the
The optimal classification threshold \( T \) was first obtained by calculating the mean threshold that minimized the difference between sensitivity (true positive rate) and specificity (true negative rate) using spatially blocked cross-validated predictions from each of the contributing models. This approach to identifying the optimal threshold was selected based on the recommendations of Jiménez-Valverde and Lobo (2007) but could be estimated by other means (e.g. maximize sum of sensitivity and specificity). The Standard Threshold Score (STS) is subsequently calculated for all pixels as...
Monthly mean climate variables (MM; 1981-2010 mean; \( n = 6 \) variables)

- Annual mean maximum temperature (a)
- Maximum temperature seasonality (d - a)
- Annual mean minimum temperature (b)
- Minimum temperature seasonality (e - b)
- Annual mean quarterly precipitation (c)
- Quarterly precipitation seasonality (f - c)

Time-series derived climate variables (TS; 1981-2019 monthly values; \( n = 9 \) variables)

- Annual mean maximum temperature (g)
- Maximum temperature seasonality (j - g)
- Maximum temperature extreme (m - j)
- Annual mean minimum temperature (h)
- Minimum temperature seasonality (k - h)
- Minimum temperature extreme (n - k)
- Annual mean quarterly precipitation (i)
- Quarterly precipitation seasonality (l - i)
- Quarterly precipitation extreme (o - l)

Seasonality and extremes defined by maximum temperature of warmest month (d), minimum temperature of coldest month (e) and minimum precipitation of the driest quarter (f) for each year in the time series

Seasonality: median (j,k,l)
Extreme: 1 in 15 year event (m,n,o)

FIGURE 1  Climate variables derived from monthly mean (MM; \( n = 6 \)) and time-series (TS; \( n = 9 \)) datasets used for species distribution modelling for a point location southeast of Melbourne, Australia (145.101°E, 37.942°S). Letters in parenthesis associated with each response variable correspond to the specific (or subtraction of one from another) value as indicated on graphical representations of the source data.
where $\mu$ and $\sigma$ are the population mean and population standard deviation of predictions from all contributing models at pixel $j$. The STS calculation provides a spatially continuous surface which expresses the ensemble mean prediction in standard deviations from $T$.

The Chebyshev–Cantelli inequality

$$P \left( X - \mu \geq k \right) \leq \frac{\sigma^2}{\sigma^2 + k^2}$$

(2)

gives the probability that random variable $X$ will be greater than $k$ standard deviations from the mean (see Appendix S4 for illustration) and was used to calculate the Threshold Agreement Index ($\text{TAI}$), which describes the agreement of individual models above or below the optimal classification threshold. The TAI is calculated as

$$\text{TAI} = \begin{cases} \frac{1}{1 + \text{STS}^2} - 1 & \text{if STS} < 0 \\ 1 - \frac{1}{1 + \text{STS}^2} & \text{if STS} \geq 0 \end{cases}$$

(3)

such that the sign of the difference from the threshold is preserved and values are bound between $-1$ and $1$. For example, TAI values of $-0.8$ ($\text{STS} = -2$) or $0.8$ ($\text{STS} = 2$) indicate that a minimum of 80% of the individual model probabilities are below or above the selected threshold respectively.

The Threshold-scaled Standard Deviation ($\text{TSD}$) uses the STS to scale variability of contributing model predictions around the optimal threshold. This penalizes prediction variability in regions where there is strong agreement for suitable or unsuitable environmental conditions and emphasizes variability in model predictions closest to the ensemble threshold. The TSD is calculated as

$$\text{TSD} = \frac{\sigma}{1 + \text{STS}^2}$$

(4)

3 | RESULTS

Ensembles of SDMs compiled using all model algorithms and MM climate datasets performed significantly better than ensembles developed with either climate or algorithms alone in both Victoria (climate, $\Delta \text{AUC} = -1.7 \pm 0.4, p < 0.001$; algorithm, $\Delta \text{AUC} = -1.3 \pm 0.2, p < 0.001$, Figure 2a, b) and Bhutan (climate, $\Delta \text{AUC} = -2.0 \pm 0.7, p < 0.001$; algorithm, $\Delta \text{AUC} = -3.6 \pm 1.1, p < 0.001$, Figure 2c, d). When directly compared against one another, there were no significant differences in performance between SDM ensembles compiled using alternative MM climate variables or algorithms.

Highly significant ($p < 0.01$) differences in AUC were only found for 14 comparisons in Victoria and none in Bhutan (Table 2) when evaluating models fitted with distinct climate datasets against one another. Pairwise comparisons of models fitted with MM climate datasets showed fewer differences. VECD performed significantly better than either WC1 or CHELSA in Victoria and was also a significant improvement upon VTCD and VTCD + M for MM models. The mean magnitudes of difference in AUC were typically higher in Bhutan; however, they were not significant due to variability in species-specific performance. Models fitted with TS variables also frequently achieved significantly higher AUC than MM models in Victoria. Each of the regionally calibrated datasets (VECD, VTCD and VTCD + M) performed significantly better than CHELSA for TS models.

Ensembles were formed using different combinations of models (Table 3, rows). The MM, TS or full (ALL) ensembles consistently increased the mean AUC relative to alternative climate variable sets; however, the differences were not always significant for individual comparisons. Comparable improvements were found for TSS, Kappa and OA (Appendix S5). The ensemble of all models improved the AUC/TSS/Kappa/OA by an average of 1.5/2.0/1.8/1.0 in Victoria and 3.9/5.7/5.9/2.8 in Bhutan across all climate variable sets. There were no significant relationships between species prevalence and statistical performance of single datasets or ensembles (Appendix S6). Except for the change in AUC for WC1 in Bhutan, there were no significant relationships between the area of habitat suitability (where 50% of models agree or TAI > 0.5) and statistical performance of single datasets or ensembles (Appendix S7).

The difference in ensemble AUC compared to alternative climate datasets is summarized per species and per dataset in Figure 3 (see Appendix S8 for TSS, Kappa and OA). Ensembles achieved higher AUC/TSS/Kappa/OA than separate climate datasets for over 85/68/68/68% and 86/69/69/69% of cases in Victoria and Bhutan respectively. There were no consistent trends in which of the climate datasets performed best in either study region. Cross-validated classification and calibration performance statistics for all species and climate variable combinations, in addition to ensembles, are tabulated in Appendix S3.

The spatial congruence and uncertainty among individual model predictions for *Eucalyptus obliqua* and *Abies densa*, as represented by the TAI and TSD are illustrated alongside the ensemble mean and standard deviation in Figures 4a–d and 5a–d respectively. Supplementary occurrence records obtained for *E. obliqua* (see Appendix S9) demonstrate strong model performance despite the lack of calibration data in many locations. The ensemble mean was above the optimal classification threshold for 81.3% of these occurrence records across the study domain (optimal threshold = 0.25, mean ensemble probability = 0.45 and mean TAI = 0.51). The TAI indicates increasing model agreement as it approaches −1 (unsuitable) or 1 (suitable), and highlights areas of potential disagreement where the ensemble mean falls near the optimal classification threshold. The TAI acts like a binary classification but retains numerically continuous information on the level of congruence between competing models (Figures 4c, 5c, 6a, c). The TSD provides a measure of uncertainty which is most sensitive to regions close to the optimal classification threshold (Figures 4d, 5d, 6b, d). The standard deviation is penalized as the STS increases, and therefore acts to filter out regions with the highest level of congruence in modelled predictions.
Our results show that climate-based ensembles provide a simple but effective method of incorporating predictive uncertainty into SDMs where there are several candidate representations of the same variables available. Information describing the key differences between the contributing models is retained when coupled with supporting metrics such as the TAI and TSD, supporting interpretation of the ensemble mean by identifying where predictions best agree and where they are most uncertain. Studies by Bedia et al. (2013), Soria-Auza et al. (2010) and Karger et al. (2017) have recommended specific datasets that best represent the physical processes occurring in a region based on the performance of species distribution models. While a better representation of local, fine-scale climatology can be achieved empirically (e.g., Karger et al., 2017; Kearney et al., 2020) or with the support of remote sensing data (Fick & Hijmans, 2017; Stewart & Nitschke, 2017e), the performance characteristics of these datasets does not consistently correspond to better SDMs.

This is not to say that better climate data products are not worth pursuit, but that a direct correspondence with improved SDM performance cannot be assumed. The variability of spatial predictions when using alternative contemporary climate datasets that has been previously reported in SDM studies (Baker et al., 2016; Jiménez-Valverde et al., 2021) is consistent with our finding that none of the candidate datasets consistently performed best. In the context of this study, ensembles provide a mechanism to trade-off the inherent uncertainties of different data sources while characterizing the congruence and variability of contributing model predictions.

The improvement in statistical performance achieved with climate-based SDM ensembles was comparable with what was achieved by using different algorithms. This finding is significant, as algorithm selection is known to be a large source of uncertainty for SDMs (Araújo & New, 2007; Elith et al., 2006; Thuiller et al., 2019). Furthermore, the combination of alternative climate datasets and algorithms was significantly better than either approach alone. This suggests that additive gains can be achieved by viewing

**FIGURE 2** Cross-validated area under the receiver operating characteristic (AUC) for plant species distribution models fitted using alternative algorithms, monthly mean (MM) climate datasets and species distribution model (SDM) ensembles in Victoria (38 species) and Bhutan (12 species). Tile plots (a, c) show the mean AUC for each combination of climates, algorithms, and ensembles. Stacked histograms (b, d) show the difference in AUC for SDM ensembles compiled using only climate or algorithm alternatives, relative to the ensemble developed using all models.

### 4 | DISCUSSION

This is not to say that better climate data products are not worth pursuit, but that a direct correspondence with improved SDM performance cannot be assumed. The variability of spatial predictions when using alternative contemporary climate datasets that has been previously reported in SDM studies (Baker et al., 2016; Jiménez-Valverde et al., 2021) is consistent with our finding that none of the candidate datasets consistently performed best. In the context of this study, ensembles provide a mechanism to trade-off the inherent uncertainties of different data sources while characterizing the congruence and variability of contributing model predictions.

The improvement in statistical performance achieved with climate-based SDM ensembles was comparable with what was achieved by using different algorithms. This finding is significant, as algorithm selection is known to be a large source of uncertainty for SDMs (Araújo & New, 2007; Elith et al., 2006; Thuiller et al., 2019). Furthermore, the combination of alternative climate datasets and algorithms was significantly better than either approach alone. This suggests that additive gains can be achieved by viewing
and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertainty independently. In this study we included high-resolution, locally calibrated interpolations and addressing both sources of uncertain
TABLE 3  Mean pairwise differences in the cross-validated area under the receiver operating characteristic (AUC) for plant distribution models fitted using discrete monthly mean (MM) or time-series (TS) climate datasets (columns) relative to ensemble means (µ; rows) across all species

|        | WC1 (MM) | WC2 (MM) | CHELSA (MM) | VECD (MM) | VTCD (MM) | VTCD + M (MM) | CHELSA (TS) | VECD (TS) | VTCD (TS) | VTCD + M (TS) |
|--------|----------|----------|-------------|-----------|-----------|---------------|-------------|-----------|-----------|---------------|
| Victoria |          |          |             |           |           |               |             |           |           |               |
| µ GL (MM) | **1.6**  | **1.6**  | **1.6**     | **1.6**   | **1.6**   | **1.6**       | **1.6**     | **1.6**   | **1.6**   | **1.6**       |
| µ GL (ALL) | **1.6**  | **1.6**  | **1.6**     | **1.6**   | **1.6**   | **1.6**       | **1.6**     | **1.6**   | **1.6**   | **1.6**       |
| µ VIC (MM) | **1.9**  | **1.9**  | **1.9**     | **1.9**   | **1.9**   | **1.9**       | **1.9**     | **1.9**   | **1.9**   | **1.9**       |
| µ VIC (ALL) | **1.9**  | **1.9**  | **1.9**     | **1.9**   | **1.9**   | **1.9**       | **1.9**     | **1.9**   | **1.9**   | **1.9**       |
| µ VIC (TS) | **2.1**  | **2.1**  | **2.1**     | **2.1**   | **2.1**   | **2.1**       | **2.1**     | **2.1**   | **2.1**   | **2.1**       |
| µ (ALL) |          |          |             |           |           |               |             |           |           |               |
| Bhutan |          |          |             |           |           |               |             |           |           |               |
| µ GL (MM) | **1.6**  | **1.6**  | **1.6**     | **1.6**   | **1.6**   | **1.6**       | **1.6**     | **1.6**   | **1.6**   | **1.6**       |
| µ GL (ALL) | **1.6**  | **1.6**  | **1.6**     | **1.6**   | **1.6**   | **1.6**       | **1.6**     | **1.6**   | **1.6**   | **1.6**       |
| µ BHU (MM) | **1.9**  | **1.9**  | **1.9**     | **1.9**   | **1.9**   | **1.9**       | **1.9**     | **1.9**   | **1.9**   | **1.9**       |
| µ BHU (ALL) | **1.9**  | **1.9**  | **1.9**     | **1.9**   | **1.9**   | **1.9**       | **1.9**     | **1.9**   | **1.9**   | **1.9**       |
| µ (MM) |          |          |             |           |           |               |             |           |           |               |
| µ (TS) |          |          |             |           |           |               |             |           |           |               |
| µ (ALL) |          |          |             |           |           |               |             |           |           |               |

Note: VIC includes VECD, VTCD and VTCD + M; BHU includes BECD, BTCD, BTCD + M; GL = includes WC1, WC2 and CHELSA.

Statistical significance: *p < 0.1; **p < 0.05; ***p < 0.01.
Statistical comparisons performed using two-tailed pairwise t-tests. Positive values indicate higher AUC for ensembles listed by row.
FIGURE 3  Difference in the cross-validated area under the receiver operating characteristic (AUC) for plant species distribution models fitted using alternative climate datasets (n = 10, boxes) relative to the ensemble mean of all models (red line). Results are presented both per dataset (a, b) and per species (c, d)
including the comparatively coarse resolution of LST (approx. 1 km) and its sensitivity to soil moisture, limited number of species assessed in Bhutan, available extent of regional data products, and the exclusion of other important environmental gradients that are also be important for defining the niche but can be difficult to accurately quantify without in-situ measurements (e.g. edaphic properties; Bennett et al., 2020).

The time-series derived variable sets improved performance in many pairwise comparisons in Victoria and were highly effective when compiled as ensembles and compared against separate climate datasets. The improvement is consistent with several recent studies showing how interannual variability in climate can improve SDMs (Karger et al., 2021; Perez-Navarro et al., 2021; Stewart, Elith, et al., 2021; Zimmermann et al., 2009). Importantly, this improvement was also found when using contrasting methodologies to produce temporal variation. The local interpolations use an anomaly plus climatology technique to generate time-series data (Stewart, Elith, et al., 2021; Stewart & Nitschke, 2017e), whereas CHELSA uses statistical downscaling techniques applied to 3-hourly ERA5 reanalysis data. Although the monthly mean surfaces are likely less prone to error accumulation as they are often interpolated directly, the ability to characterize interannual variability can improve predictive performance while allowing for more flexibility in the design and selection of climate predictors. Even though the time-series models often performed better than the monthly mean alternatives, further improvements were achieved using ensembles.

**FIGURE 4** Spatial distribution of the ensemble mean (a), ensemble standard deviation (b), threshold agreement index (TAI; c) and threshold-scaled standard deviation (TSD; d) from distribution models fitted for *Eucalyptus obliqua* using 10 alternative sets of climate variables across Victoria, southeast Australia.
The proposed TAI and TSD metrics characterize congruence and variability across any number of predictions, supporting the ensemble mean by ensuring that information describing the contributing models is retained. The TAI is computed directly from modelled probabilities rather than the input climate surfaces, meaning that it embeds species-specific responses to variables and can be applied to machine learning models that can fit non-linear trends and variable interactions. The TAI could also be applied in a wide range of analyses where potential niche overlap is of interest, such as the evaluation of climate change scenarios. When congruence is low or the ensemble prediction is close to the optimal classification threshold (i.e. TAI ≅ 0), the standard deviation of the predictions is assigned the highest weight for calculating the TSD. This means that the TSD describes regions that have the highest level of predictive uncertainty relative to the optimal classification threshold. The TSD can therefore be indicative of a poorly performing model or strong disagreement among models. This may indicate that additional sampling is required to better characterize the species-response, or that a subset of the models are producing erroneous predictions and should be considered for exclusion. The spatial coherence of each of the proposed metrics, often around the periphery of the predicted distribution, strongly suggest that their use could be valuable for evaluating risk across competing management and climate change scenarios.

The results of this study indicate that there can be significant advantages to considering multiple sources of climate data when predicting species distributions. While we have focused on climate datasets these methods could also be applied in other situations, for example, where there are multiple representations of a response variable (e.g. remotely sensed biophysical attributes sourced from different sensors) or when evaluating ensembles of competing algorithms. We suggest the following steps be taken to build reliable ensemble models with corresponding measures of uncertainty where several candidate data sources are available:

1. Consider the climatic and non-climatic variables that are of ecological significance to the target species.
2. Identify and evaluate datasets that are available for the target study region that quantify these variables. Both regional and global products should be considered where available. Exclude products known to be unsuitable for the study domain.
3. Consider any post-processing methods that should be applied to the available products to improve suitability of the data for the target species. Statistical downscaling against fine-grain digital elevation model may be useful in regions of complex topography where the available products are comparatively coarse (Guisan et al., 2007). Microclimate modelling (Kearney et al., 2020; Maclean et al., 2019) using macroclimatic data sources may better represent environmental conditions for the target species.

4. Calibrate alternative models and generate spatial predictions using multiple representations of the same variables provided by the available range of datasets. Where post-processing has been applied, it may be worth considering various combinations of input data sources (e.g. vegetation cover, precipitation, air temperature and wind speed).

5. Evaluate the spatial and statistical performance of the models. Consider removing models that are unsuitable or perform poorly relative to the alternatives (e.g. below a statistical threshold; Thuiller et al., 2020). Visual inspection for spatial artefacts and expert consultation may assist in identifying models with erroneous predictions.

6. Calculate the TAI and TSD (Equations 3 and 4; code examples provided in supplementary data; Stewart et al., 2021a) to summarize model predictions as an ensemble. This can be done using weighted values (e.g. by statistical performance; Araújo & New, 2007; Marmion et al., 2009), or assuming equal weights (Thuiller et al., 2009) to mitigate the potential for introducing unknown biases (Claeskens et al., 2016). The TAI indicates the level of agreement in niche suitability across models, relative to the ensemble threshold. Conversely, the TSD will indicate areas where the largest material differences (i.e. high variability close to the optimal classification threshold) in niche suitability are predicted.

7. Evaluate variability across model predictions using the TSD. This step will help to identify areas that are indicative of uncertainty due to divergent model predictions and may require additional sampling to improve model calibration. Where
additional samples can be obtained and included in these regions, return to step 4.
8. Calculate the ensemble mean of model predictions, retaining both the TAI and TSD to support the interpretation of the ensemble.

5 | CONCLUSION

The results of this study demonstrate that ensemble modelling is an advantageous strategy when faced with uncertainty in the selection of competing climate datasets for predicting species distributions. The ongoing development of new and better techniques for generating, downscaling, and modifying climate surfaces, and need to evaluate the response of ecosystems to projected climate change, means that these uncertainties are likely to become even greater over time. The procedure described can minimize the risk of unknowingly selecting a single dataset with suboptimal performance and introducing bias into modelled species distributions. The value of ensembles becomes clear when coupling the mean with supporting measures of congruence and uncertainty. The proposed TAI and TSD metrics provide spatially continuous surfaces that characterize the level of agreement between contributing predictions and indicate where uncertainty is greatest. There are many potential sources of error in SDM; however, the inherent uncertainties in the spatial variables used for prediction are not often considered directly. This approach to ensemble modelling provides a simple mechanism to enhance the quality of predicted species distributions by mitigating the risks of erroneous predictions and by providing spatially explicit measures of uncertainty.

6 | DATA SOURCES

Species records for Victoria were collated from the following sources: S. C. Baker et al. (2019); Chick et al. (2018); Fairman et al. (2017); M. Fedrigo et al. (2019); Haslem et al. (2016); Haywood et al. (2018); Kasel et al. (2017); Leonard et al. (2016); Nelson et al. (2017).

ACKNOWLEDGEMENTS

We would like to thank the National Center for Hydrology and Meteorology (NCHM), Ministry of Economic Affairs, Bhutan for providing weather station observations that were essential in developing the climate surfaces used in this study. We thank Jane Eith from the University of Melbourne, Chris Ware and Sam Andrew from the CSIRO, and each of the anonymous reviewers for providing valuable and constructive feedback.

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

PEER REVIEW

The peer review history for this article is available at https://pubons.com/publon/10.1111/ddi.13515.

DATA AVAILABILITY STATEMENT

The datasets developed as part of this study are available on the CSIRO Data Access Portal (https://data.csiro.au). See Stewart et al. (2021a) for the climate variables, calibration records, modelled species distributions, ensembles, metrics and code used to calculate the TAI and TSD. See Stewart et al. (2021b) for the monthly and monthly mean climate datasets interpolated for Bhutan.

ORCID

Stephen B. Stewart https://orcid.org/0000-0002-1835-6211
Melissa Fedrigo https://orcid.org/0000-0002-6883-0647
Stephen H. Roxburgh https://orcid.org/0000-0001-5021-3459
Karma Tenzin https://orcid.org/0000-0001-9594-2815
Craig R. Nitschke https://orcid.org/0000-0003-2514-9744

REFERENCES

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. Sci Data, 5, 170191. https://doi.org/10.1038/sdata.2017.191
Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology & Evolution, 22(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010
Ashcroft, M. B., & Gollan, J. R. (2012). Fine-resolution (25 m) topoclimatic grids of near-surface (5 cm) extreme temperatures and humidities across various habitats in a large (200 x 300 km) and diverse region. International Journal of Climatology, 32(14), 2134–2148. https://doi.org/10.1002/joc.2428
Baker, D. J., Hartley, A. J., Butchart, S. H. M., & Willis, S. G. (2016). Choice of baseline climate data impacts projected species’ responses to climate change. Global Change Biology, 22(7), 2392–2404. https://doi.org/10.1111/gcb.13273
Baker, D. J., Hartley, A. J., Pearce-Higgins, J. W., Jones, R. G., & Willis, S. G. (2017). Neglected issues in using weather and climate information in ecology and biogeography. Diversity and Distributions, 23(3), 329–340. https://doi.org/10.1111/ddi.12527
Baker, S. C., Kasel, S., van Galen, L. G., Jordan, G. J., Nitschke, C. R., & Pryde, E. C. (2019). Identifying regrowth forests with advanced mature forest values. Forest Ecology and Management, 433, 73–84. https://doi.org/10.1016/j.foreco.2018.10.048
Bedía, J., Herrera, S., & Gutiérrez, J. M. (2013). Dangers of using global bioclimatic datasets for ecological niche modeling. Limitations for future climate projections. Global and Planetary Change, 107, 1–12. https://doi.org/10.1016/j.gloplacha.2013.04.005
Bennett, L. T., Hinko-Najera, N., Aponte, C., Nitschke, C. R., Fairman, T. A., Fedrigo, M., & Kasel, S. (2020). Refining benchmarks for soil organic carbon in Australia’s temperate forests. Geoderma, 368, 114246. https://doi.org/10.1016/j.geoderma.2020.114246
Bobrowski, M., Weidinger, J., & Schickhoff, U. (2021). Is new always better? Frontiers in global climate datasets for modeling treeline species in the Himalayas. Atmosphere, 12(5), 543. https://doi.org/10.3390/atmos12050543
Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). bioclim: the first species distribution modelling package, its early applications and relevance to most current MaxEnt studies. Diversity and Distributions, 20(1), 1–9. https://doi.org/10.1111/ddi.12144
Jiménez-Valverde, A., Rodríguez-Rey, M., & Peña-Aguilera, P. (2021). Climate data source matters in species distribution modelling: the case of the Iberian Peninsula. *Biodiversity and Conservation, 30*(1), 67–84. https://doi.org/10.1007/s10531-020-02075-6

Jones, D. A., Wang, W., & Fawcett, R. (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal, 58*, 233–248. https://doi.org/10.22499/2.5804.003

Karger, D. N., Conrad, O., Böhner, J., Krahulec, T., & Murati, A. (2018). Ensemble species distribution modelling with translocation. *Methods in Ecology and Evolution*.

Kasem, S., Bennett, L. T., Aponte, C., Fedrigo, M., & Nitschke, C. R. (2017). Environmental heterogeneity promotes floristic turnover in temperate forests of south-eastern Australia more than dispersal limitation and disturbance. *Landscape Ecology, 32*(8), 1613–1629. https://doi.org/10.1007/s10980-017-0526-7

Kearney, M. R., Gillingham, P. K., Bramer, I., Duffy, J. P., & Maclean, I. M. D. (2020). A method for computing hourly, historical, terrain-corrected microclimate anywhere on earth. *Methods in Ecology and Evolution, 11*(1), 38–43. https://doi.org/10.1111/2041-210X.13330

Kearney, M. R., Shamakh, A., Tingley, R., Karoly, D. J., Hoffmann, A. A., Briggs, P. R., & Porter, W. P. (2014). Microclimate modelling at macro scales: a test of a general microclimate model integrated with gridded continental-scale soil and weather data. *Methods in Ecology and Evolution, 5*(3), 273–286. https://doi.org/10.1111/2041-210X.12148

Kindt, R. (2018). Ensemble species distribution modelling with transformed suitability values. *Environmental Modelling & Software, 100*, 136–145. https://doi.org/10.1016/j.envsoft.2017.11.009

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate, 23*(10), 2739–2758. https://doi.org/10.1175/2009jcli3361.1

Krishnamurti, T. N., Kishawal, C. M., LaRow, T. E., Bachioki, D. R., Zhang, Z., Williford, C. E., & Surendran, S. (1999). Improved weather and seasonal climate forecasts from multimodel superensemble. *Science, 285*(5433), 1548. https://doi.org/10.1126/science.285.5433.1548

Kriticos, D. J., Webber, B. L., Leriche, A., Ota, N., Macadam, I., Bathols, J., & Scott, J. K. (2012). CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution, 3*(1), 53–64. https://doi.org/10.1111/j.2041-210X.2011.00134.x

Lembrechts, J. J., Lenoir, J., Roth, N., Hattab, T., Milbau, A., Haider, S., Pellasová, L., Pauchard, A., Ratier, Backes, A., Dimarco, R. D., Núñez, M. A., Aalto, J., Nijs, I. (2019). Comparing temperature data sources for use in species distribution models: From in-situ logging to remote sensing. *Global Ecology and Biogeography, 28*(11), 1578–1596. https://doi.org/10.1111/geb.12974

Lembrechts, J. J., van den Hoogen, J., Aalto, J., Ashcroft, M. B., De Frenne, P., Kemppinen, J., & Lenoir, J. (2021). Global maps of soil temperature. *Global Change Biology, 27*(1), 1628–1643. https://doi.org/10.1111/gcb.15606

Leonard, S. W. J., Bruce, M. J., Christie, F., Di Stefano, J., Haslem, A., & Holland, G. J., Kelly, L., Loy, R., MacHunter, J., Rumpff, L., Stamatien, K., Bennett, A., Clarke, M., & York, A. (2016). *Foothills Fire and Biota, Fire and Adaptive Management Report no. 96. Retrieved from Department of Environment, Land Water and Planning, Melbourne.*

Maclean, I. M., Mosedale, J. R., & Bennie, J. J. (2019). Microclima: An R package for modelling meso- and microclimate. *Methods in Ecology and Evolution, 10*(2), 280–290. https://doi.org/10.1111/2041-210X.13093

Marmion, M., Parvainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions, 15*(1), 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Meinert, E., & Hylander, K. (2017). Fine-grain, large-domain climate models based on climate station and comprehensive topographic information improve microrefugia detection. *Ecography, 40*(8), 1003–1013. https://doi.org/10.1111/ecog.02494

Morales-Barbero, J., & Vega-Álvarez, J. (2019). Input matters matter: Bioclimatic consistency to map more reliable species distribution models. *Methods in Ecology and Evolution, 10*(2), 212–224. https://doi.org/10.1111/2041-210X.13124

Nelson, J. L., Durkin, L. K., Cripps, J. K., Scroggie, M. P., Bryant, D. B., Macak, P. V., & Lumsden, L. F. (2017). Targeted surveys to improve Leadbeater’s possum conservation. Arthur Rylah Institute for Environmental Research Technical Report Series No. 278. Retrieved from Department of Environment, Land Water and Planning, Melbourne.

Oyler, J. W., Ballantyne, A., Jencko, K., Sweet, M., & Running, S. W. (2015). Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *International Journal of Climatology, 35*(9), 2258–2279. https://doi.org/10.1002/joc.4127

Perez-Navarro, M. A., Bremmenn, O., Esteve, M. A., Moya-Perez, J. M., Carreño, M. F., Guisan, A., & Lloret, F. (2021). Temporal variability is key to modelling the climatic niche. *Diversity and Distributions, 27*(3), 473–484. https://doi.org/10.1111/dad.13207

Poggio, L., Simonetti, E., & Gimona, A. (2018). Enhancing the WorldClim data set for national and regional applications. *Science of the Total Environment, 625*, 1628–1643. https://doi.org/10.1016/j.scitotenv.2017.12.258

Raupach, M. R., Briggs, P. R., Haverd, V., King, E. A., Pegat, M., & Trudinger, C. M. (2012). *Australian Water Availability Project (AWAP): CSIRO Marine and Atmospheric Research Component: Final Report for Phase 3. CAWCW Technical Report No. 013*. Retrieved from Canberra, Australia.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Soria-Auza, R. W., Kessler, M., Bach, K., Barajas-Barbosa, P. M., Lehnert, M., Herzog, S. K., & Böhner, J. (2010). Impact of the quality of climate models for modelling species occurrences in countries with poor climatic documentation: a case study from Bolivia. *Ecological Modelling, 221*(8), 1221–1229. https://doi.org/10.1016/j.ecolmodel.2010.01.004

Stewart, S. B., Choden, K., Fedrigo, M., Roxburgh, S. H., Keenan, R. J., & Nitschke, C. R. (2017). The role of topography and the northern Indian monsoon on mean monthly climate interpolation within the Himalayan Kingdom of Bhutan. *International Journal of Climatology, 37*(S1), 897–909. https://doi.org/10.1002/joc.5045

Stewart, S. B., Elith, J., Fedrigo, M., Kasel, S., Roxburgh, S. H., Bennett, L. T., Chick, M., Fairman, T., Leonard, S., Kohout, M., Cripps, J. K., Durkin, L., Nitschke, C. R. (2021). Climate extreme variables generated using monthly time-series data improve predicted distributions of plant species. *Ecography, 44*(4), 626–639. https://doi.org/10.1111/ecog.05253
Stewart, S. B., Fedrigo, M., Kasel, S., Roxburgh, S. H., Choden, K., Tenzin, K., Allen, K., Nitschke, C. R. (2021a). Climate-based ensemble predictions and uncertainty metrics (TAI, TSD) for plant species in Victoria, southeast Australia (38) and the Himalayan Kingdom of Bhutan (12) [Raster]. Retrieved from: https://doi.org/10.25919/2e5h-2J38

Stewart, S. B., Fedrigo, M., Kasel, S., Roxburgh, S. H., Choden, K., Tenzin, K., Allen, K., Nitschke, C. R. (2021b). Interpolated climate variables for the Himalayan Kingdom of Bhutan [Raster]. Retrieved from: https://doi.org/10.25919/m8yh-gt42

Stewart, S. B., Fedrigo, M., Roxburgh, S. H., Kasel, S., & Nitschke, C. R. (2020). Climate Victoria: Precipitation (9 second, approx. 250 m) [Raster]. Retrieved from: https://doi.org/10.25919/5e5701b5df4d8

Stewart, S. B., & Nitschke, C. R. (2017a). Climate Victoria: Maximum Temperature (3DS-M; 9 second, approx. 250 m) [Raster]. Retrieved from: https://doi.org/10.25919/5e5701b5df4d8

Stewart, S. B., & Nitschke, C. R. (2017b). Climate Victoria: Maximum Temperature (3DS; 9 second, approx. 250 m) [Raster]. Retrieved from: https://doi.org/10.25919/5e5d9c823b16

Stewart, S. B., & Nitschke, C. R. (2017c). Climate Victoria: Minimum Temperature (2DS-E; 9 second, approx. 250 m) [Raster]. Retrieved from: https://doi.org/10.25919/5e5d4a54398a

Stewart, S. B., & Nitschke, C. R. (2017d). Climate Victoria: Minimum Temperature (3DS-T; 9 second, approx. 250 m) [Raster]. Retrieved from: https://doi.org/10.25919/5e5dafa50749

Stewart, S. B., & Nitschke, C. R. (2017e). Improving temperature interpolation using MODIS LST and local topography: a comparison of methods in south east Australia. International Journal of Climatology, 37(7), 3098–3110. https://doi.org/10.1002/joc.4902

Stoklosa, J., Daly, C., Foster, S. D., Ashcroft, M. B., & Warton, D. I. (2015). A climate of uncertainty: accounting for error in climate variables for species distribution models. Methods in Ecology and Evolution, 6(4), 412–423. https://doi.org/10.1111/2041-210X.12217

Suggitt, A. J., Platt, P. J., Barata, I. M., Bennie, J. J., Burgess, M. D., Bystråkova, N., Duffield, S., Ewing, S. R., Gillingham, P. K., Harper, A. B., Hartley, A. J., Hemming, D. L., Maclean, I. M. D., Maltby, K., Marshall, H. H., Morecroft, M. D., Pearce-Higgins, J. W., Pearce-Kelly, P., Phillimore, A. B., ... Hill, J. K. (2017). Conducting robust ecological analyses with climate data. Oikos, 126(11), 1533–1541. https://doi.org/10.1111/oik.04203

Thuiller, W., Georges, D., Engler, R., & Breiner, F. (2020). bioclim: Ensemble Platform for Species Distribution Modeling. R package version 3.4.6. (Version R package version 3.4.6). Retrieved from https://CRAN.R-project.org/package=bioclim

Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N., & Zimmermann, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. Nature Communications, 10(1), 1446. https://doi.org/10.1038/s41467-019-09519-w

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. Ecography, 32(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution, 10(2), 225–232. https://doi.org/10.1111/2041-210X.13107

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., & Elith, J. (2021). Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. Ecological Monographs, 91(1), e01486. https://doi.org/10.1002/ecm.1486

Wahba, G., & Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross validation. Monthly Weather Review, 108(8), 1122–1143. https://doi.org/10.1175/1520-0493(1980)108<1122:Smnfve>2.0.co;2

Zimmermann, N. E., Yoccoz, N. G., Edwards, T. C., Meier, E. S., Thuiller, W., Guisan, A., Schmatz, D. R., Pearman, P. B. (2009). Climatic extremes improve predictions of spatial patterns of tree species. Proceedings of the National Academy of Sciences, 106(Supplement 2), 19723–19728. https://doi.org/10.1073/pnas.0901643106

BIOSKETCH

Stephen B. Stewart is an experimental scientist at the Commonwealth Scientific Industrial Research Organisation. He applies his expertise in spatial modelling, digital technologies and data analysis techniques to a range of applications, including climate science, ecology, applied remote sensing and natural capital accounting. Prior to working at CSIRO, Stephen worked in the sustainability industry, including the development of carbon offsetting projects under the UNFCCC’s Clean Development Mechanism (CDM).

Author contributions: SBS, and CRN conceived the ideas and designed the methodology. SBS analysed the data and led writing of the manuscript. SBS developed the climate surfaces for Bhutan using data provided by KC, KT and KA. All authors contributed critically to drafts.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Stewart, S. B., Fedrigo, M., Kasel, S., Roxburgh, S. H., Choden, K., Tenzin, K., Allen, K., & Nitschke, C. R. (2022). Predicting plant species distributions using climate-based model ensembles with corresponding measures of congruence and uncertainty. Diversity and Distributions, 28, 1105–1122. https://doi.org/10.1111/dad.13515