GUASNR

Climate change would enlarge suitable planting areas of sugarcanes in China

C.R. Mi[a], Q. Zu[a], L. He[b], F. Huettmann[c], N. Jin[d], J. Li[a,*]

[a]Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, University of Chinese Academy of Sciences, Beijing 100101, China.
[b]National Meteorological Center, Beijing 100081, China.
[c]EWHALE Lab, Department of Biology and Wildlife, Institute of Arctic Biology, University of Alaska Fairbanks (UAF), 419 Irving I, P.O. Box 757000, Fairbanks AK 99775, USA.
[d]State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi 712100, China.
*Corresponding author. E-mail: lijun@igsnrr.ac.cn

Abstract

China's sugar production and consumption continues to increase. This process is already ongoing for over 15 years and over 90% of the sugar production comes from sugarcane (*Saccharum officinarum*). Most of the sugarcane is planted in the south (e.g. the Chinese provinces of Yunnan, Guangxi, Guangdong and Hainan) and it represents there a major economic crop in these landscapes. As found virtually worldwide, climate change is generally expected to influence such suitable planting areas. Here we started a first empirical assessment how climate change would influence the spatial distribution of those current and future suitable planting areas of this strategic crop in China. We employed an ensemble machine learning algorithm (Random Forest; bagging) and increasingly used and robust species distribution models (SDMs). These are based on our compiled and best publicly available crop data sampled from the Chinese sugarcane industry map. They were linked with bioclimate variables from the Worldclim database. This powerful concept allowed us to project sugarcane's current and future (2070) suitable distributions based on the climate niche. Our results were extrapolated to three Global Circulation Models (GCMs; BCC-CSM1-1, CNRM-CM5 and MIROC-ESM) under three representative concentration pathways (RCPs of 2.6, 4.5 and 8.5). The evaluations of these models indicated that our results had a powerful performance (AUC=0.97, TSS=0.96) for robust inference. Bioclimatic variables related to temperature were the most important predictors for sugarcane planting. All models showed similar increasing spatial trends in suitable distribution area and just a few original suitable areas would be lost. Our finding puts emphasize on new growing areas, their soil and management. It is the first to provide the necessary background in the future to safely cultivate sugarcane in climate-suitable areas and to obtain more sugar production for farmers and the industry; it is of large and strategic importance for food security and national autonomy of this central commodity.

Keywords: Sugarcane; Climate change; China; Species distribution model (SDMs); Random forest (bagging) and machine learning; Food security.

Introduction

Sugar represents these days a basic commodity in human society. During the past fifteen years, sugar production and consumption in China are on the rise, growing app. 2.8% - 6.4% per year. It reaches 11.87 million tons and 14.50 million tons in 2011

respectably and the Chinese import currently stands at 2.63 million tons (Li and Liu, 2013). In 2011 over 90% of the sugar production and 85% of the sugar crops planting areas are derived from sugar cane (Fu, 2013; Li and Liu, 2013). Climate is among the most dominant factor that affects species distribution across broad spatial scales (Woodward, 1987; Pearson and Dawson, 2003). The Intergovernmental Panel on Climate Change (IPCC; http://www.ipcc.ch/) (2012) predicts in its scenarios global temperatures to increase by about 1 to 3 °C by the mid 21st century and by the late 21st century an increase of 2 to 5 °C is predicted. How such a climate would impact sugarcane's suitable distribution and its change in China remains currently unknown, so far. Equally elusive remains it influence on national agricultural policy, food security and the world economy.

Sugarcane makes for a major economic crop in southern China and it represents a staple industry in some areas. It's also the main raw material of sugar production and the dominating plant in the Guangxi, Yunnan, Guangdong and Hainan Provinces. In recent years, with the new adjustment of the country's industrial structure, the central-south of Guangxi, southwest of Yunnan, west of Guangdong and north of Hainan will be specifically developed into three major sugarcane-planting areas (Fu, 2013).

Several researchers have already made in-depth studies elsewhere regarding the effects of climate change on sugarcane growth and water requirements from irrigation. For instance, Knox et al. (2010) researched this issue in Swaziland and these authors found it will fail to meet the irrigation demand in the future with unfolding climate change. In South Brazil, Marin et al. (2013) found that sugarcane yield will increase 15 to 59% for 2050. Long et al. (1994) found that China will benefit from climate change, but needs to pay attention to precipitation change in the future. More locally, Chen (2003) showed that sugarcan-production will be negatively affected by climate change in the Fujian Province. Xie et al. (2006) noted the inprovement of Autumn temperature, decrease of Autumn precipitation and increase of Winter precipitaion in the current climate condition would act negatively for sugar accumulation in Xuwen, Guangdong Province. To regionalize climate models for sugarcane growth investigations in China, reseachers usually applied a so-called 'small grid' climate analysis method. They used this approach and combined it with physiology knowledge of sugarcane in a Geographic Information System (GIS) to divide land into different suitable classes (see for instance Xie and Yan (2004) for Fujian Province; Tu et al. (2003) and Su et al. (2006) for Guangxi Province; Liu et al. (2009) for Hainan Province).

As a quite novel method in precision farming, species distribution models (SDMs) include process-based and bioclimatic envelope approaches and they are able to successfully quantify the relationship between species distribution and climate (Guisan and Zimmermann, 2000; Elith et al., 2006; Drew et al., 2011; Mi et al., 2016). SDMs have become an essential tool in ecology, biogeography, evolution and, more recently, in conservation biology (Drew et al., 2011). Increasing attention has been given to project potential species distributions under various climate change scenarios, all based on those methods (Dyer, 1995; Iverson and Prasad 1998; Hijmans and Graham, 2006; Prasad et al., 2006; Wu et al., 2012). While SDMs perform pretty high for wildlife questions (e.g. Kandel et al., 2015), they tend to perform even better for stationary processes like plants (Elith et al., 2006; Ohse et al., 2009). In this study, we chose Random Forest (Breiman, 2001) as our species distribution model machine learning algorithm because of its good performance and common usage (Zhai and Li, 2003; Elith et al., 2006; Drew et al., 2011; Mi et al., 2014). Random Forest is a relatively new entry

to the field of data mining, it employs bagging and it is designed to produce accurate predictions that usually do not overfit the data (Breiman, 2001; Prasad et al., 2006). This approach is not hinging on parametric assumptions. This is achieved by the use of 'partition trees' and with the model aggregation (ensemble) concept, for both classification and regression problems, as officially introduced by Breiman (2001). In machine learning, it belongs to the wider family of ensemble methods, a rather powerful group of algorithms appearing at the end of nineties (Dietterich, 2000). The general principle of Random Forest is to combine many binary decision trees built using several bootstrap samples of rows coming from the learning sample and then choosing randomly at each node a subset of explanatory variables (predictors), hence the name 'random' forests (Prasad et al., 2006; Genuer et al., 2010). A method called 'voting' creates the final tree. Out-of-bag (OOB) samples are used to calculate an unbiased error rate and variable importance, eliminating the need for a test set or cross-validation. These steps are usually referred to as 'bagging'. Details vary on the actual software implication; here we are using the Leo Breiman one with Salford Systems Ltd (SPM; https://www.salford-systems.com/).

Though research has been done on the impacts of climate change for sugarcane growth and yield locally and abroad, in China such study usually is restricted to a provincial scale or even a county scale. Some research has reported on incorporating climate change into crop distribution (Beck, 2013; Estes et al., 2013; Evangelista et al., 2013). However, no research works with the appropriate macro-ecology approach and reported on the effects of climate change on sugarcane distribution, especially applying SDMs explicit in space ad time. Using those methods, the purpose of this investigation here was to explore whether and how climate change would influence the spatial distribution of the current climate-suitable planting areas of sugarcane in China. This topic was identified as being of large importance for food security beyond just China but with a global impact! It's a worthwhile scheme nationally because it certainly matters for governors and farmers in providing them with the necessary background to cultivate sugarcane plants and breed species in climate-suitable areas to obtain a valid sugar harvest and a sustainable income.

Materials and Methods

Study area and data source

Our first-step study area encompassed the whole country of China (Figure 1). Crop data for sugarcanes were derived from the layout of the Chinese Sugarcane Industry (Zhang, 2013), which mapped the three major sugarcane-planting areas (central-south of Guangxi, southwest of Yunnan, west of Guangdong and north of Hainan in related four provinces (Fu, 2013)). Then we selected the sugarcane-planting counties from maps and geo-referenced two samples (planting sites) in each county (farmland region) with ArcGIS10.1. Though random sampling could be used to determine the planting sites in a representative fashion and satisfies our aim, we did not use this method because of our geographic projection scale, which was the whole of China. We determined the GPS sites from Google Earth by locating the points in the farmland. Overall, we obtained 176 geo-referenced locations in our study area (see Figure 1 and data made publicly available as Open Access data in Appendix 1) which is one the best sugarcane databases

for farming in China to now. One thing one needs to notice here is that our samples just represent the location of major sugarcane-planting areas but not the whole region of sugarcane plants. So our model estimates are carefully conservative underestimates.
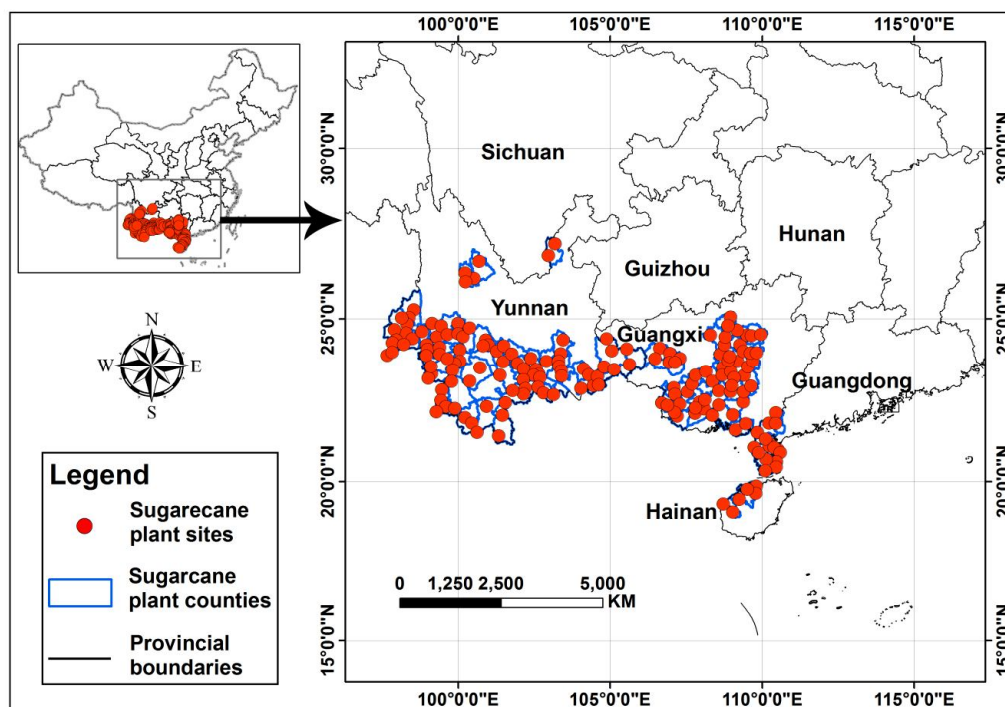

Figure 1. The Study area for predicting the climate-suitable planting area and zoom in for sugarcanes across China.

Environmental variables

Current climate conditions (1950–2000) were represented by 19 bioclimatic variables at a 30 seconds resolution acquired from the public open access Worldclim database (Hijmans et al., 2005; http://www.worldclim.org/). We chose these 19 variables to develop our model as stated in Evangelista et al. (2013). To obtain a valid future estimate we used three future climate scenarios for 2070 (average for 2061–2080). Future climate scenarios were also obtained from the Worldclim database. The data applied here are the most recent IPCC-CMIP5 climate projections from three GCMs (BCC-CSM1-1, CNRM-CM5 and MIROC-ESM, hereafter BC, CN and MR) under three representative concentration pathways (RCPs of 2.6, 4.5 and 8.5, which are named after a possible range of radiative forcing values in the year 2100, relative to pre-industrial values (+2.6, +4.5 and +8.5 W/m^2, respectively).

Model development and evaluation

We chose Random Forest (Breiman, 2001) by Salford Systems Ltd as our species distribution model (SDM) algorithm and model software platform. Random Forest in SPM runs with a convenient GUI and even more importantly, it includes a number of sophisticated optimizations, descriptive results and graphics which are not readily

available in the R version (Herrick, 2013 for a comparison). This algorithm is among the best modeling algorithms available (Elith et al., 2006) and performs so well due to their inherent optimizations in Salford Predictive Modeler (SPM) (Herrick et al., 2013; Kandel et al., 2015). For more details on Random Forest, we refer readers to read the user guide (https://www.salford-systems.com/ products/spm/userguide) and references within (see also Drew et al., 2011).

For model construction, about 10,000 representative pseudo-absence points were taken by random sampling across all of China using the freely available Geospatial Modeling Environment software (GME; Hawth's Tools; http://www.spatialecology.com/ htools/tooldesc.php).

We extracted the climate information from environmental layers for 'presence' (176 plant locations) and the pseudo-absence points. Then we constructed a sugarcane model in SPM with 80% of the dataset (140 presence points and 8,000 pseudo-absence points with climate information). The left-out 20% (36 presence points and 20 pseudo-absence points with climate information) were used for testing model performance. In addition, we used balanced class weights; 1,000 trees were built for all models to find an optimum within, other settings used the default which is known to perform very strong.

Area under the receiver operating characteristic curves (AUC) (Swets, 1988; Fielding and Bell, 1997) and the True Skill Statistic (TSS) (Allouche et al., 2006) were used as metrics to evaluate sugarcane model performance applying testing data. AUC and TSS are commonly used to evaluate SDMs performance (Manel et al., 2001; McPherson et al., 2004; Mi et al., 2016). AUC takes values between 0 and 1, with 0.5 meaning no relevant agreement (essentially random), 0 an inverse relationship (errors better predicted) and 1 perfect agreement (Randin et al., 2006). TSS ranges from -1 to +1, where 1 indicates perfect agreement and values of <0 indicate models that perform worse than random. AUC and TSS in our study were calculated with the 'SDMTools' package in R 3.1.0.

Current and future projection

We created 10 'lattice' data cubes (equally spaced points across China; approximately 5×5 km spacing for the study area). One is for the current climate condition in China and 9 others are for future climate conditions of three GCMs in three scenarios we want to predict to. For each of these latticeswe extracted climate information from the same environmental layers (19 bioclimate variables) as described above for each point. We then used the model to predict ('score') sugarcane presence for each of the regular lattice points. For visualization, we imported the dataset of spatially referenced predictions ('score file') into GIS as a raster file and interpolated for visual purposes between the regular points using inverse distance weighting (IDW) to obtain a smoothed predictive map of all pixels in a high resolution for the sugarcane distributions (as performed in Ohse et al., 2009, Booms et al., 2010).

Next, we used the average predicted index of occurrence across the three GCMs (BCC-CSM1-1, CNRM-CM5 and MIROC-ESM) for each grid as our consensus forecast (named BCM, taking the first letter of each of the three GCMs). This step was considered as one of the best methods for developing an ensemble forecast (Hole et al., 2009). Subsequently, we chose the threshold when the absolute value of 'sensitivity-specificity' get the minimum value (which usually is zero) as the suitable distribution threshold to define the presence–absence distribution of sugar cane's suitable cultivating areas. This method has been found to be a robust approach (Liu et al., 2005, Mi et al., 2016).

Spatial analysis of potential effects of climate change

We applied ArcGIS 10.1 to calculate the suitable areas of sugar cane for two time periods (current, as well as 2070) under three scenarios (RCPs of 2.6, 4.5 and 8.5) from three GCMs (BC, CN and MR) and their average (BCM). We also used the overlay analysis (in ArcGIS 10.1) to assess the potential distribution changes of this crop. This allowed us to identify areas of the planting range that is projected to be lost, gained or remaining under future climate scenarios (Mi et al., 2016).

Results

Model performance and current distribution of sugarcane

Our model obtained a rather high performance measured by AUC and TSS (0.97 and 0.96, respectively), which means our model was classified as very accurate (based on Swets, 1988; Allouche et al., 2006).

Figure 2 shows the current climate suitable distribution map of sugarcane across China, as well as the zoom-in figures for major sugarcane-planting areas for a better overview. We found the most suitable planting ranges were the Yunnan, Guangxi, Guangdong and Hainan Provinces (Figure 2a). Besides, we also found that some places in Tibet, Sichuan, Guizhou, Fujian and Taiwan Province were also suitable planting regions for sugarcane, but not as suitable as the former four provinces. Figure 2(b) indicates that most parts of highly suitable places predicted in Yunnan have already been covered by sugarcanes; Guangxi Province still offers numerous suitable areas in the northwest and southeast, as well as the center-west of Guangdong which could grew sugarcanes (Figure 2c, 2d); From Figure 2e, we observed that almost half of the lands in Hainan Province were suitable for sugarcane cultivation, especially in the northwest direction.

Variable importance

Table 1 presents the top five predictor variables ranking and importance values obtained from the Random Forest metric (the detailed variable rankings and importance values for all 19 predictors can be seen in Appendix 2). We found the top five predictor variables were all temperature-related. The most important predictor to sugarcane distribution was BIO11 (Mean Temperature of Coldest Quarter), followed by BIO7 (Temperature Annual Range) and then was BIO6 (Min Temperature of Coldest Month), the last two were BIO9 (Mean Temperature of Driest Quarter) and BIO4 (Temperature Seasonality).
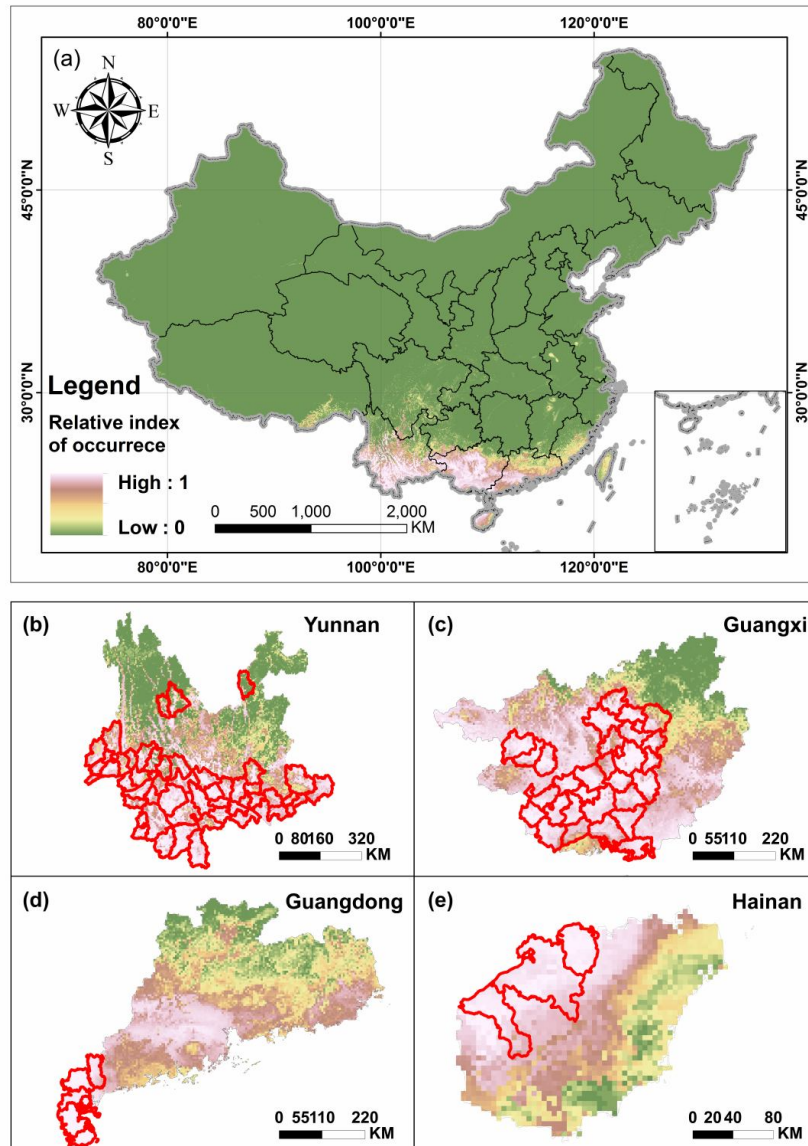
Figure 2. The current climate-suitable planting area of sugarcane predictions obtained from Random Forest. Zones within red line represent the current major planting counties of sugarcanes. (a) current sugarcanes climate-suitable planting area across China, (b) current sugarcanes climate-suitable planting area in Yunnan Province, (c) current sugarcanes climate-suitable planting area in Guangxi Province, (d) current sugarcanes climate-suitable planting area in Guangdong Province, (e) current sugarcanes climate-suitable planting area in Hainan Province.

Table 1. The top five variables ranking and importance value of Random Forest.

Variable	Score
BIO11 (Mean Temperature of Coldest Quarter)	100.00
BIO7 (Temperature Annual Range (BIO5-BIO6))	81.19
BIO6 (Min Temperature of Coldest Month)	60.58
BIO9 (Mean Temperature of Driest Quarter)	60.03
BIO4 (Temperature Seasonality (standard deviation *100))	47.66

Partial dependence plots of Random Forest

   Partial dependence plots provide a tool for interpreting the functional effects of each variable in the model by representing a variable's marginal effects on the response, after accounting for the average effects of other variables in the model (Friedman, 2001; Johnstone et al., 2010). We created these plots in R 3.1.0 (Figure 3) for Random Forest describing the generic relationship between sugarcane distribution and each variable. Figure 3a indicates the preference of sugarcanes for BIO11 (Mean Temperature of Coldest Quarter) >85 (10*ºC). BIO7 (Temperature Annual Range (BIO5-BIO6)) <270 (10* ºC) also appears to be important for the distribution of this crop (Figure 3b). According to Figure 3c, sugarcane favors BIO6 (Min Temperature of Coldest Month) larger than 0 (10*ºC). For the fourth important variable, BIO9 (Mean Temperature of Driest Quarter, Figure 6d), sugarcane prefers to be > 90 (10*ºC). Sugar cane also favors BIO4 (Temperature Seasonality (standard deviation *100)) <7000.

(a)

BIO11(Mean Temperature of Coldest Quarter)

(b)

BIO7 (Temperature Annual Range (BIO5-BIO6))
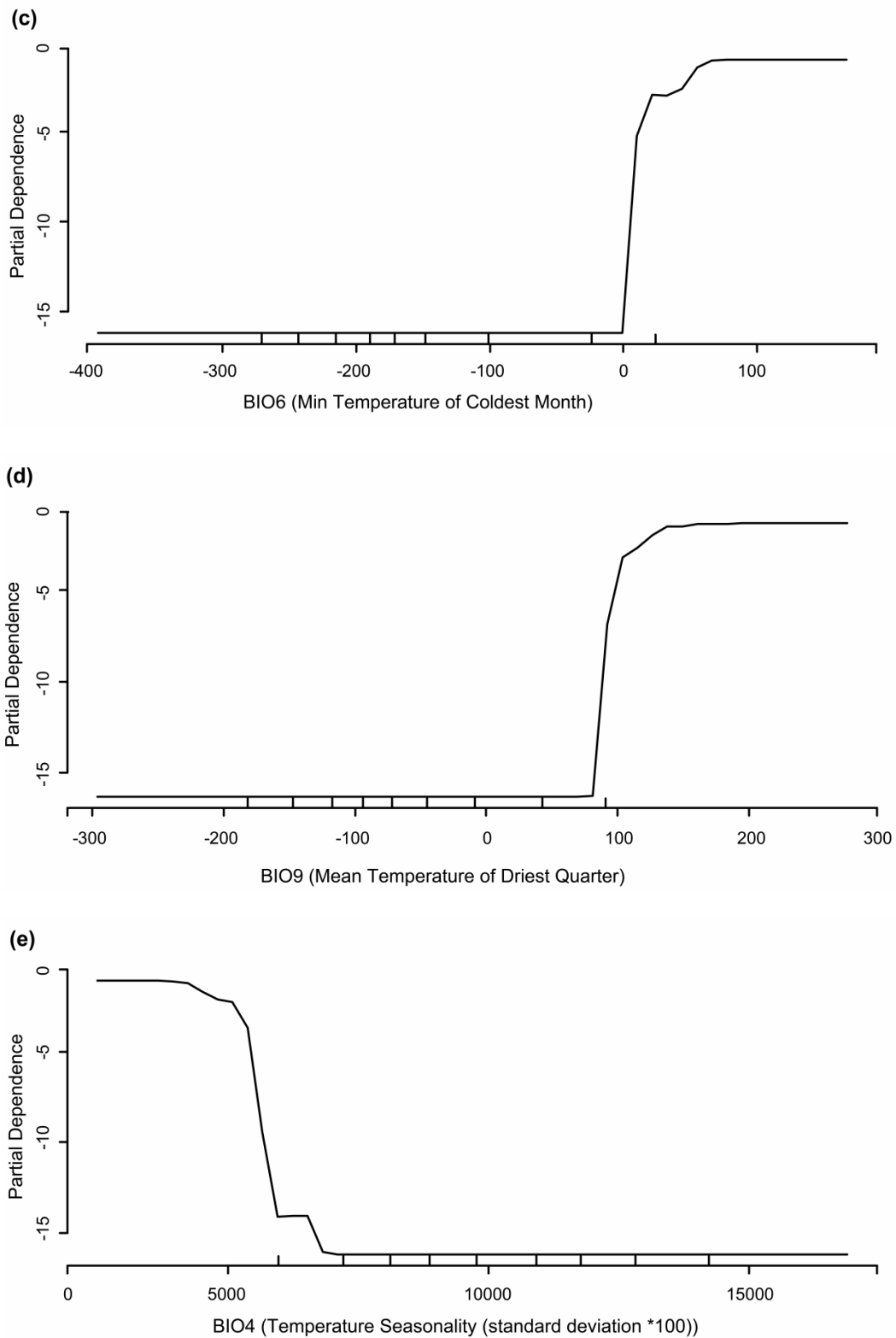
(c)

(d)

(e)
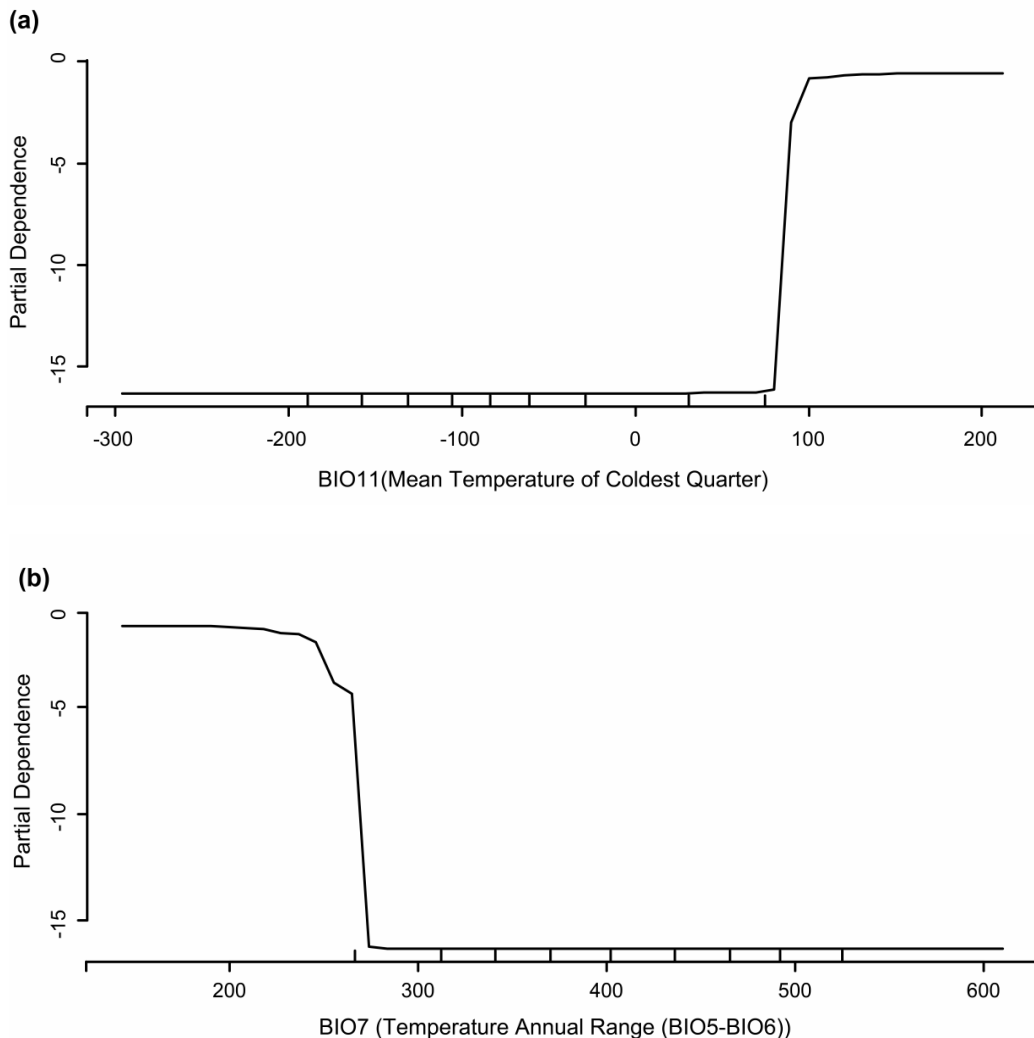
Figure 3. Partial dependence plots of the top five important variables obtained by Random Forest (created in R 3.1.0). The following panels show the relationships (a) with BIO11 (Mean Temperature of Coldest Quarter), (b) BIO7 (Temperature Annual Range (BIO5-BIO6)), (c) BIO6 (Min Temperature of Coldest Month), (d) BIO9 (Mean Temperature of Driest Quarter), (e) influence of BIO4 (Temperature Seasonality (standard deviation *100)).

Spatial analysis of potential effects of climate change

The threshold of 0.75 to define the presence-absence distribution was obtained by the sensitivity-specificity equality approach by (Liu et al., 2005). We then transformed four continuous distribution maps (Current, RCP 2.6, RPC 4.5, RCP 8.5) to binary presence-absence maps (see Appendix 3 for all details). The results indicated that when solely judged by climate change envelopes, the suitable distribution of sugarcanes would enlarge (Figure 4 and Table 2). Depending on RCPs of 2.6, 4.5 and 8.5 scenarios, the climate-suitable planting areas was projected to increase between 44.2% and 104.6% to 2070 (Table 2).
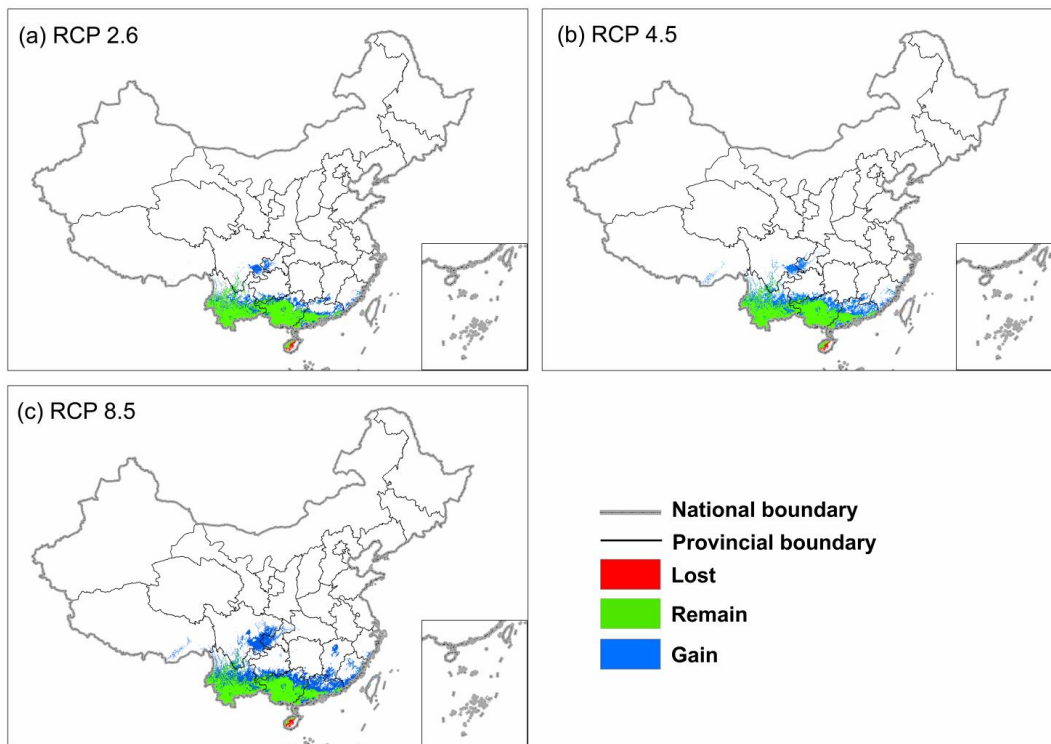

Figure 4. Projected change of sugarcanes climate-suitable plating areas based on a consensus forecast (BCM) from three GCMs by 2070 under (a) RCP 2.6, (b) RCP 4.5, (c) RCP 8.5. The projected current sugarcanes climate-suitable planting areas were overlaid with future projections to identify areas that would be lost, gained, or remain.

Furthermore, Table 2 and Figure 3 show that some of the original suitable planting areas would be lost. This makes for app. 2.5% to 3.7% of the current distribution areas, depending on RCP scenario (Table 2). It mainly would happen in the Hainan Province (see Figure 4). Meanwhile, most parts of the currently suitable areas would still remain. Our model also offered the information that the junction of Sichuan and Chongqing would gradually become a suitable place and the north boundary of suitable areas would shift to further north (Figure 4).

Table 2. Projected change in the total area of sugarcanes suitable distribution areas based on consensus forecast from three GCMs by 2070. Areas are given in (km^2), with the percent of the current total given in parentheses.

Area Scenario	Area Lost (%)	Area Remain (%)	Area Gain (%)	New total habitat (%)
Current	-	-	-	345,580
RCP 2.6	9,910 (2.9)	335,670 (97.1)	172,860 (47.1)	498,620 (144.2)
RCP 4.5	8,510 (2.5)	337,070 (97.5)	247,580 (69.1)	576,150 (166.7)
RCP 8.5	12,700 (3.7)	332,890 (96.3)	387,090 (108.3)	707,280 (204.6)

Discussion

Spatial variations of sugarcane areas with climate change

Our model based on Random Forest offered a great accuracy (AUC=0.97, TSS=0.96) allowing for robust inference. Our model results indicated that the suitable planting areas for sugarcanes in the current distribution would increase by app. 44.2%, 66.6%, 104.6% (under RCP 2.6, RCP 4.5 and RCP 8.5, respectively) during the 21st century (Table 2 and Figure 4). However, some currently suitable habitats will experience a reduction, which is manly distributed in the Hainan Province. And most parts of the current sugarcane distribution areas would still remain (>96.3%) under each scenario. So strictly judged from the aspect of climate-suitable areas enlarging, China would benefit from climate change for sugarcane planting. We cannot speculate about other factors, yet, but likely more synergies are to occur as well.

According to our model predictions (Figure 4), Guangxi (excluding the northeast parts), Yunnan (the south parts) and all of the Guangdong Province would be the major suitable planting areas. The junction of Sichuan and Chongqing would be an additional suitable area. The climate boundary for cultivating sugarcane would shift northwards. The reasons for suitable areas enlarging and the boundary shifting north may be due to climate warming, which makes more regions in the north satisfy the climate demand especially for temperature conditions of sugarcane's life history. Based on our obtained results, we argue that the adjustment of the country's industrial structure plan, developing central-south of Guangxi, southwest of Yunnan, west of Guangdong and north of Hainan into three major sugarcane-planting areas (Fu, 2013) would contribute to the production of sugarcane in China and its economic development.

Environmental variables

From the variable ranking and importance value for sugarcane (Table 1 and Appendix 2), we found that the top six important variables for sugarcane growth were all temperature-relevant (BIO11, BIO7, BIO6, BIO9, BIO4, BIO1). Then from the 7th to 11th place (BIO12, BIO16, BIO18, BIO19, BIO13) were precipitation-relevant variables. Therefore and in the absence of other predictors than the 19 Worldclim ones, we inferred for now that temperature condition is more important than precipitation for sugarcane growth at the present time in China. Also, the temperature of the coldest month and quarter were more important than for the warmest ones. We

think that's because the environment need to satisfy the lowest temperature demand for growth (live) and for high temperature influence, plants could presumably regulate it through transpiration. Our results are consistent with (Su et al., 2006), who found sugarcane could not tolerate low temperature. We also found the precipitation of wettest quarter and month (BIO16 and BIO13) contributed more than the driest quarter and month (BIO17 and BIO14). That's probably because the driest quarter and month is in winter with low temperature, sugarcane don't grow much during this period (Zhang, 2013) While the wettest quarter and month are in summer with high temperature, sugarcane grows fast then with the required amount of water. Su et al. (2006) noted that little precipitation in Spring and Autumn would influence the production of sugarcane a lot.

Species distribution models with climate variables projection versus traditional climate regionalization

In the recent 15 years, Chinese researchers tried to apply small grid climate analysis methods in GIS to divide land into different suitable classes. This method was based on the relationship of the sugarcane growth and climate conditions (see Tu et al., 2003; Su et al., 2006). The advantage of this method is that it appears to be simple and it's easy to handle and within experts' physiology and ecology knowledge. Our projection of the crops' distribution with species distribution models uses a similar concept, but obtains physiology and ecology knowledge from the real cultivating locations using latest statistics methods (supervised classification, machine learning algorithms, bagging). It is based on the ecological niche and allows for interactions and synergies. Furthermore, the model approach we employed here doesn't take all variables as same weight. Instead it can calculate the individual contributions of variables to species distribution to achieve a best-possible outcome. And also, it could offer partial dependence plots of each predictor variable, which could display the relationship between the species distribution and individual predictors. These plots could help to increase people's physiology and ecology knowledge of species, not just from experience but from powerful statistics which is known to 'outsmart' experts (Elith et al., 2006, Drew et al., 2011). Finally, our method follows Open Access principles and thus allows for a repeatable and transparent science, to be tested in a quantitative fashion explicit in space and time (Drew and Perera, 2011). We would like to encourage investigators to use our data and output, assess it and push it forward. Our work offers itself easily for testing and improvements and once better and additional data come online. We hope our work triggers the creation and release of new data relevant for sugarcane and global food security. Therefore, our model method presents an advanced and more rigorous version of traditional climate regionalization based on small grids based on best available science.

Climate suitable areas versus real planting areas

Our model predicted the climate suitable distribution of sugar cane for the current as well as for 2070. We found the suitable planting areas in the future under three scenarios were larger than current estimates. Our findings do not mean necessarily that China would grow sugarcane in more areas and in new climate suitable places; namely, suitable areas don't automatically equal planting areas e.g. due to soil and other real-

world constraints. It depends on multiple factors whether farmers would plant sugarcane in a certain place, such as soil condition, sugarcane varieties improvement, water and fertilizer management, other crops planting, economic benefits, industrial structure, national policy (e.g. Returning Farmland to Forest Policy, subsidies), etc. However, overall and solely judged by the growing area, China will benefit from sugarcane's suitable planting area increase and will presumably have more choices to decide how to grow and obtain more sugar products. We have not addressed here other impacts of this prediction result and we suggest more studies about those synergies.

Limitations and future work

One probable shortcoming of our model is that we could not use all of the underlying surface conditions, such as soil status, nutrition availability etc. Another limitation is that we did not consider the irrigation situation, which could make unsuitable dry land to satisfy the demand of sugarcane growing. That's because here we started this investigation using species distribution models and for now, just considered the effects of climate change on crop distribution. For future work, we could try to predict the actual production in the suitable areas and integrate water and fertilizer management, economic development into a complicated sustainable development system. We believe considering an added tele coupling perspective would help.

Conclusion

Our sugarcane distribution model with the Random Forest algorithm had rather high AUC (0.97) and TSS (0.96) values indicating excellent performance. Our model results showed that climate change would generally extend sugarcane's suitable planting areas in China in the 21st century under all of the three conservative emission scenarios (RCPs of 2.6, 4.5 and 8.5). This includes a loss of few areas in the Hainan Province. Using WorldClim predictors, we found that the distribution of sugarcane largely depended on temperature predictors rather than precipitation as such. Based strictly on three future climate projections, our results indicated that China may benefit from climate change on sugar-production.

Acknowledgments

References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J. appl. ecol. 43, 1223-1232.
Beck, J., 2013. Predicting climate change effects on agriculture from ecological niche modeling: who profits, who loses? Climatic change. 116, 177-189.
Booms, T.L., Huettmann, F., Schempf, P.F., 2010. Gyrfalcon nest distribution in Alaska based on a predictive GIS model. Polar biol. 33, 347-358.
Breiman, L., 2001. Random forests. Mach. learn. 45, 5-32.

Chen, H., 2003. Impacts of climate change on sugarcane production in Fujian. Meteorological science and technology innovation and development of Atmospheric Science in the new century-Annual meeting of the Chinese Meteorological Society in 2003. (In Chinese)

Dietterich, T.G., 2000. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Mach. learn. 40, 139-157.

Drew, C.A., Perera, A.H., 2011. Expert knowledge as a basis for landscape ecological predictive models. Predictive Species and Habitat Modeling in Landscape Ecology, pp. 229-248. Springer.

Drew, C.A., Wiersma, Y., Huettmann, F., 2011. Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer.

Dyer, J.M., 1995. Assessment of climatic warming using a model of forest species migration. Ecol. Model. 79, 199-219.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Jin, L., Lohmann, L.G, Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend, P.A., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., S. Wisz, M., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 29, 129-151.

Estes, L.D., Beukes, H., Bradley, B.A., Debats, S.R., Oppenheimer, M., Ruane, A.C., Schulze, R., Tadross, M., 2013. Projected climate impacts to South African maize and wheat production in 2055: a comparison of empirical and mechanistic modeling approaches. Glob. change biol. 19, 3762-3774.

Evangelista, P., Young, N., Burnett, J., 2013. How will climate change spatially affect agriculture production in Ethiopia? Case studies of important cereal crops. Climatic change. 119, 855-873.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. conserv. 24, 38-49.

Friedman, J.H. 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics. pp. 1189-1232.

Fu, M., 2013. Optimization of the Sugar Cane industry chain of Guangdong [D]. Hainan University. (In Chinese)

Genuer, R., Poggi, J.M., Tuleau-Malot, C., 2010. Variable selection using random forests. Pattern Recogn. Lett. 31, 2225-2236.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol model. 135, 147-186.

Herrick, K.A., Huettmann, F., Lindgren, M.A., 2013. A global model of avian influenza prediction in wild birds: the importance of northern regions. Vet. Res. 44, 308-311.

Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob. change biol. 12, 2272-2281.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Andy, J., 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. 25, 1965-1978.

Hole, D.G., Willis, S.G., Pain, D.J., Fishpool, L.D., Butchart, S.H., Collingham, Y.C., Rahbek, C., Huntley, B., 2009. Projected impacts of climate change on a continent‐wide protected area network. Ecol Lett. 12, 420-431.

Iverson, L.R., Prasad, A.M., 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecol. Monogr. 68, 465-485.

Johnstone, J.F., Hollingsworth, T.N., Chapin, F.S., Mack, M.C., 2010. Changes in fire regime break the legacy lock on successional trajectories in Alaskan boreal forest. Global Change Biology. 16, 1281-1295.

Kandel, K., Huettmann, F., Suwal, M.K., Regmi, G.R., Nijman, V., Nekaris, K., Lama, S.T., Thapa, A., Sharma, H.P., Subedi, T.R., 2015. Rapid multi-nation distribution assessment of a charismatic conservation species using open access ensemble model GIS predictions: Red panda (Ailurus fulgens) in the Hindu-Kush Himalaya region. Biol. Conserv. 181, 150-161.

Knox, J.W., Díaz, J.A.R., Nixon, D.J., Mkhwanazi, M., 2010. A preliminary assessment of climate change impacts on sugarcane in Swaziland. Agri. Syst. 103, 63-72.

Li, R., Xu, M., Wong, M.H.G., Qiu, S., Li, X., Ehrenfeld, D., Li, D., 2015. Climate change threatens giant panda protection in the 21st century. Biological Conservation. 182, 93-101.

Li, T., Liu, W., 2013. Safety status of sugar industry in China. World Agriculture. 8, 89-91. (In Chinese)

Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography. 28, 385-393.

Liu, J., Peng, Y., Chen, Q., 2009. Climate Suitability and Regionalization of Sugar Cane Cultivation in Hainan. Chinese J. Agrometeo. 30, 254-256. (In Chinese)

Long, G., Li, G., Chen, Y., 1994. Impacts of climate change on sugar cane production in China. Chinese J. Agrometeo. 4, 23-25.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. appl. Ecol. 38, 921-931.

Marin, F.R., Jones, J.W., Singels, A., Royce, F., Assad, E.D., Pellegrino, G.Q., Justino, F., 2013. Climate change impacts on sugarcane attainable yield in southern Brazil. Climatic Change. 117, 227-239.

McPherson, J., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? J. appl. ecol. 41, 811-823.

Mi, C., Huettmann, F., Guo, Y., 2014. Obtaining the best possible predictions of habitat selection for wintering Great Bustards in Cangzhou, Hebei Province with rapid machine learning analysis. Chinese Sci. Bull. 59, 4323-4331.

Mi, C., Huettmann, F., Guo, Y., 2016. Climate envelope predictions indicate an enlarged suitable wintering distribution for Great Bustards (*Otis tarda dybowskii*) in China for the 21st century. Peer J. 4, 16-30.

Ohse, B., Huettmann, F., Lckert-Bond, S.M., Juday, G.P., 2009. Modeling the distribution of white spruce (Picea glauca) for Alaska with high accuracy: an open access role-model for predicting tree species in last remaining wilderness areas. Polar biol. 32, 1717-1729.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global ecol. biogeogr. 12, 361-371.

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 9, 181-199.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., Guisan, A., 2006. Are niche-based species distribution models transferable in space? J. Biogeogr. 33, 1689-1703.

Su, Y., Li, Z., Sun, H., 2006. Climate Division of Sugarcane Planting Based on GIS in Guangxi. Chinese J. Agrometeo. 27, 252-255. (In Chinese)

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science. 240, 1285-1293.

Tu, F., Kuang, X., Huang, X., 2003. Using small grids climate analysis method to make climate regionalization of sugarcane in Guangxi Province. GUANGXI SUGARCANE & CANESUGAR. pp. 8-11. (In Chinese)

Woodward, F.I., 1987. Climate and Plant Distribution. Geogr. J. 69, 189-197.

Wu, W., Gu, S., Wu, J., Cao, M., Juncheng, L., Xu, H., 2012. Impact of Climate Change on Distribution of Breeding Sites of Red-Crowned Crane in China. J. Ecol Rural Environ. 28, 243-248. (In Chinese)

Xie, P., Zhang, Y., Chen, S., 2006. Climate change and its effect on sugar cane production and sugar content in Xuwen. GUANGXI SUGARCANE & CANESUGAR, pp. 10-14. (In Chinese)

Xie, X., Yan, L., 2004. GIS -Based Small Grids Reckoning of Agricultural Climate Resources in Fujian Province. Journal of Fujian Normal University (Natural Science Edition). 20, 90-93. (In Chinese)

Zhai, T., Li, X., 2012. Climate change induced potential range shift of the crested ibis based on ensemble models. ACTA ECOLOGICA SINICA. 9, 1353-1362. (In Chinese)

Zhang, Y., 2013. Modern Sugar Industry. Science Press. (In Chinese)