Title: Quantifying uncertainties of ecological forecasting under climate change

Authors:

Xin Chen[1*], Ye Liang[2], Xiao Feng[1*]

1 Department of Geography, Florida State University, Tallahassee, FL, United States

2 Department of Statistics, Oklahoma State University, Stillwater, OK, United States

*Corresponding authors:

Xin Chen, Email: xin.chenxz@gmail.com

Xiao Feng, Email: fengxiao.sci@gmail.com

Classification: BIOLOGICAL SCIENCES (major); Ecology (minor)

Key words: climate change, predictor novelty, collinearity, collinearity shift, ecological forecasting, extrapolation

This PDF file includes:

Main Text
Fig. 1 to 4
SI Appendix Fig. 1 and 2

Abstract

Ecological forecasting is critical in understanding of ecological responses to climate change and is increasingly used in climate mitigation plans. The forecasts from correlative models can be challenged by novel environmental conditions and predictor collinearity that are common during model extrapolation. However, there is still a lack of comprehensive knowledge about how these factors interactively affect forecasting. We conducted modeling experiments to mimic a wide range of extrapolation scenarios commonly seen under climate change. We modeled three functional relationships using general linear model (GLM) and Random Forests algorithm (RF). We assessed how predictor novelty, training collinearity, collinearity shift, and model complexity interactively influenced model predictions. We found that predictor novelty and collinearity shift were two major factors for inflated errors. The prediction errors were doubled with moderate predictor novelty (predictors increased by 0.4) or considerable collinearity shift (correlation coefficient r changed > 0.9). Interestingly, we also found negative interactions between predictor novelty and collinearity shift. Model predictions became more erroneous as model complexity increased. Calibrating models using variables correlated with |r| < 0.7 has been a rule-of-thumb; building upon that, our study further recommends a threshold of < 0.4 increased predictor novelty and/or < 0.9 collinearity shift for forecasting models to avoid a substantial loss of model performance. GLM is a safer option than RF during forecasting because it is more tolerant to predictor novelty and collinearity shift when true predictors are known. One will be cautious with forecasting beyond our identified threshold regardless of modeling algorithm.

Introduction

Under future climate change, ecologists inevitably have to forecast species range shifts, invasion of non-native species, potential of vector-borne diseases, colonization and extinction risks, and many other ecological responses to altered ecological stressors. Correlative ecological models are widely used in forecasting ecological responses to various global environmental drivers such as climate change, anthropogenic land use and land cover change, and habitat fragmentation over time and space. Population models can be used for assessing the dynamics of biological populations by modeling population growth or decline (1, 2), planning and evaluating management actions by providing different conservation scenarios (3), and understanding how demographic rates contribute to population dynamics (4,5). Ecological niche models or species distribution models that relate ecological responses to environments have been widely used in forecasting range shifts and changes in species composition (6-8), species invasion (9-12), and risk of extinction (13-16). Moreover, ecological niche models have incorporated the Quaternary fossil record to forecast (or hindcast) species distributions and biodiversity over evolutionary time scales (17-19). Among those extensive studies, more models were found for explanation and prediction in the same time period and geographic region as those of calibration data rather than for projecting models in the past or future and/or in different regions (20).

However, ecological forecasting based on correlative models is always associated with varying errors and uncertainties (21-25). Projecting ecological models involves model interpolation and extrapolation (26, 27). Interpolation is referred as to predicting the response within the ranges of the environmental conditions used to calibrate a model while extrapolation is projecting a model to the conditions exceeding the ranges of its calibration conditions (21, 28-30). In most of cases, predicting ecological responses is not necessarily limited to either interpolation or extrapolation since one can be intrinsically accompanied by the other (31, 32), for example, predicting future species distributions, ecological patterns, or biodiversity in new geographic areas under climate change (6, 20, 33-36), raising the concerns about the reliability of model projections (37). Thus, it is of great significance to identify which factors potentially affect the model projections onto new environments and quantify the magnitude of these effects.

The ecological forecasting to future or new geographic regions under climate change is mainly induced by climate novelty. The climate is not changing evenly across space (38), neither over time. In the eastern United States, the last century has witnessed increased mean annual temperatures in the Midwest and Northeast while the Southeast had areas with not only increased but decreased temperature (8). In the meanwhile, climate shifts of more than 100 km across the Northeast and Upper Midwest in terms of spatial velocity have been observed (39). Most of this region experienced growing season temperatures during the 1971 to 2000 period that were cooler than those during the 1911 to 1940 period (40). As climates continue to shift across time and space, exceeding the ranges of those experienced in the past, no-analog climates are likely to exist in many regions across the globe (41, 42). Given the future climate novelty, the ecological forecasting relying on correlative models is prone to increased prediction errors and more uncertainties (6, 23, 28, 30, 42, 43).

The other major factor that contributes to forecasting error is collinearity. Collinearity usually refers to the linear relationship among predictor variables in a statistical model (22). Collinearity shift exists when a model is calibrated on data from one region or time, and predicted to another region or time with different structure of collinearity among predictors. The former could influence parameter estimation in correlative models while the latter may affect the accuracy of model prediction. Model interpolation may be reliable as long as the collinearity between variables remains constant (44), but models would become more erroneous when predicting to the changed collinearity structures that are different from calibration data (22, 43). Consequently, increased errors may emerge for ecological forecasting dependent on correlative models because the collinearity among the climatic predictor variables may change under future climate change. When the change in collinearity structure from calibration to testing data set (i.e. collinearity shift and hereafter) was small, the degradation of model performance was not pronounced (22). Nonetheless, the model performance moderately decreased as the collinearity slightly became less and further exerted a considerable loss when collinearity shifted to very small degrees (22). Unlike the negative impact induced by collinearity shifting to lower degrees on model performance, Feng et al. (43) also found that significantly higher collinearity shifts (i.e. collinearity shifting to higher degrees) led to decreased performance when models were projected to new geographic regions that were different from calibration areas. Thus, the relationship between collinearity shift and model performance needs more investigation.

Apart from predictor novelty and collinearity shift, some other factors have been suggested to have an influence on model performance, such as training collinearity, model complexity, and modeling algorithms. As for training collinearity, high training collinearity leads to poor prediction accuracy mainly by affecting parameter estimation for traditional regression models based on ordinary least squares, for example, the general linear model (GLM). However, it might not be the case for machine learning algorithms that do not rely on parametric approaches, for example, Random Forests (RF) can reduce the influence of collinearity among predictors by randomly sampling part of the whole data set and randomly selecting a subset of predictors at each node when building each regression or classification tree during model fitting process (45). Dorman et al. (22) demonstrated that machine learning methods such as RF, Boosted Regression Trees (46), and Multivariate Adaptive Regression Splines (47) worked reasonably well under moderate collinearity. Feng et al. (43) also found no correlation between model performance and training collinearity for another machine learning algorithm, Maximum Entropy algorithm (48) due to regularization during its model fitting process (48, 49). In spite of their tolerance to training collinearity, machine learning algorithms generally have limited predictive ability in model extrapolation (50, 51). Therefore, whether the training collinearity influences model performance may depend on modeling algorithms (22). Even though some strategies, for example integrating penalized parametric regression, for example Lasso, into Random Forests algorithm to allow for effective prediction outside the range of the calibration data, the effect of those strategies on model predictions remains quantitatively unknown under the conditions with changed predictor novelty and collinearity shift. Regarding model complexity, simpler models may be easier to be generalized and they work better than more sophisticated models when being projected to different times or new regions. Taking regression models as an example, as long as the parameter estimation has been affected by high training collinearity, larger prediction errors can be made from more sophisticated models rather than simpler models because the former has more parameters to estimate. Despite the potential influence of all the factors discussed above, there is still a lack of knowledge regarding how model complexity, training collinearity, collinearity shift, predictor novelty, and modeling algorithm affect predictive performance (i.e. prediction accuracy) of ecological model forecasting. Thus, there is a need to assess their relative importance to model prediction. Many studies focus mainly on how predictor novelty influences model performance in model extrapolation (23, 28, 51). One of the most recent evidence has illustrated the weak correlation between predictor novelty and collinearity by investigating the role of predictor novelty, collinearity shift, and training collinearity when projecting models to new regions (43). By comparing the explained variation of model performance, Watling et al. (52) found that modeling algorithm was more important than collinearity among predictor variables. More interestingly, for models with severe training collinearity, the influence of collinearity shift has been proved greater than that of training collinearity in decreasing model performance (22), suggesting that the influence of collinearity shift may be associated with the magnitude of training collinearity. Overall, these evidence motivated our interest in conducting a comprehensive assessment to disentangle the relative roles of each of the five factors including model complexity, training collinearity, collinearity shift, predictor novelty, and modeling algorithms in affecting predictive abilities of ecological forecasting models.

Here, we designed a modeling experiment that mimics model extrapolation scenarios that are common during ecological forecasting under climate change. We assessed model prediction errors using simulated 3 levels of training collinearity, 19 levels of collinearity shift, 7 levels of predictor novelty, 3 levels of model complexity, and 2 different modeling algorithms. We evaluated the robustness of ecological forecasting models to changed environmental conditions induced by climate change and quantified how much change in training collinearity and novelty the correlative models could tolerate for reliably projecting ecological responses to future climate change. Specifically, we ask (1) how much magnitude of changes in model performance could be induced by collinearity shift and increased predictor novelty? (2) Would the influences of collinearity shift and predictor novelty be dependent on model complexity, training collinearity, and modeling algorithms? (3) Would there be any interactions between the influences of collinearity shift and predictor novelty? By quantifying the prediction errors under different combinations of collinearity shift and predictor novelty scenarios, we aim to derive the rules of thumb for making reliable forecasts to different time and space with increased predictor novelty and changed correlations among predictors. Our comprehensive assessment will provide important insights into projecting correlative models to different times or new regions accompanied by collinearity shift and predictor novelty by quantifying how much prediction errors will increase under different simulated scenarios of collinearity shift and predictor novelty.

Results

Collinearity shift. We analyzed the Root Mean Squared Error (RMSE) of GLM and RF models for model predictions along four gradients including model complexity, training collinearity, collinearity shift, and predictor novelty.

The RMSE of GLM models increased when the correlation between X_1 and X_2 decreased or when the sign of correlation changed (e.g. from positive to negative). The RMSE of GLM models with high training collinearity were higher than those with medium and low training collinearity when collinearity shift occurred. Moreover, the RMSE increased and became much more variable along the gradients of model complexity from Linear to Product, training collinearity from Low to High, and collinearity shift with the correlation changed from -0.9 to 0.9) (Fig. 1A). Thus, collinearity shift negatively influenced prediction accuracy when the collinearity shifted to positive lower degrees or negative higher degrees. For GLM models with medium and high training collinearity, larger errors were found when predictors became more correlated.

Likewise, we detected the trends of increased RMSE for RF models along the gradients of model complexity from Linear to Product, training collinearity from Low to High, and collinearity shift with the correlation changed from -0.9 to 0.9) (*SI Appendix*, Fig. S2). When the training collinearity was high, RMSE was found to be more variable and larger compared to those of medium and low training collinearity (*SI Appendix*, Fig. S2). In contrast, for the most complex model, the RMSE of GLM and RF models had different responses to the collinearity shifted to positive higher degrees. The former did not change while the latter became larger and more variable after the correlation became higher than that in training data. Overall, RF models had relatively higher RMSE than GLM models across the gradients of model complexity, training collinearity, and collinearity shift given that the underlying relationship between response variable and predictor variables is multivariate linear relationship (Fig. 3 and *SI Appendix*, Fig. S2; see also Recommendations in the Discussion).

Predictor Novelty. Higher degrees of predictor novelty led to more variability and larger magnitude of RMSE, and this pattern was consistent across the gradients of model complexity, training collinearity, and collinearity shift for both GLM and RF models (Fig.1B and *SI Appendix*, Fig. S2; see also Recommendations in the Discussion). Again, RF models had relatively higher RMSE induced by the change in predictor novelty than GLM models for all gradients of interest. Therefore, predictor novelty affects prediction accuracy negatively under all testing scenarios for both GLM and RF models.

Variance partitioning. To quantify the contribution of training collinearity, collinearity shift, predictor novelty, and the interactions between collinearity shift and predictor novelty to RMSE, we ran ANOVAs to obtain the Type III Sum of Squares and then we calculated the partial R-squared for each factor that may have an influence on model performance measured by prediction accuracy (Fig. 2). Regardless of the modeling algorithms, the partial R-squares of predictor novelty increased as more gradients of novelty were considered in variation partitioning and predictor novelty had the higher partial R-squares than any other factors when the larger gradients of predictor novelty were considered in variation partitioning. For GLM models, when only accounting for two levels of predictor novelty including 0 and 0.2 in variation partitioning, training collinearity had the highest partial R-squared, followed by that of collinearity shift. In contrast, for RF models, predictor novelty consistently had the highest partial R-squares regardless of model complexity and the levels of novelty considered in variation partitioning. Collinearity shift had higher partial R-squares than training collinearity for the models with quadratic and product terms while collinearity shift had lower partial R-squares than training collinearity for the models with only multiple linear terms.

Interaction between collinearity shift and predictor novelty. We further assessed whether the effect of collinearity shift on prediction accuracy was dependent on predictor novelty for GLM and RF models by comparing the single effect between two gradients of collinearity shift along the gradient of predictor novelty. As Fig. 3 shows, there was a consistent negative interaction between collinearity shift and predictor novelty across the gradients of training collinearity when the degree of predictor novelty was less than 0.4 for RF models, suggesting that predictor novelty reduced the influence of collinearity shift on prediction errors. The similar trend was also found for GLM models under high and medium training collinearity. When the degree of predictor novelty was higher than 0.4, the negative interaction between collinearity shift and predictor novelty no longer existed for RF models mainly because RF models predicted the response variable as the same value as long as the values of predictors exceed their ranges in training data so that the prediction errors were mainly driven by the values of response variable.

Discussion

Under future climate change, reliable forecasts of different ecological responses such as species range shifts, invasion of non-native species, potential of vector-borne diseases, colonization and extinction risks to different changed ecological stressors is crucial to climate mitigation plans. We designed a simulated experiment to mimic the ecological forecasting under climate change, and investigated the effect of model complexity, training collinearity, collinearity shift, predictor novelty, and modeling algorithm on model predictive performance in a combined framework. Both collinearity shift and predictor novelty resulted in degradation of predictive performance of models in our simulation, suggesting that collinearity shift and predictor novelty would increase the risk of projecting ecological responses to climate change. When the testing became more novel than training data, predictor novelty caused higher prediction errors than collinearity shift. Interestingly, the interaction between collinearity shift and predictor novelty was negative because the effect of predictor novelty reduced the effect of collinearity shift on model predictions. Our results still highlights that GLM is robust to low and moderate collinearity and predictor novelty when the largest correlation between predictors was less than the rule-of-thumb $|r| < 0.7$ and the predictor novelty was less than 0.6. Machine learning models like RF were no more tolerant of collinearity shift than statistical models like GLM. The correlative models calibrated with low and medium training collinearity are more robust to collinearity shift and predictor novelty. Therefore, we encourage the choice of ecological forecasting models to avoid high collinearity (e.g. $|r| < 0.7$) and with parsimonious statistical association if the models have to be projected to the conditions with changed collinearity structure and predictor novelty. However, when the predictor novelty exceeds 1.5 times as large as the ranges of predictors in training data, the ecological forecasts based on correlative models will become unreliable (e.g. RMSE is 3-fold as much as that of no shift; Fig. 4) with large variation in model predictions regardless of the change in collinearity structure.

Extrapolation becomes problematic when ecological models are projected beyond the geographic or environmental range of the training data (22), leading to increased prediction errors especially for correlative models (57). This issue is mainly attributed to the change in the collinearity structure (22), collinearity shift, and predictor novelty (43) that are assumed to have negative influence on model predictions. Dormann et al. (22) also found the varying responses of prediction errors to different levels of training collinearity and modeling algorithms. Therefore, our study delved more into the investigation on the roles of different factors including model complexity, training collinearity, collinearity shift, predictor novelty, and modeling algorithms in affecting predictive performance of correlative models. Compared to the previous comprehensive investigation on collinearity only considering positive correlation among predictors (22, 58), we extended the gradient of collinearity shift by introducing the testing scenarios with the correlation of predictors shifting from high positive (r = 0.9) to high negative (r = -0.9), allowing us to identify not only the influence of collinearity shift on prediction errors but also how the magnitude of collinearity shift and the change in the signs of correlation between predictors affected prediction errors. In addition, we separated the changes in predictor novelty from those in collinearity to assess their effects on prediction errors, respectively. By simulating over one hundred testing scenarios across the gradients of collinearity shift and predictor novelty, we illustrated whether the interaction between these two factors existed under different degrees of training collinearity and model complexity, to our knowledge, which is the first one testing their interaction in simulation study.

Predicting species distributions or ecological communities using correlative models in new geographic regions or changed climatic conditions have encountered more errors (23, 30, 33, 35, 51). Similarly, we found the increased prediction errors of models along the gradients of collinearity shift and increased predictor novelty because the statistical relationship captured by the models may not apply well to new conditions. This pattern was more pronounced when the correlation of predictors shifted from positive to negative and the ranges of predictors became much higher than those of training data, respectively. More complex the models were and higher training collinearity the predictors had, the more degraded the predictive performance became. Moreover, our results show that the influence of predictor novelty was much greater than that of collinearity shift. This may be because predictor novelty results in more data points beyond the ranges of training data. The pattern of higher prediction errors due to increased predictor novelty rather than collinearity shift was more pronounced for predictions derived from RF models, suggesting that RF models might be more prone to errors when extrapolating ecological responses to new environmental conditions. RF models could be over-fitted on training data (22), or the predictive ability of RF models may be limited in linear extrapolation (50). Last but not the least, there was a negative interaction between collinearity shift and predictor novelty, suggesting that the influence of collinearity shift on prediction errors was reduced by increased predictor novelty.

Collinearity shift on model prediction. The negative influence of collinearity shift on prediction errors varied among low, medium, and high training collinearity among predictors. The highest prediction errors with the most variation induced by collinearity shift were found for the models with the most complexity and high training collinearity. However, collinearity shift did not considerably increase the prediction errors for the less complex models with less training collinearity (Fig.1A). Training collinearity is likely to result in imprecise parameter estimates even if they are unbiased (58) because of the inflated standard errors of model parameters (22). When the correlation among predictors were high, the estimates became less precise and scattered more widely (58), which may lead to higher and more variant prediction errors in the face of collinearity shift. In contrast, low correlations among predictors had little effect on parameter estimates (58, 59). Therefore, the correlative models with high training collinearity with $|r| > 0.7$ would be problematic when forecasting ecological responses under collinearity shift, which commonly occurs during model forecasts across space and time (54).

More interestingly, the prediction errors were dependent on the magnitude of collinearity shift and the change in the signs of correlation among predictors. The more the collinearity shifted, the higher the prediction errors became. This pattern is evident that, for example, in our results, the prediction error substantially increased when the correlation between predictors shifted from $r = 0.9$ to $r = -0.9$ (Fig.1A). Likewise, Dormann et al (22) found that the prediction errors consistently inflated when the collinearity shifted from high to almost zero regardless of modeling algorithms and the complexity of functional relationships, supporting our findings. In the context of changes in collinearity structure including of magnitude and sign of correlations, the correct parameter estimates might be not likely to be identified, thus the substantial changes in collinearity structure would lead to a detrimental effect on prediction accuracy (22). Accordingly, two highly correlated variables may not necessarily encounter more correlation shift than a pair of less correlated variables because the magnitude of collinearity shift depends not only the training collinearity but also the collinearity in the test data set (43). However, for the same magnitude of collinearity shift or shifting to the same collinearity structure, we expect higher prediction errors for the models with high training collinearity rather than low one. Further, the prediction errors would become increasingly larger if the sign of correlations among predictors change (i.e., positive correlations shifting to negative correlations, or vice versa).

Predictor novelty on model prediction. Similarly, we found that predictor novelty negatively affected predictive performance regardless of modeling algorithms and higher degrees of predictor novelty resulted in larger prediction errors than lower degrees of predictor novelty, illustrating that the increase in prediction errors depends on the magnitude of predictor novelty. The negative effect of novelty on model performance has been illustrated by previous studies modeling the relationship between species distributions or assemblages to the changing environments (23, 30, 51).

The impact of collinearity shift was even worse than that of training collinearity (22). This is well supported by our findings that the partial R-squares of collinearity shift was larger than that of training collinearity (Fig. 2). This might only apply to the testing scenarios where not much predictor novelty could be found because the prediction errors induced by predictor novelty were much greater than those by training collinearity and collinearity shift. We also found a negative interaction between the influence of predictor novelty and collinearity shift, suggesting that the increased predictor novelty could reduce the effect of collinearity shift on prediction errors and eventually masked its effect. Regardless of modeling algorithms, we found consistent larger influence caused by increased predictor novelty on prediction errors than that attributed to training collinearity and collinearity shift. Therefore, in the context of climate change, ecological forecasting based on correlative models would become more error-prone to increases in climate novelty rather than the changed correlation among climatic variables.

Implications for future studies. Ecological forecasting will be accompanied by much uncertainty because non-analog conditions are common under future climate (42), which forces ecologists to extrapolate correlative models for predicting ecological patterns, such as projecting responses of species distributions climate, species invasions to novel environment, and disease emergence in new areas. Some of the collinearity shift and predictor novelty we simulated could possibly emerge under non-analog climate change. By the end of 2090 CE, the range of future climate novelty is around 0.5 climatic distance across North America (23), which would lead up to more than 3-fold prediction errors inferred from our simulation (Fig. 1 and *SI Appendix*, Fig. S2). To mediate the negative impact of collinearity shift and predictor novelty on the reliability of ecological forecasting, one may consider not extrapolating further than 1/10 of the predictor range as a rule-of-thumb (60). However, this rule may not offer practical and straightforward guidelines for spatial planners and resources managers (61). Identifying the locations and time periods where model extrapolation with departures from reference conditions will occur before applying correlative models to ecological forecasting may make one aware of spatial and temporal limits in model prediction (62), or identifying where models are extrapolating outside of the training data by calculating the novelty (63). Another solution is to quantify 'forecast horizon' in terms of spatial, temporal, phylogenetic, and environmental dimensions by defining a measure of prediction quality and a threshold for acceptable forecast proficiency (64). Using mechanism models rather than correlative models for forecasting future ecological responses is also encouraged whenever possible (23).

Concluding remarks. To obtain reliable predictions < 2-fold change of RMSE, we derive the following recommendations for forecasting ecological models under climate change and many other altered ecological stressors inducing different degrees of increased predictor novelty and changed collinearity structure.

1) GLM models will be preferred rather than RF models for forecasting the models to novel conditions and varied collinearity structures, if the true set of predictors and the type of each relationship is known. For example, the net primary productivity (NPP) is only linearly associated with temperature, precipitation, and squared precipitation. When forecasting NPP under future climate, one may calibrate a GLM model because RF model may overfit on training data and is less tolerant to model extrapolation.

2) RF models will be only safe for reliable forecasts under very restricted conditions. Even with the true predictors correctly specified, RF models cannot make reliable forecasts when novelty is > 0.2 (Fig. 4 and *SI Appendix*, Fig. S2). However, when novelty is < 0.2 and collinearity shift < 0.9, RF models can make acceptable predictions.

3) Under low or medium (i.e. $|r| < 0.7$) training collinearity, we encourage the use of GLM models rather than RF models for model forecasting, because GLM models can stand more predictor novelty and collinearity shift. The prediction errors of GLM models are < 2-fold within 0.6 of predictor novelty across all gradients of correlation change. In contrast, RF models will only produce the forecasts with < 2-fold errors when the predictor novelty does not go beyond 0.3 with collinearity shift.

4) One should be cautious with forecasting regardless of modeling algorithms (GLM and RF) when training collinearity is high ($|r| \geq 0.7$), because most commonly used correlative models will yield degraded predictions under change in collinearity structure (22). Moreover, higher degrees of predictor novelty are expected to make model predictions more erroneous, for example, leading to > 5-fold change in RMSE (Fig. 1 and 4 and *SI Appendix*, Fig. S2). In order to produce reliable forecasts with < 2-fold errors, GLM models are only allowed for < 0.4 predictor novelty with collinearity shift (Fig. 4 and *SI Appendix*, Fig. S2). However, RF models will be more restricted to the scenarios with < 0.2 predictor novelty and < 0.9 changed r.

Overall, our study agrees with the rule-of-thumb of using variables correlated with $|r| < 0.7$, and provide a forward step by recommending a threshold of < 0.4 increased predictor novelty (increment > 0.4 of predictor range) and/or < 0.9 collinearity shift (correlation coefficient r changed < 0.9) for forecasting models to avoid substantially distorted model performance. Within the thresholds, while both GLM and RF models have acceptable predictions, our recommendation is to use GLM models rather than RF models for model forecasting. Because GLM models are more tolerant to predictor novelty and collinearity shift during model extrapolation given that the true set of predictors and their relationship with the response variable is known while RF models may fail in extrapolation ((50) Zhang, Nettleton, & Zhu 2017). Otherwise, making predictions with > 0.4 increased predictor novelty (increment > 0.4 of predictor range) or > 0.9 collinearity shift (correlation coefficient r changed > 0.9 will always be risky regardless of modeling algorithms.

Materials and Methods

We conducted simulations to investigate how each factor including degree of training collinearity, collinearity shift and predictor novelty affected predictive performance measured by prediction accuracy in different prediction scenarios. Further, we also investigated if their influences were associated with modeling contexts such as model complexity and modeling algorithms. To do so, we first simulated ecological responses using three different functional relationships associated with linear terms, linear and quadratic terms, and linear, quadratic, and product terms, respectively. We fitted models using GLM and RF algorithms for the data sets under 3 levels of training collinearity and calculated the prediction errors when predicting the true responses under different collinearity shift and predictor novelty scenarios. Using a variance partitioning approach, we ran an analysis of variance (ANOVA) to assess the relationship between model performance and four different factors including degree of training collinearity, collinearity shift, predictor novelty, and the interaction between collinearity shift and predictor novelty. Last but not the least, based on the simulation results, we derived the rules of thumb for making reliable forecasts to different time and space with increased predictor novelty and changed correlations among predictors. We described more details of our simulation in the following sections.

Modeling complexity. To conduct our simulation experiment, we created each training and testing data set that included 1,000 data points from a data-generating model given by

$Y = f(X) + \epsilon$, (1)

where *Y* is a continuous variable and X is a p-dimensional vector of predictors. The number of dimensions is 3 for f_{ln}, 4 for f_{quad} and 6 for f_{prdct} (see equations below). The independent random errors ϵ follow $N(0, 0.5^2)$. We considered three different functional relationships to represent different model complexities:

1. f_{ln} (Linear): $f(X) = 25 + 2X_1 + 1.5X_2 - 3X_3$, i.e. multiple linear terms

2. f_{quad} (Quadratic): $f(X) = 25 + 2X_1 - 2X_1^2 + 1.5X_2 - 3X_3$, i.e. multiple linear terms plus a quadratic term

3. f_{prdct} (Product): $f(X) = 25 + 2X_1 - 2X_1^2 + 1.5X_2 - 3X_3 + 1.5X_1X_2$, i.e. multiple linear, quadratic and a product term of two different predictors, where all predictor variables X_1, X_2, and X_3 follow multivariate normal distribution with zero mean and variance equal to one that allowed for manipulating correlations among all predictor variables. To note, we specified the three functional relationships with the coefficients at the same scale between 1.5 to 3 to make sure the effects of the three predictor variables on the response variable were at similar magnitude.

Training collinearity. To evaluate the influence of training collinearity on model performance, we introduced three levels of training collinearity by drawing values of the three predictors from multivariate normal distributions with different variance-covariance matrices. We defined the high training collinearity based on the widely used Pearson correlation coefficient threshold r = |0.7| (22, 53) as the pairwise Pearson correlation coefficients among the three predictors r = 0.9, 0.8, and 0.7, respectively. The medium training collinearity was set as the pairwise correlation coefficients being equal to 0.6, 0.5, and 0.4. We used a more restrictive threshold |r| = 0.4 (22) to define the low training collinearity with r = 0.3, 0.2, 0.1 for the three predictors.

Collinearity shift and predictor novelty. We produced the gradients of collinearity shift to mimic changes in collinearity structure when ecological forecasting models were projected over different time and space (also see Fig. 1 of Dorman et al. (22) and APPENDIX I of Jiménez-Valverde et al. (54))**.** To simply illustrate the influence of collinearity shift on predictive performance, we assumed that collinearity shift was mainly induced by the changes in the correlation between X_1 and X_2 throughout the simulations so that we set their pairwise Peason correlation coefficient r ranging from -0.9 to 0.9 with 0.1 increment simulating the gradient of collinearity shift in the testing data sets. As a result, collinearity shift was denoted by the relative difference between these 19 levels of correlation coefficient r in testing data and r of training data (*SI Appendix*, Fig. S1). The scenario in which r was the same as that of training data was deemed as no collinearity shift.

Practically, the values of X_1 and X_3 were held constant while the values of X_2 were resampled from $N(0, 1^2)$ in each testing data set to satisfy the given correlation coefficient between X_1 and X_2. We expected that larger magnitude of collinearity shift induced greater prediction errors for models. To account for the effect of predictor novelty on model performance, we simultaneously increased the values of one predictor by 0.2, 0.4, 0.6, 2, 4, 6 multiplied by its original range in the training data set, resulting in a gradient of 7 levels in total for predictor novelty with the original ranges being considered as no change. The multiples of increased range was used to denote the level of novelty, for example, 0.4 novelty means that the values of one predictor increased by 0.4 timed by its original range. Since we attempted to provide insights into ecological forecasting under climate change, we chose such gradient of predictor novelty to mimic climatic novelty faced in the real world; for example the climatic distance in North America with the maximum value around 0.6 (23), and the Euclidean distance of dissimilarities between 20th- and 21st- century climates across the globe with the maximum value around 6 (25). We assumed that higher predictor novelty led to higher prediction errors.

Modeling algorithms. For each level of model complexity and training collinearity, we fitted statistical models using general linear model (GLM) and Random Forests (RF; 45, 55) algorithm to predict their true responses. We chose these two modeling algorithms in order to compare the difference between a statistical method and a machine learning approach in dealing with training collinearity, collinearity shift, and predictor novelty. Both GLM and RF were useful in ecology for dealing with moderate collinearity (22). The parameter estimation may be biased by training collinearity for GLM models while RF models can tolerate correlations among predictors (45). Therefore, under severe training collinearity, machine learning approaches, for example, RF outperformed GLM-like methods yielding lower-error models (22). Nonetheless, GLM models may outperform RF models in linear extrapolation because tree-based models like RF models cannot make accurate predictions when predictor values are beyond their bounds in training data (50). Thus, we hypothesized that the predictive abilities of GLM models and RF models would not be considerably influenced by collinearity shift when training collinearity was low, but they would be affected by medium to high training collinearity and high levels of predictor novelty.

Data simulation and analysis. We implemented a full combination of every experimental dimensions: 3 levels of model complexity (f_{ln}, f_{quad}, and f_{prdct}), 3 levels of training collinearity (high, medium, low), 19 levels of collinearity shift, 7 levels of predictor novelty (0, 0.2, 0.4, 0.6, 2, 4, 6), and 2 modeling algorithms (GLM and RF). This led to a total of 2,394 different experimental setups. We produced the training data sets for 3 levels of model complexity f_{ln}, f_{quad}, and f_{prdct} by 3 levels of predictor novelty low, medium, and high training collinearity using 1000 simulated data points to train general linear models (GLM) and Random Forests (RF; 45, 55) models. To note, RF algorithm can intrinsically consider the interactions among predictors due to its tree-based characteristic so that it is not necessary to specify the quadratic and product terms in model statement, but we retained the same quadratic and product terms in RF models as those in GLM models when training the models for f_{quad}, and f_{prdct} allowing for a fair comparison. We then predicted the true response variable (Y) on 2,394 testing data sets using these two models, each representing a unique combination of levels of the three factors (3 levels of training collinearity x 19 levels of collinearity shift x 7 levels of predictor novelty x 3 model complexity x 2 modeling algorithms). Then the prediction errors were calculated across all testing data sets with the formula $Root\ Mean\ Squared\ Error\ (\mathrm{RMSE}) = \sqrt{\frac{\sum(Y-\hat{Y})^2}{n}}, n = 1000$. These were repeated 1000 times with random seeds to ensure robust prediction of response variable under each testing scenario.

Variance partitioning. Using a variance partitioning approach, we ran an analysis of variance (ANOVA) to assess the relationship between model performance and four different factors including degree of training collinearity, collinearity shift, predictor novelty, and the interaction between collinearity shift and predictor novelty. The ANOVA was performed at each level of model complexity and modeling algorithms to test whether these relationships were dependent on model complexity and modeling algorithm. All analyses were done in R 4.1.2 (56).

Acknowledgements

X.C. acknowledges the support for post-doctoral associate from Department of Geography at Florida State University.

Reference

1. H. Caswell, Matrix Population Models. *Encyclopedia of Environmetrics* (2012) https:/doi.org/10.1002/9780470057339.vam006m.

2. J. Millspaugh, F. R. Thompson, *Models for Planning Wildlife Conservation in Large Landscapes* (Academic Press, 2011).

3. D. E. Lee, M. K. L. Strauss, Giraffe Demography and Population Ecology. *Reference Module in Earth Systems and Environmental Sciences* (2016) https:/doi.org/10.1016/b978-0-12-409548-9.09721-9.

4. D. N. Koons, T. W. Arnold, M. Schaub, Understanding the demographic drivers of realized population growth rates. *Ecol. Appl.* **27**, 2102–2115 (2017).

5. S. Wilson, K. C. Gil-Weir, R. G. Clark, G. J. Robertson, M. T. Bidwell, Integrated population modeling to assess demographic variation and contributions to population growth for endangered whooping cranes. *Biological Conservation* **197**, 1–7 (2016).

6. J. Elith, J. R. Leathwick, Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697 (2009).

7. M. C. Fitzpatrick, A. D. Gove, N. J. Sanders, R. R. Dunn, Climate change, plant migration, and range collapse in a global biodiversity hotspot: the *Banksia* (Proteaceae)

of Western Australia. *Global Change Biology* **14**, 1337–1352 (2008).

8. K. Zhu, C. W. Woodall, J. S. Clark, Failure to migrate: lack of tree range expansion in response to climate change. *Global Change Biology* **18**, 1042–1052 (2012).

9. C. Bellard, *et al.*, Will climate change promote future invasions? *Glob. Chang. Biol.* **19**, 3740–3748 (2013).

10. R. Tingley, M. Vallinoto, F. Sequeira, M. R. Kearney, Realized niche shift during a global biological invasion. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10233–10238 (2014).

11. M. P. Hill, B. Gallardo, J. S. Terblanche, A global assessment of climatic niche shifts and human influence in insect invasions. *Global Ecology and Biogeography* **26**, 679–689 (2017).

12. C. Liu, C. Wolter, W. Xian, J. M. Jeschke, Species distribution models have limited spatial transferability for invasive species. *Ecol. Lett.* **23**, 1682–1692 (2020).

13. W. Thuiller, S. Lavorel, M. B. Araújo, M. T. Sykes, I. C. Prentice, Climate change threats to plant diversity in Europe. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 8245–8250 (2005).

14. D. Nogués-Bravo, R. Ohlemüller, P. Batra, M. B. Araújo, Climate predictors of late quaternary extinctions. *Evolution* **64**, 2442–2449 (2010).

15. M. B. Araújo, D. Alagador, M. Cabeza, D. Nogués-Bravo, W. Thuiller, Climate change threatens European conservation areas. *Ecol. Lett.* **14**, 484–492 (2011).

16. S. Jarvie, J.-C. Svenning, Using species distribution modelling to determine opportunities for trophic rewilding under future scenarios of climate change. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373** (2018).

17. C. E. Myers, A. L. Stigall, B. S. Lieberman, PaleoENM: applying ecological niche modeling to the fossil record. *Paleobiology* **41**, 226–244 (2015).

18. E. Martínez-Meyer, A. Townsend Peterson, W. W. Hargrove, Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography* **13**, 305–314 (2004).

19. D. Nogués-Bravo, J. Rodríguez, J. Hortal, P. Batra, M. B. Araújo, Climate change, humans, and the extinction of the woolly mammoth. *PLoS Biol.* **6**, e79 (2008).

20. M. B. Araújo, *et al.*, Standards for distribution models in biodiversity assessments. *Sci Adv* **5**, eaat4858 (2019).

21. J. S. Clark, *et al.*, Ecological forecasts: an emerging imperative. *Science* **293**, 657–660 (2001).

22. C. F. Dormann, *et al.*, Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46 (2013).

23. M. C. Fitzpatrick, *et al.*, How will climate novelty influence ecological forecasts? Using the Quaternary to assess future reliability. *Glob. Chang. Biol.* **24**, 3575–3586 (2018).

24. K. A. Heilman, *et al.*, Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty. *Glob. Chang. Biol.* **28**, 2442–2460 (2022).

25. J. W. Williams, S. T. Jackson, J. E. Kutzbach, Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 5738–5742 (2007).

26. A. Townsend Peterson, *et al.*, *Ecological Niches and Geographic Distributions (MPB-49)* (Princeton University Press, 2011).

27. R. K. Heikkinen, M. Marmion, M. Luoto, Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* **35**, 276–288 (2012).

28. M. C. Fitzpatrick, W. W. Hargrove, The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation* **18**, 2255–2261 (2009).

29. D. Zurell, J. Elith, B. Schröder, Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions* **18**, 628–634 (2012).

30. H. L. Owens, *et al.*, Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* **263**, 10–18 (2013).

31. A. T. Peterson, A. Townsend Peterson, M. Papeş, M. Eaton, Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* **30**, 550–560 (2007).

32. R. Muscarella, *et al.*, ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity forMaxentecological niche models. *Methods in Ecology and Evolution* **5**, 1198–1205 (2014).

33. W. Thuiller, Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology* **10**, 2020–2027 (2004).

34. A. Guisan, W. Thuiller, Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009 (2005).

35. M. B. Araújo, C. Rahbek, How Does Climate Change Affect Biodiversity? *Science* **313**, 1396–1397 (2006).

36. S. Ferrier, A. Guisan, Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* **43**, 393–404 (2006).

37. A. Morán-Ordóñez, J. J. Lahoz-Monfort, J. Elith, B. A. Wintle, Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions? *Global Ecology and Biogeography* **26**, 371–384 (2017).

38. A. Prasad, *et al.*, Combining US and Canadian forest inventories to assess habitat suitability and migration potential of 25 tree species under climate change. *Diversity and Distributions* **26**, 1142–1159 (2020).

39. S. R. Loarie, *et al.*, The velocity of climate change. *Nature* **462**, 1052–1055 (2009).

40. R. W. McEwan, J. M. Dyer, N. Pederson, Multiple interacting ecosystem drivers: toward an encompassing hypothesis of oak forest dynamics across eastern North America. *Ecography* **34**, 244–256 (2011).

41. A. Ordonez, J. W. Williams, J.-C. Svenning, Mapping climatic mechanisms likely to

favour the emergence of novel communities. *Nature Climate Change* **6**, 1104–1109 (2016).

42. J. W. Williams, S. T. Jackson, Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment* **5**, 475–482 (2007).

43. X. Feng, D. S. Park, Y. Liang, R. Pandey, M. Papeş, Collinearity in ecological niche modeling: Confusions and challenges. *Ecol. Evol.* **9**, 10365–10376 (2019).

44. F. E. Harrell, *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis* (Springer Science & Business Media, 2013).

45. L. Breiman, Machine Learning. **45**, 5–32 (2001).

46. J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *The Annals of Statistics* **28** (2000).

47. J. H. Friedman, Multivariate Adaptive Regression Splines. *The Annals of Statistics* **19** (1991).

48. S. J. Phillips, R. P. Anderson, R. E. Schapire, Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231–259 (2006).

49. J. Elith, *et al.*, A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**, 43–57 (2011).

50. H. Zhang, D. Nettleton, Z. Zhu, Regression-Enhanced Random Forests (2019) https:/doi.org/10.48550/arXiv.1904.10416 (November 19, 2022).

51. H. Qiao, *et al.*, An evaluation of transferability of ecological niche models. *Ecography* **42**, 521–534 (2019).

52. J. I. Watling, *et al.*, Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. *Ecological Modelling* **309-310**, 48–59 (2015).

53. B. L. Welch, *et al.*, *Identifying Proxy Sets in Multiple Linear Regression: An Aid to Better Coefficient Interpretation* (1994).

54. A. Jiménez-Valverde, Y. Nakazawa, A. Lira-Noriega, A. Townsend Peterson, ENVIRONMENTAL CORRELATION STRUCTURE AND ECOLOGICAL NICHE MODEL PROJECTIONS. *Biodiversity Informatics* **6** (2009).

55. D. R. Cutler, *et al.*, RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology* **88**, 2783–2792 (2007).

56. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for StatisticalComputing, Vienna, 2022).

57. N. J. Briscoe, *et al.*, Forecasting species range dynamics with process-explicit models: matching methods to applications. *Ecol. Lett.* **22**, 1940–1956 (2019).

58. H. Schielzeth, *et al.*, Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution* **11**, 1141–1152 (2020).

59. M. B. Morrissey, G. D. Ruxton, Multiple Regression Is Not Multiple Regressions: The Meaning of Multiple Regression and the Non-Problem of Collinearity. *Philosophy,*

Theory, and Practice in Biology **10** (2018).

60. C. F. Dormann, Response to Comment on “Methods to account for spatial autocorrelation in the analysis of species distributional data: a review.” *Ecography* **32**, 379–381 (2009).

61. K. L. Yates, *et al.*, Outstanding Challenges in the Transferability of Ecological Models. *Trends Ecol. Evol.* **33**, 790–802 (2018).

62. P. J. Bouchet, *et al.*, dsmextra: Extrapolation assessment tools for density surface models. *Methods in Ecology and Evolution* **11**, 1464–1469 (2020).

63. J. Elith, M. Kearney, S. Phillips, The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1**, 330–342 (2010).

64. O. L. Petchey, *et al.*, The ecological forecast horizon, and examples of its uses and determinants. *Ecol. Lett.* **18**, 597–611 (2015).

Figures

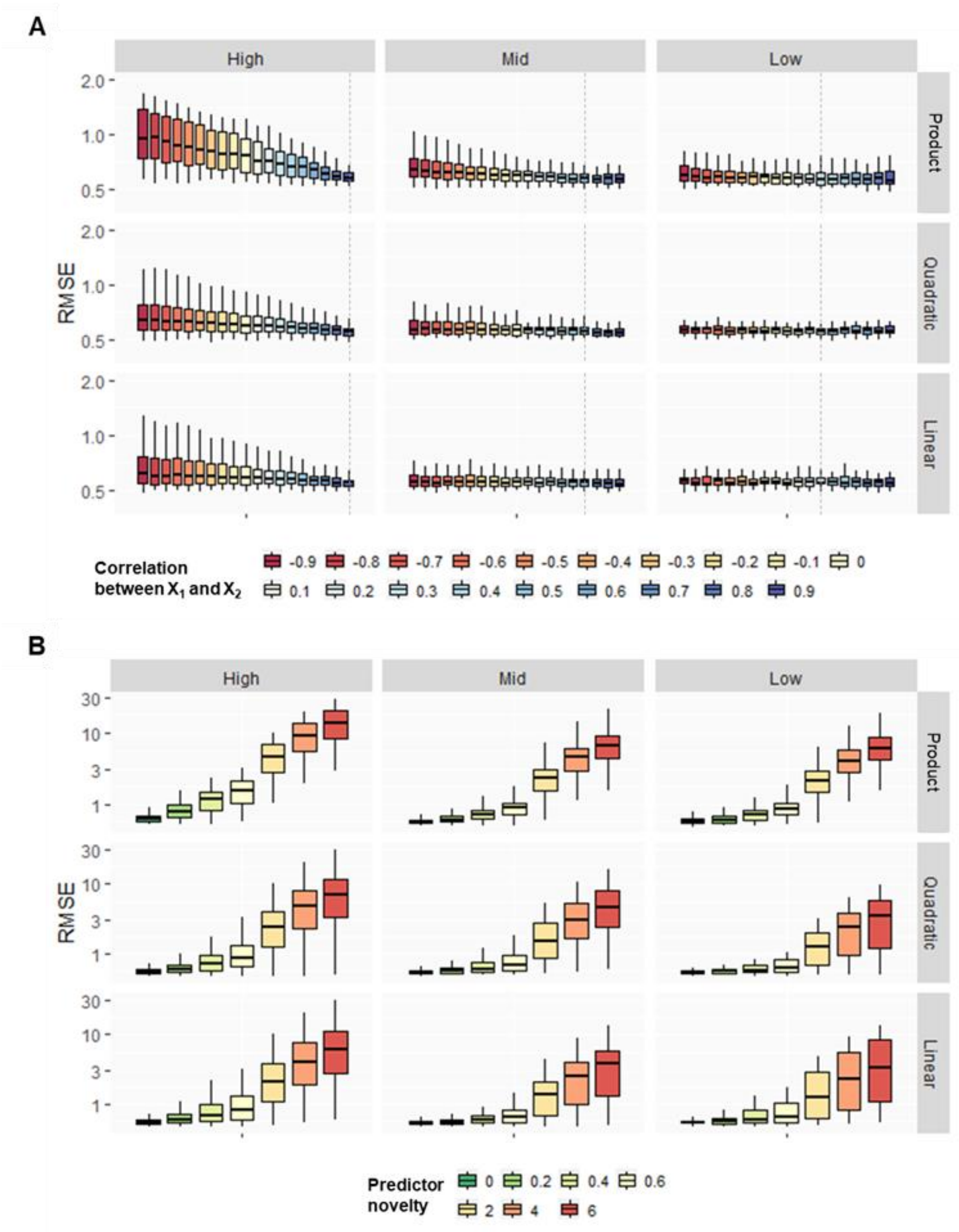

Fig. 1. Root Mean Square Errors (RMSE) derived from the prediction using GLM models across all simulations along the gradient of correlation between X_1 and X_2 (A) and predictor novelty (B) grouped by the three different levels of training collinearity (High, Medium, and Low) and model complexity (Product, Quadratic, and Linear). The collinearity shift was denoted by the correlation between X_1 and X_2 in the testing data sets ranging from -0.9 to 0.9. The vertical dashed lines in each panel indicate the correlation between X_1 and X_2 in the

training data sets. The y-axis was log-transformed. The predictor novelty was represented by the magnitude of increased ranges of X_2 in the testing data sets. The y-axis was log-transformed in each panel.

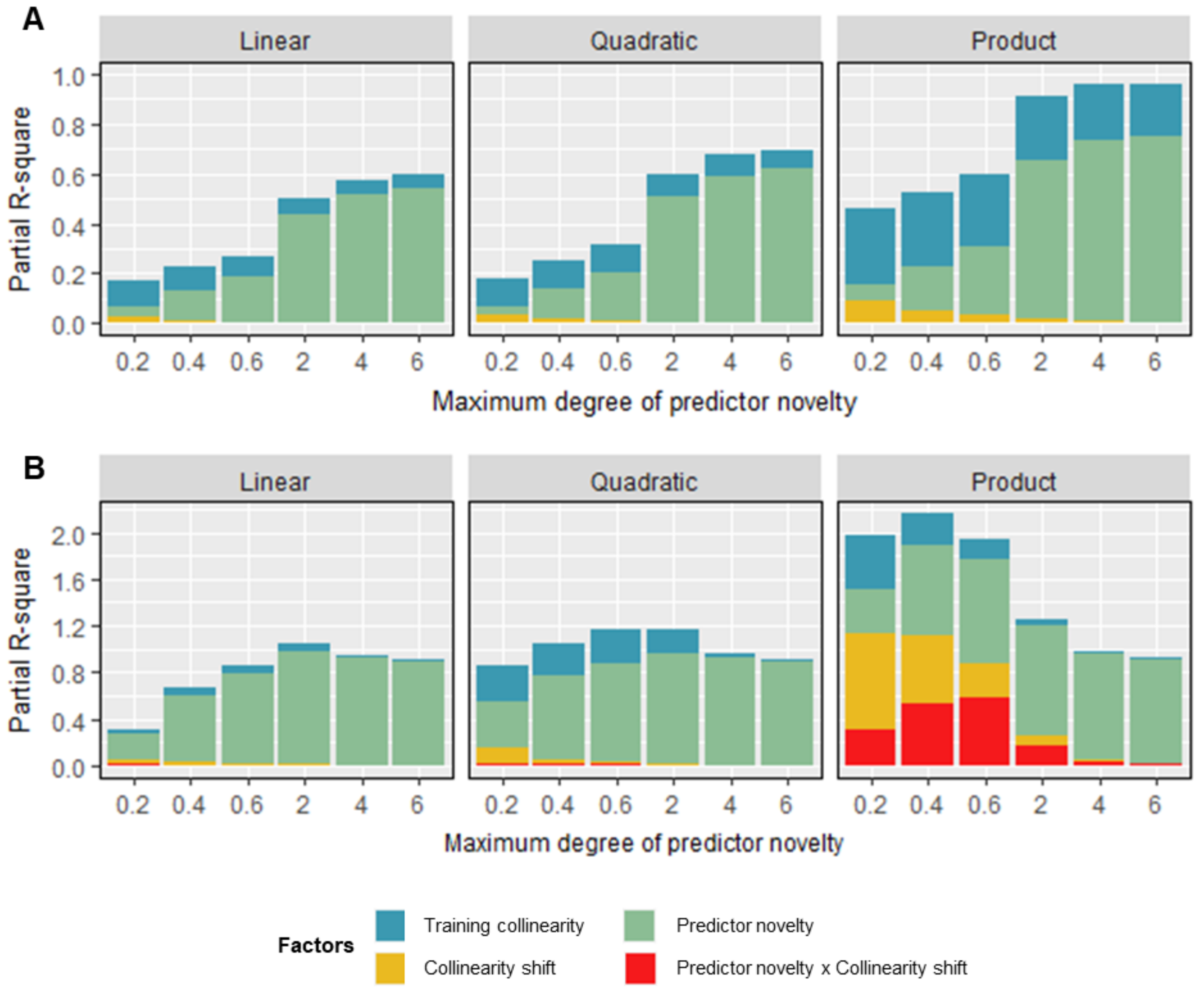

Fig. 2. Partial R-squares of each of four different factors including collinearity shift, training collinearity, predictor novelty, and interaction between predictor novelty and collinearity shift conditional on the rest of other three factors derived from Type-III Sum of Squares for GLM models (A) and RF models (B) with three different levels of complexity (Linear, Quadratic, and Product). A higher partial R-square suggests greater importance of a factor to Root Mean Square Errors. X-values depict the highest degree of predictor novelty used in variation partitioning through ANOVAs. That is, 0.2 on x-axis integrated 0 and 0.2 predictor novelty in ANOVA while 6 included all levels of predictor novelty from 0 to 6.

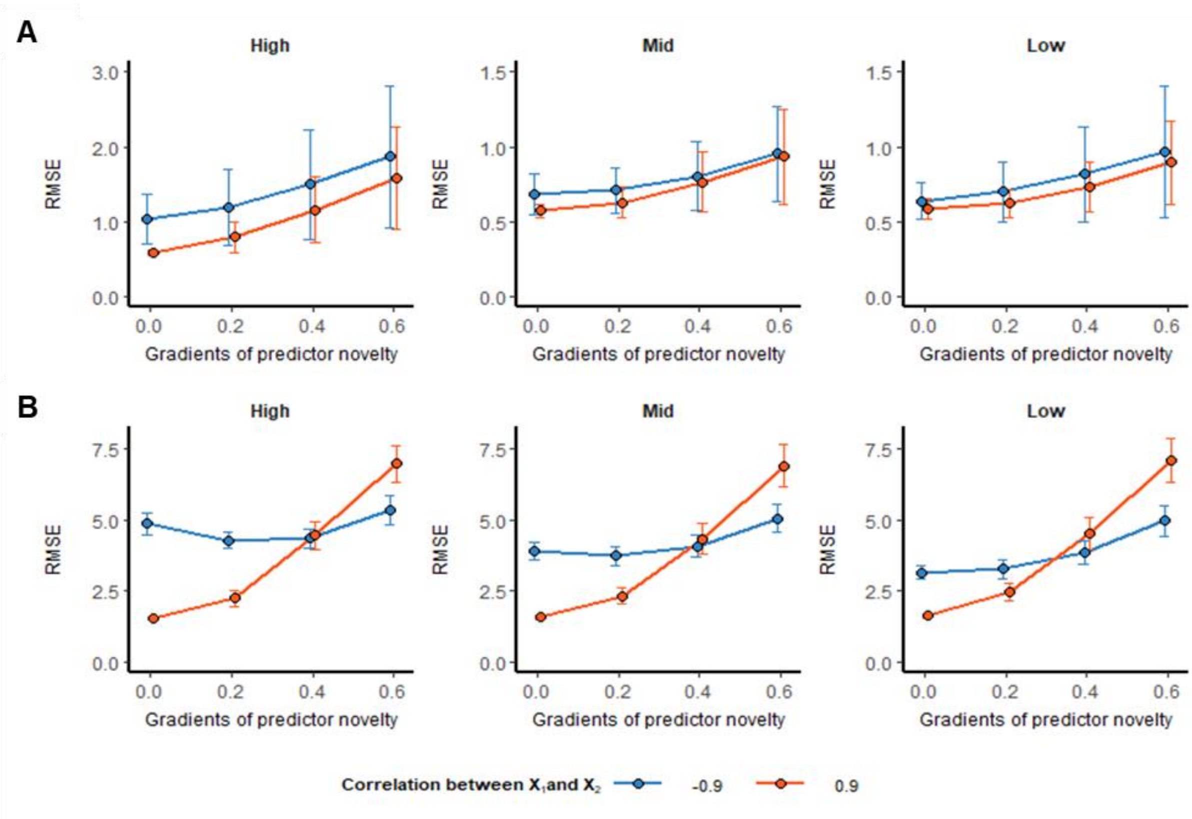

Fig. 3. Root Mean Square Errors (RMSE) derived from the predictions using GLM models (A) and RF models (B) fitting Product functional relationship across High, Medium, and Low training collinearity. The RMSE were compared along the four gradients of predictor novelty (x-axis) between two degrees of collinearity shift with the correlation between X_1 and X_2 shifted from 0.9 (red) to -0.9 (blue). The decreased difference in RMSE between two degrees of collinearity shift along the gradients of predictor novelty indicates negative interaction between predictor novelty and collinearity shift, suggesting that predictor novelty reduces the influence of collinearity shift on prediction accuracy.

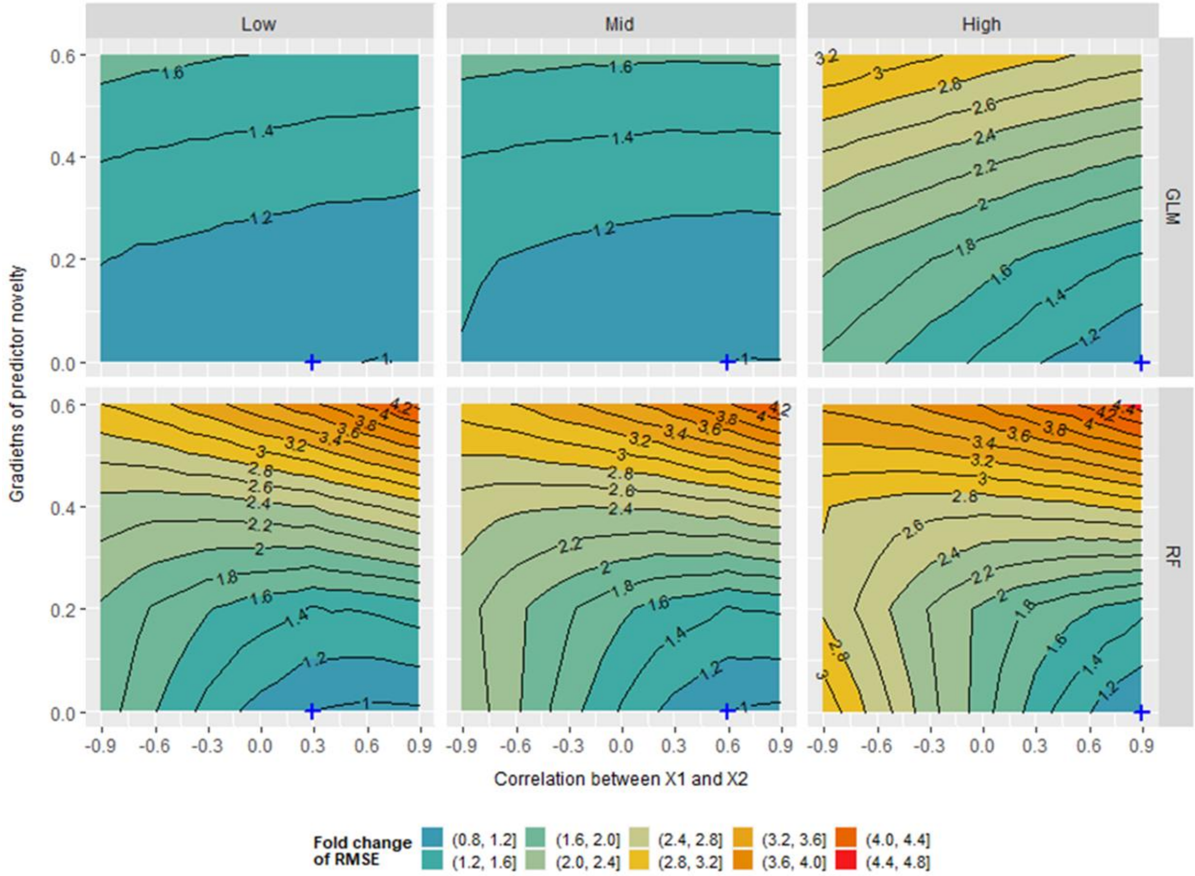

Fig. 4. Fold change of Root Mean Square Errors (RMSE) in the scenarios with increased predictor novelty and collinearity shift relative to the scenario with no change for GLM and RF models fitting Product functional relationship. Shading color indicates the fold change of RMSE under different test scenarios compared to that without any predictor novelty and collinearity shift. More intense colors denote higher prediction errors. The gradient of collinearity shift is denoted by the correlation between X_1 and X_2 along the x-axis. The location of the blue cross refers to the scenario with no predictor novelty and collinearity shift (i.e. correlation between X_1 and X_2 remained the same as that of the training data).

Supplementary Materials

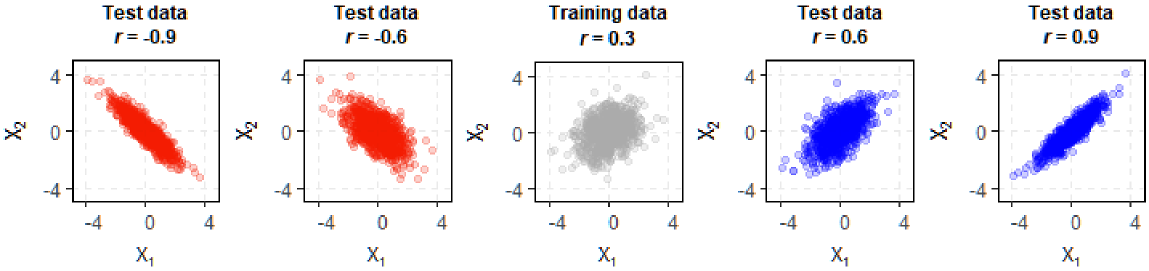

Fig. S1. Illustration of collinearity shift denoted by the Pearson correlation coefficient r between X_1 and X_2 in the training and testing data. The coefficient r between X_1 and X_2 in the training data set is 0.3. Blue dots represent the testing data with collinearity shifted by 0.3 and 0.6 while red dots depict those with collinearity shifted by -0.9 and -1.2).

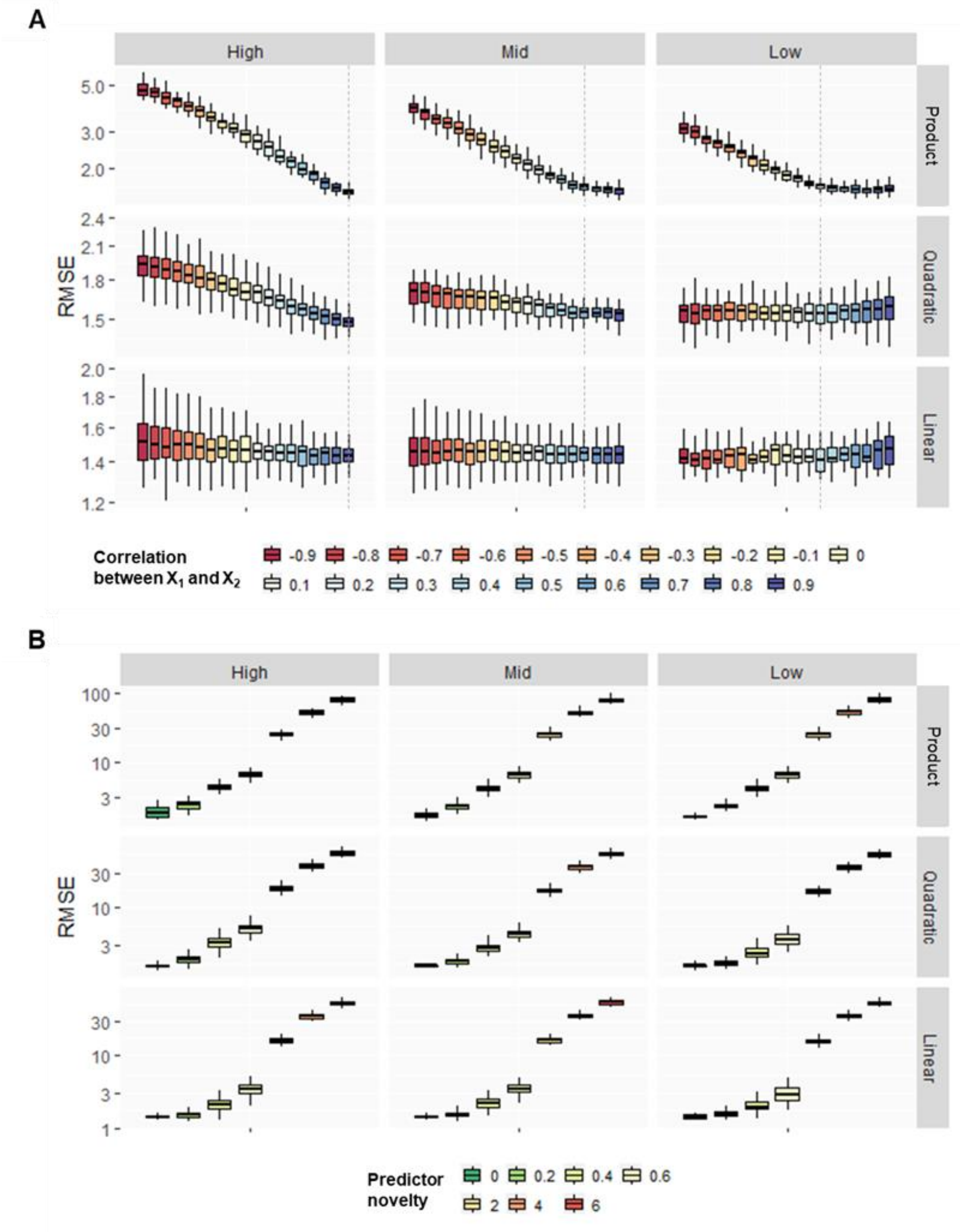

Fig. S2. Root Mean Square Errors (RMSE) derived from the prediction using RF models across all simulations along the gradient of correlation between X_1 and X_2 (A) and predictor novelty (B) grouped by the three different levels of training collinearity (High, Medium, and Low) and model complexity (Product, Quadratic, and Linear). The collinearity shift was represented by the correlation between X_1 and X_2 in the testing data sets ranging from -0.9 to 0.9. The vertical dashed lines in each panel indicate the correlation between X_1 and X_2 in the training data sets. The y-axis was log-transformed. The predictor novelty was represented by the magnitude of increased ranges of X_2 in the testing data sets. The y-axis was log-transformed in each panel.