# Prediction modeling of isoscapes for stable isotopes in China's rice based on environmental similarity

**Meiling Sheng**
**Weixing Zhang**
China National Rice Research Institute
**Jing Nie**
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products
**Chunlin Li**
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products
**A-Xing Zhu**
Key Laboratory of Virtual Geographic Environment (Nanjing Normal University)
**Hao Hu**
Zhejiang Academy of Agricultural Sciences
**Weidong Lou**
Zhejiang Academy of Agricultural Sciences
**Xunfei Deng**
Zhejiang Academy of Agricultural Sciences
**Xiaonan Lyu**
Zhejiang Academy of Agricultural Sciences
**Zhouqiao Ren**
Zhejiang Academy of Agricultural Sciences
**Karyne M. Rogers**
Zhejiang Academy of Agricultural Sciences
**Yongzhi Zhang**
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products
**Yuwei YUAN** ( ywytea@163.com )
State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products https://orcid.org/0000-0001-6707-2373

**Article**

**Keywords:** isoscapes, rice, origin, environmental similarity, stable isotope, third law of geography

**Posted Date:** December 29th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-1198139/v1

# Prediction modeling of isoscapes for stable isotopes in China's rice based on environmental similarity

Meiling Sheng[a, c, d], Weixing Zhang[#b], Jing Nie[a, c, d], Chunlin Li[a, c, d], A-Xing Zhu[e], Hao Hu[c, d], Weidong Lou[c, d], Xunfei Deng[c, d], Xiaonan Lyu[c, d], Zhouqiao Ren[c, d], Karyne M. Rogers[c,f], Yongzhi Zhang[a, c, d**], Yuwei Yuan[a, c, d*]

[a] State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Hangzhou 310021, China

[b]China National Rice Research Institute, Hangzhou, 310006, China

[c] Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

[d] Key Laboratory of Information Traceability for Agricultural Products, Ministry of Agriculture and Rural Affairs, P.R. China, Hangzhou 310021, China

[e] Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China

[f] National Isotope Centre, GNS Science, 30 Gracefield Road, Lower Hutt 5040, New Zealand

**Abstract**

Rice quality is directly related to human health, so it is important to have traceability systems that can trace inferior or contaminated rice back to its geographical origin. This ensures farming practices in substandard regions become better regulated to improve rice quality, origin labelling and consumer trust. However, tracing the origin of rice on the marketplace requires an accurate database benchmarking the isotope distribution over areas of rice production. Large stable isotope data sets can be used to determine the geographical origin of rice through predictive isoscape models. This study presents the first rice isoscape based on environmental similarity to predict the geospatial distribution of $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ values of Chinese rice and provides uncertainty at every location such prediction is made. For this study, 794 rice samples were collected in 2017 from primary rice production regions of China. An independent verification shows that the predicted isotope distribution from this new approach is of high accuracy, with a root mean squared error (RMSE) of 0.51 ‰, 7.09 ‰ and 2.06 ‰ for $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ values respectively. In addition, uncertainty in the spatial distribution of isotopes can be used to indicate the prediction accuracy and to guide future sampling. Our results indicate that an isoscape prediction method based on environmental similarity is effective to predict the spatial distribution of stable isotope in rice,

and is an effective tool for building isotope distribution in rice over large areas with complex environment. This method could also be used to predict potential isotopic variations in future years due to climate change.

**Key words:** isoscapes; rice; origin; environmental similarity; stable isotope; third law of geography

Food production and security play an important role in regional social stability and economic development of China and the rest of the world[1,2]. Rice is the most widely consumed food for more than half of the world's population[3]. Rice quality is attributed to human well-being and health[4,5]. As global living standards improve, consumers are increasingly interested to know the geographical origin of rice they consume[6]. This growing interest in the origin of rice provides impetus for developing methods to accurately verify the geographical origin rice and support its safe consumption[7].

One method used to investigate authenticity and geographical origin of rice is stable isotope analysis[8-13]. This method assumes that the isotope ratios of the plant reflect the features of its growing environment[14]. Specifically, the stable isotope composition of different elements in organic plant tissues retain important geographic and climatic information that describes the conditions of the organic matter formation[15]. Therefore, carbon($\delta^{13}$C), hydrogen ($\delta^{2}$H) and oxygen ($\delta^{18}$O) stable isotope values found in rice reveal their native regions[13].

Given the wide distribution of rice, a stable isotope traceability method requires comprehensive sampling from various environments across China. However, the associated costs of field sampling and laboratory analysis limits the actual number of rice sample points that can be collected. Recently developed geospatial modeling tools provide a novel method in the field of food authentication that can overcome spatially limited datasets. West et al.[16] proposed a term called "isoscapes," reflecting a combination of "isotopes" and "landscapes". Isoscapes are geospatial models based on independent environmental variables that predict the isotopic composition of elements such as C, H, O and N (that are integral in biogeochemical processes)[17].

Over the last decade, a number of studies have investigated the use of isoscapes in the fields of water quality[18], ecology[19-21] and food authentication[14,22,23]. However, to our knowledge there are no published studies of isocapes undertaken for the purpose of estimating rice provenance. Recent methods for predicting isoscapes mainly focus on the use of kriging techniques[18,20,22]. Chiocchini et al.[14] used a

59 combined regression-geostatistical approach to predict $\delta^{13}C$ and $\delta^{18}O$ values of Italian extra-virgin olive
60 oils. These methods are limited by the requirement of stationarity assumption over the study area. But
61 the stationarity requirement is often not met for spatial prediction across complex geographic processes[24].
62 To overcome the stationarity requirement, Zhu et al.,[24] proposed a new technique for spatial prediction
63 based on the Third Law of Geography, where the spatial prediction is based on the similarity of
64 geographic environments between a sample and a prediction point.
65 The aim of this work was to develop and test a geospatial model for carbon, oxygen and hydrogen
66 stable isotope variability in rice based on the similarity of geographic environment (Third Law of
67 Geography[24]), and to examine its suitability for the authentication of the geographical origin of Chinese
68 rice.
69
70 **Results and Discussion**
71 **Correlations between stable isotopes and environmental factors.** The environmental variables
72 selected in this study are described in Table S1. Pearson's correlation coefficients (r) between the
73 environmental variables and stable isotopes ($\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$) are shown in Table 1. All environmental
74 variables except SSD (SunShine Duration) and $SSD_g$ (SunShine Duration during growing season (from
75 June to October)) revealed a strong negative correlation relationship with $\delta^{13}C$ and $\delta^{18}O$. The strongest
76 negative correlation coefficient was found between covariates $\delta^{13}C$ and annual precipitation (PRE). SSD
77 and $SSD_g$ variables showed strong positive correlations with $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$. SSD showed a stronger
78 positive correlation than $SSD_g$ for $\delta^{13}C$ and $\delta^{2}H$ with coefficients of 0.62 and 0.48 respectively. RHU
79 (Relative humidity) showed the strongest negative correlation for $\delta^{2}H$ and $\delta^{18}O$ with coefficients of -0.40
80 and -0.49, respectively. $GDD_g$ (Growing degree days base 10 ℃) and $T_g$ (Average annual temperatures
81 during growing season (from June to October)) exhibited no significant correlations with $\delta^{2}H$. Some
82 studies have shown that the carbon isotope values of plant carbohydrates record an impact of
83 photosynthesis[25,26] while the oxygen and hydrogen isotope values record water-related effects on plant
84 carbohydrates[27,28]. Camin et al.[29] found that the $\delta^{18}O$ values of Italian wine had the strongest correlation
85 with climate information. Xia et al.[30] noted that stable isotopes of green tea were influenced by
86 temperature and light conditions, and Deng et al.[31] noted that $\delta^{2}H$ of green tea was significantly correlated
87 with temperature, sunshine duration and precipitation. Overall, the relationships found between

environmental variables and stable isotope ratios are similar to previous studies.

Table 1. Correlation coefficients between environmental variables and stable isotopes ($\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$).

| | GDD | $GDD_g$ | T | $T_g$ | RHU | $RHU_g$ | PRE | $PRE_g$ | SSD | $SSD_g$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\delta^{13}C$ | -0.63** | -0.61** | -0.63** | -0.61** | -0.69** | -0.64** | -0.74** | -0.59** | 0.62** | 0.58** |
| $\delta^{2}H$ | -0.21** | -0.07 | -0.16** | -0.06 | -0.40** | -0.22** | -0.24** | -0.19** | 0.48** | 0.37** |
| $\delta^{18}O$ | -0.36** | -0.26** | -0.35** | -0.25** | -0.49** | -0.47** | -0.43** | -0.44** | 0.43** | 0.49** |

Note: The specific explanation of the variables is shown in Table S1.

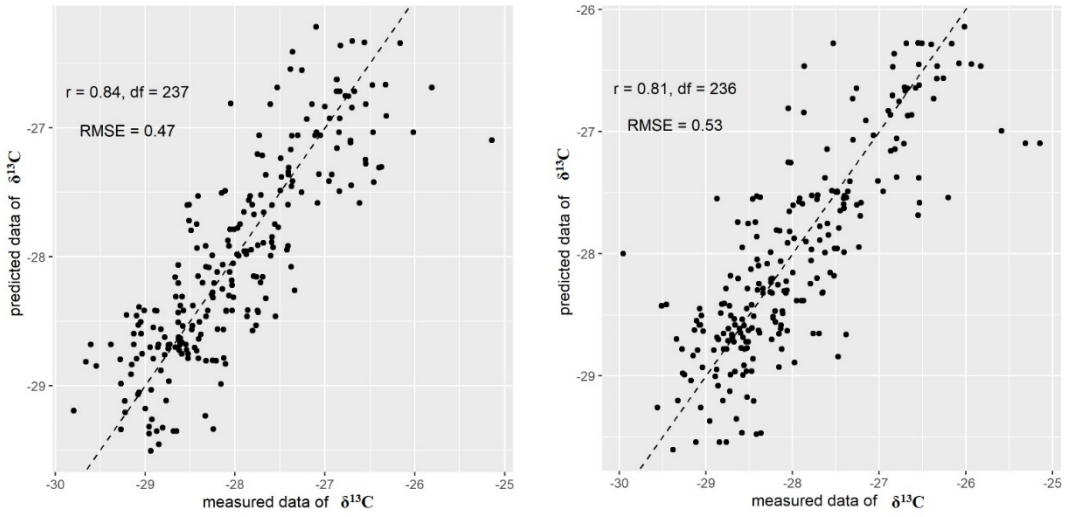
*, **: correlation is significant at the 0.05 and 0.01 levels (2-tailed), respectively.

**Stable isotope mapping and variable evaluation.** Environmental variables with significant correlation ($p < 0.01$) were selected for predicting isoscapes of $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ in rice. For $\delta^{13}C$ and $\delta^{18}O$ values, all environmental variables (Table 1) were used to predict the isoscapes. $\delta^{2}H$ isoscapes were predicted using GDD, T, RHU, $RHU_g$, PRE, $PRE_g$, SSD and $SSD_g$ environmental variables. To obtain optimal isoscape predictions, three thresholds; environmental similarity threshold (p1), target variable similarity threshold (p2) and reliability threshold (p3) were set with different combinations. These three threshold parameters ranged from 0 to 1 over intervals of 0.1.

**$\delta^{13}C$ isoscapes for rice.** The $\delta^{13}C$ variable evaluation results, based on 10 independent validation samples, are shown in Table 2. The mean RMSE was 0.51 ‰ with a range of 0.47 ‰ to 0.54 ‰. The MAE ranged from 0.36 ‰ to 0.42 ‰ with a mean MAE of 0.39 ‰. The correlation coefficient between observed and predicted $\delta^{13}C$ values ranged from 0.78 to 0.84. The mean correlation coefficient between observed and predicted $\delta^{13}C$ values was 0.82. The coefficient of determination ($R^2$) between observed and predicted $\delta^{13}C$ values ranged from 0.61 to 0.69, with a mean $R^2$ of 0.66. This reveals that the predicted values were close to the observed values for the validation dataset. P1 was uniform at 0.9 for the evaluation of 10 random samples while p2 ranged from 0.5 to 0.9 and p3 varied from 0.2 to 0.9 (Table S2). The optimal threshold value changed with different evaluation samples. Scatterplots of observed versus predicted $\delta^{13}C$ values for the sixth and tenth independent validation datasets are shown in Figure 1. The dotted line represents the 1:1 regression line. Most validation samples are evenly distributed on both sides of the 1:1 line. This suggests that the rice $\delta^{13}C$ isoscape model based on the geospatial similarity method is unbiased and performs well.

Table 2. Geographic prediction evaluation results and optimal threshold values based on 10 random samples for $\delta^{13}C$ values of rice. RMSE = root mean squared error. MAE = mean average error, $R^2$=coefficient of determination. r=Pearson's correlation coefficient. p1= the environmental similarity threshold. p2= the target variable similarity threshold. p3= the reliability threshold.

| Sampling event | RMSE (‰) | MAE (‰) | $R^2$ | r | Optimal threshold values | | |
|---|---|---|---|---|---|---|---|
| | | | | | p1 | p2 | p3 |
| 1 | 0.52 | 0.40 | 0.66 | 0.81 | 0.9 | 0.5 | 0.8 |
| 2 | 0.54 | 0.42 | 0.64 | 0.81 | 0.9 | 0.5 | 0.7 |
| 3 | 0.53 | 0.41 | 0.65 | 0.81 | 0.9 | 0.8 | 0.5 |
| 4 | 0.50 | 0.38 | 0.68 | 0.83 | 0.9 | 0.9 | 0.3 |
| 5 | 0.52 | 0.40 | 0.66 | 0.82 | 0.9 | 0.6 | 0.8 |
| 6 | 0.47 | 0.36 | 0.69 | 0.84 | 0.9 | 0.8 | 0.6 |
| 7 | 0.50 | 0.40 | 0.61 | 0.78 | 0.9 | 0.7 | 0.9 |
| 8 | 0.49 | 0.37 | 0.69 | 0.83 | 0.9 | 0.8 | 0.7 |
| 9 | 0.52 | 0.41 | 0.64 | 0.81 | 0.9 | 0.9 | 0.3 |
| 10 | 0.53 | 0.39 | 0.64 | 0.81 | 0.9 | 0.9 | 0.2 |
| Mean | 0.51 | 0.39 | 0.66 | 0.82 | 0.9 | 0.74 | 0.58 |

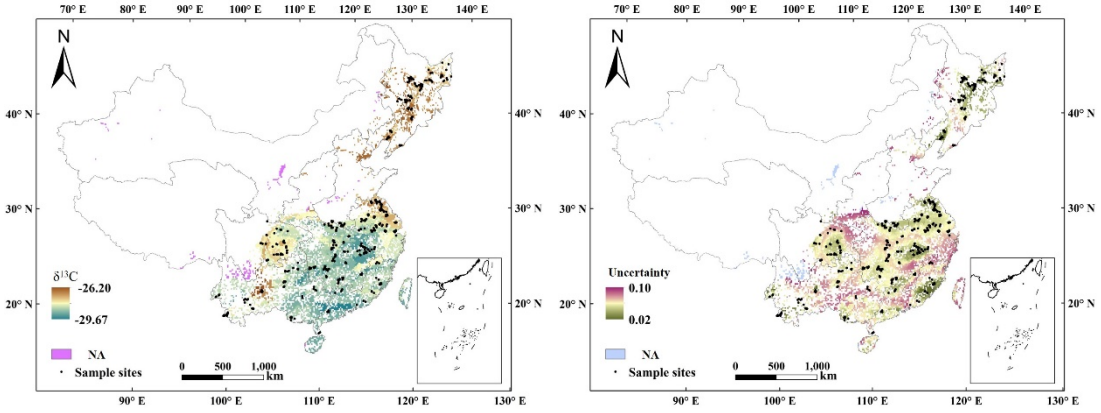



(a) (b)

Figure 1. Scatter plots of measured versus predicted $\delta^{13}C$ values for the (a) sixth and (b) and tenth independent validation datasets. High r values indicate unbiased model predictions.

Figure 2a shows the predicted spatial distribution of $\delta^{13}C$ with the similarity-based method. The predicted rice $\delta^{13}C$ values ranged from -29.67‰ to -26.20‰, , and lay between the maximum and minimum values of the actual values. The purple areas on the map indicated locations where the $\delta^{13}C$ values could not be predicted due to the absence of available data. Rice from central and southern regions (blue shaded color in Figure 2a) are characterized by lower $\delta^{13}C$ than those from northeast regions ((brown shaded color in Figure 2a). Figure 2b shows the spatial distribution of uncertainty values from

the similarity-based method. The overall uncertainty was relatively small and ranged from 0.02 to 0.10, resulting in a relatively high confidence level for the predicted values (blue shading in Figure 2a). The uncertainty values are relatively small in areas where the samples are concentrated. And in marginal areas with fewer samples, there is higher predicting uncertainty.




(a) (b)

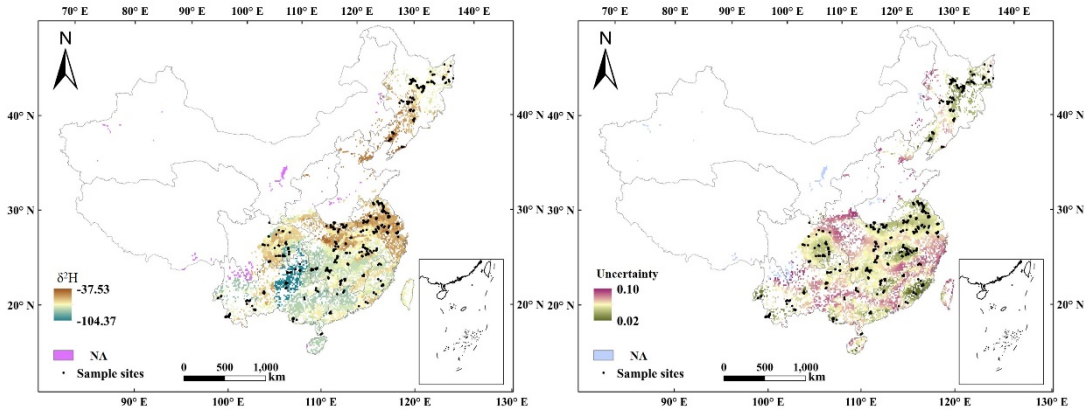
Figure 2. Spatial distribution of (a) predicted $\delta^{13}C$ isoscape for Chinese rice and (b) prediction uncertainty.

**$\delta^{2}H$ isoscapes for rice.** The $\delta^{2}H$ evaluation result based on 10 random are shown in Table S2. RMSE ranged from 6.40 ‰ to 7.38 ‰ with a mean RMSE of 7.09 ‰. MAE ranged from 4.67 ‰ to 5.58 ‰, with a mean MAE of 5.21 ‰. The correlation coefficient between observed and predicted $\delta^{2}H$ values for the validation dataset ranged from 0.77 to 0.83 with a mean correlation coefficient of 0.81. The coefficient of determination ($R^2$) between observed and predicted $\delta^{2}H$ values ranged from 0.57 to 0.68, with a mean $R^2$ of 0.62. Mean predicted $\delta^{2}H$ values were closer to the observed values for the validation dataset. The optimal value of p1 was set at 0.8 or 0.9, p2 was set at 0.6, 0.8 or 0.9 and p3 varied from 0.1 to 0.8. The optimal threshold values were similar to the predictive $\delta^{13}C$ thresholds.

Scatter-plots of the observed $\delta^{2}H$ values against predicted values for the sixth and tenth independent validation datasets are shown in Figure S1. Most validation samples were evenly distributed on both sides of the 1:1 regression line indicating that the $\delta^{2}H$ rice isoscape model based on the similarity method performed well.

Figure 3a shows the predicted $\delta^{2}H$ spatial distribution. The predicted rice $\delta^{2}H$ values ranged from -104.3 ‰ to -37.5 ‰, and lay between the maximum and minimum values of the actual samples. The purple areas on the map indicated locations where the $\delta^{2}H$ values could not be predicted due to the absence of available data. Rice was characterized by higher $\delta^{2}H$ values in the northeastern regions, middle and lower Yangtze River regions (brown shaded color in Figure 3a, conversely to the blue shaded

colour in other regions). Figure 3b shows the spatial distribution of the uncertainty values from the similarity-based method. The overall uncertainty was relatively small and ranged from 0.02 to 0.10, resulting in a relatively high confidence level for the predicted values (blue shading in Figure 3a). The uncertainty spatial distribution of predicted $\delta^2$H values was similar to the $\delta^{13}$C values. Overall we found that the sampling plan could be formulated according to the spatial distribution of uncertainty, where more samples should be taken in areas with high uncertainty and fewer samples taken in areas with low uncertainty, which reduces sampling costs.




(a) (b)

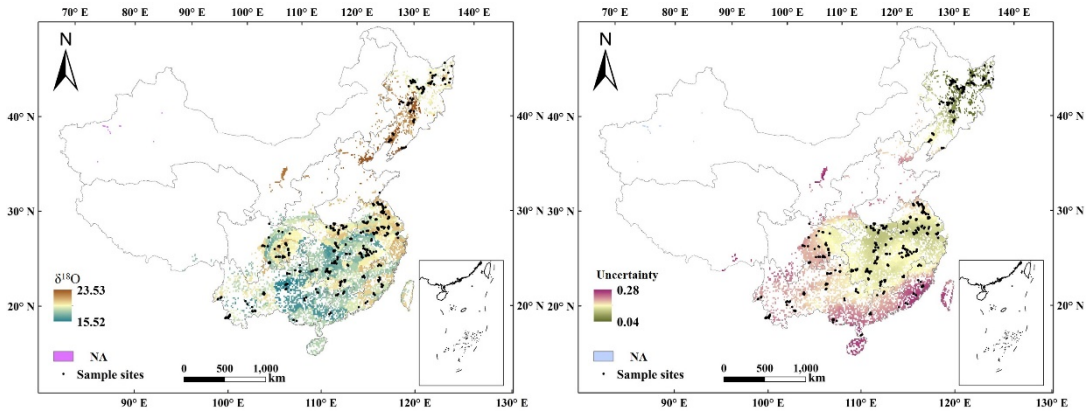
Figure 3. Spatial distribution of (a) prediction $\delta^2$H isoscape for Chinese rice (a) and (b) prediction uncertainty.

**$\delta^{18}$O isoscapes for rice.** Table S3 shows the $\delta^{18}$O results based on 10 independent validations. The mean RMSE was 2.06 ‰ with a range of 1.89 ‰ to 2.21 ‰. The MAE ranged from 1.29 ‰ to 1.45 ‰ and the mean MAE was 1.36 ‰. The correlation coefficient between observed and predicted$\delta^{18}$O values for the validation dataset ranged from 0.59 to 0.65 with a mean correlation coefficient of 0.63. The coefficient of determination ($R^2$) between observed and predicted $\delta^{18}$O values ranged from 0.34 to 0.42, and mean $R^2$ was 0.38. For the 10 independent validations, the optimal value of p1 was set 0.7, 0.8 or 0.9, p2 was set 0.1, 0.6 or 0.7 and p3 was set 0.7 or 0.8. These p1 and p2 settings to predict $\delta^{18}$O values were similar to $\delta^{13}$C and $\delta^2$H.

Figure S2 shows the scatterplots of the observed $\delta^{18}$O values plotted against predicted $\delta^{18}$O values for the sixth and tenth independent validation datasets. Most validation samples were evenly distributed on both sides of the 1:1 line, suggesting the $\delta^{18}$O rice isoscape model, based on a similarity method, performs well, similar to the $\delta^{13}$C and $\delta^2$H predictive models.

All rice samples were used to predict the $\delta^{18}$O isoscapes using the similarity-based method. The

predicted spatial distribution of $\delta^{18}O$ is shown in Figure 4a. The predicted $\delta^{18}O$ values ranged from 15.5 ‰ to 23.5 ‰. Rice was characterized by higher $\delta^{18}O$ values in northeastern and coastal regions than those from other regions. Figure 4b shows the spatial distribution of uncertainty values with the similarity-based method. The uncertainty values ranged from 0.04 to 0.28, with those from the southern marginal areas having the highest uncertainty.




(a) (b)

Figure 4. Spatial distribution of (a) predicted $\delta^{18}O$ isoscape for Chinese rice and (b) prediction uncertainty.

**Comparison with regression-geostatistics method.** The similarity-based method was compared with the regression-geostatistics method[14]. The same environmental covariates and validation samples were used in regression-geostatistics method to ensure consistency and comparability between methods.

The prediction accuracies from the validation samples for the similarity-based method and regression-geostatistics method are shown in Table 3. The mean RMSE and MAE values of predicted $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ values for the similarity-based method were lower than those from the regression-geostatistics method. The mean r and $R^2$ values of predicted $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ values for the similarity-based method were higher than those from the regression-geostatistics method. The similarity-based method used the individual sample representativeness to reflect stable isotope spatial heterogeneity. Consequently, the similarity-based method had a higher accuracy using the same number of samples and environmental covariates.

Table 3. Prediction accuracy of similarity-based and regression-geostatistics methods. RMSE = root mean squared error. MAE = mean average error. $R^2$=coefficient of determination. r=Pearson's correlation coefficient.

| | Similarity-based method | | | | Regression-geostatistics method | | | |
|---|---|---|---|---|---|---|---|---|
| | RMSE (‰) | MAE (‰) | $R^2$ | r | RMSE (‰) | MAE (‰) | $R^2$ | r |
| $\delta^{13}C$ | 0.51 | 0.39 | 0.66 | 0.82 | 0.54 | 0.43 | 0.62 | 0.79 |
| $\delta^{2}H$ | 7.09 | 5.21 | 0.62 | 0.81 | 8.83 | 6.70 | 0.42 | 0.65 |
| $\delta^{18}O$ | 2.06 | 1.36 | 0.38 | 0.63 | 2.11 | 1.53 | 0.34 | 0.59 |

In this study, we investigated spatial $\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$ variations in Chinese rice and environmental climate data combined with a similarity-based prediction method. The method not only provided a high accuracy of origin from samples in the dataset, it was also able to quantify the prediction uncertainty at unvisited locations. Using this similarity-based method to generate the prediction uncertainty, sampling and analytical costs can be reduced, as it is only necessary to undertake more intensive sampling in under-represented regions. This environmental similarity isoscape method can be used to define an origin verification isotope protocol in order to certify the origin of rice. Future research will include predicting temporal stable isotope variations and investigating the number of sampling points required to influence the prediction result.

## Methods

**Background theory.** The isoscape prediction method developed for this study uses the similarity in geographic environment which is theoretically based on the third law of geography[24]. It can be simply summarized as "the more similar the geographic environment is between two locations, the more similar the geographic features are between them". It has been used in different research fields, such as mapping soil types[32], soil organic matter[33,34], soil salinity[35], landslide risk assessment[36], a spatial prediction of fleas transmitting plague[37] and so on.

Carbon, oxygen and hydrogen isotope ratios reflect the environmental effects of photosynthesis and water-related processes in plants[38, 39]. Therefore, based on the relationships between isotopic ratios and environmental factors, the more similar the environmental conditions are between two locations, the more similar the isotopic ratios are likely to be.

**Detailed design of the proposed method.** There are four steps used to construct isoscapes for spatial prediction of the carbon, oxygen and hydrogen isotope ratios based on environmental similarity.

(1) Characterization the geographic environment using stable isotopes

Carbon isotope values of plant carbohydrates record the environmental effects from

photosynthesis[25,26]. Oxygen and hydrogen isotope values of plant carbohydrates reflect water-related processes in plants[27,28]. It is necessary and important to quantitatively characterize the environmental conditions using stable isotopes at each location. Presently, relevant studies have only explored the correlation between stable isotopes and environmental factors in agricultural products[29-31, 40].This research uses climate information for the first time to characterize the stable isotope environmental conditions.

(2) Calculating environmental similarities

Environmental similarities can be computed in two steps[34]. The first step computes the similarity values based on individual environmental covariates. In this study, we used the Gower similarity coefficient[41] (Eq. 1). The second step integrates similarities from all environmental variables into one variable representing the similarity at a given sample location. A minimum operator was adopted[42] in this research.

$$E(e_{vi},e_{vj})=1-\frac{|e_{vi}-e_{vj}|}{Range(v)} \quad (1)$$

Where E(•) represents the function evaluating similarity of the individual environmental variable level; $e_{vi}$ and $e_{vj}$ are the values of the $v$-th environmental variable at location $i$ and location $j$; $Range(v)$ represents the value range of the $v$-th environmental variable.

(3) Calculating the credibility of each sample

The relationship between samples can be divided into supporting samples, contradictory samples and supplementary samples according to the environmental similarity and target variable ($\delta^{13}$C, $\delta^{2}$H or $\delta^{18}$O) similarity. If the environmental conditions of the samples are similar and their target variable values are also similar, then the relationship between the two samples is supportive. If a sample has many supporting samples, the reliability of this sample is high[43].

The environmental similarity and target variable similarity between samples are calculated based on the second step. The environmental similarity threshold (p1) and the target variable similarity threshold (p2) determine the relationship between the sample points. If the sample has only contradictory points and no support points, then the reliability of the sample is 0; if there are neither support points nor contradictory points, then the reliability of the point is unknown and set to a null value (NA) (Eq. 2).

$$r_i = \begin{cases} \frac{\sum_{k=1}^{n_s} TS_{i,k}}{n_s} \times \frac{n_s}{n_s+n_c}, & n_s + n_c > 0 \text{ and } n_s \neq 0 \\ 0, & n_c > 0 \text{ and } n_s = 0 \\ NA, & n_s + n_c = 0 \end{cases} \quad (2)$$

Where $r_i$ represents the reliability of sample *i*; $n_s$ *and* $n_c$ represent the number of support samples and the number of contradictory samples for sample *i*, respectively ; $TS_{i,k}$ is the similarity of target variable between samples *i* and *k*.

(4) Predicting the stable isotope value and calculating the uncertainty value

In order to set the reliability threshold (p3) of the sample, only samples with high reliability are selected for prediction. The stable isotope value at an unsampled location point is predicted using Eq. (3).

$$V_j = \frac{\sum_{i=1}^{n'} S_{ji} \times V_i}{\sum_{i=1}^{n'} S_{ji}} \quad (3)$$

Where $n'$ is the number of samples that met the prediction conditions; $S_{ji}$ is the environmental similarity between the unvisited location point *j* and the sample *i*; $V_i$ is the target variable value ($\delta^{13}$C, $\delta^{2}$H or $\delta^{18}$O) of the sample *i*.

The uncertainty of prediction at each location is inversely related to its environmental similarities and reliability to existing samples[30,43]. The prediction uncertainty is given as the following equation:

$$U_j = 1 - \max(S_{j1} \times r_1, S_{j2} \times r_2, \ldots, S_{jn'} \times r_{n'}) \quad (4)$$

Equation (4) indicates that the prediction uncertainty at unvisited location *j* is affected by the environmental similarity ($S_{jn'}$) and reliability ($r_{n'}$) at the sample points used in the prediction.

**Study area and datasets.** The study areas included the major rice cultivation zones of China and in total 794 samples were collected in 2017 (Fig.5). A total of 17 provinces and 117 counties were sampled, covering the main rice producing regions in China. The number of samples in each region was determined according to the size of the planting area. The size of each growing area was determined from WESTDC Land Cover Products 2.0[44] using a resolution of 0.15°×0.15°.

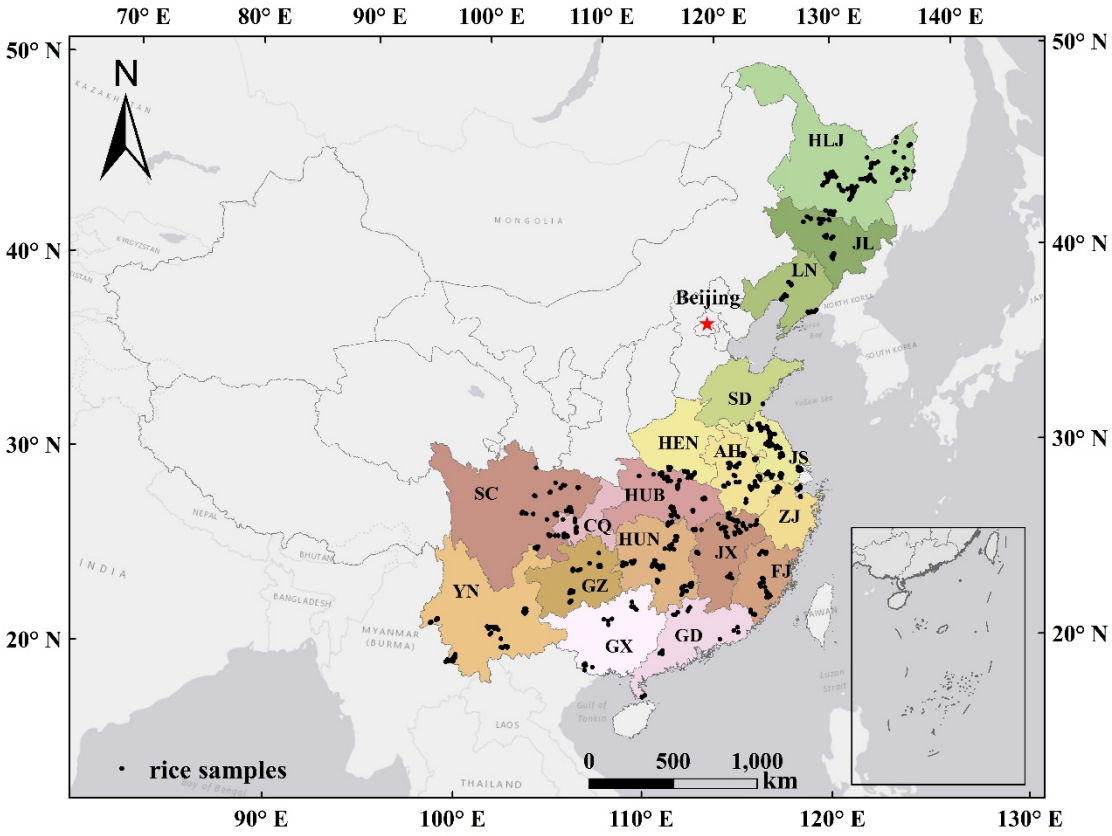



Figure 5. Location sites of 794 rice samples from the study area. Note: The colored regions represent the main rice producing provinces in China.

The isotopic composition data included carbon, hydrogen and oxygen isotope ($\delta^{13}C$, $\delta^{2}H$ and $\delta^{18}O$) values. $\delta^{13}C$ values were analyzed using an Isotope Cube elemental analyzer (Elementar, Germany) coupled with a Biovision Isotope Ratio Mass Spectrometer (Elementar, United Kingdom). $\delta^{2}H$ and $\delta^{18}O$ values were measured using an Isotope PYRO cube elemental analyzer (Elementar, Germany) connected to an Isoprime100 Isotope Ratio Mass Spectrometer (Elementar, United Kingdom).The detailed analysis of these three stable isotopes are referenced Liu et al.[45,46].

The spatial data includes coordinates of the rice sites and climate data. The sample sites were stored as point feature classes in an ArcGIS 10.6 (ESRI, Redlands, Calif. USA) geo-database (Fig.5). The climate data includes average annual temperatures (T), average annual relative humidity (RHU), annual precipitation (PRE), annual sunshine duration (SDD) and growing degree days (GDD), average temperature ($T_g$) and relative humidity ($RHU_g$) precipitation ($PRE_g$), growing degree days ($GDD_g$) and sunshine duration ($SDD_g$) during the growing season (from June to October). The climate data was derived by interpolating gauged daily temperatures, precipitation, relative humidity and sunshine

duration from 825 CMA (the Chinese Meteorological Administration, http://data.cma.cn/) stations in 2017 into a spatial resolution of 0.15°.

**Evaluation and validation.** Repeated subsampling was used for evaluation and validation. All existing samples were randomly divided into training samples (70 %) and validation samples (30 %). In this study, 794 samples were randomly divided into 555 training samples and 239 validation samples. The dataset divisions were repeated 10 times. The training samples were used to develop an isoscape model, and then the validation samples were used to evaluate the prediction model. The mean summary statistics of the 10 repeated subsamples are shown in Table S4. All samples were used for predictive mapping of the stable isotope ratios.

To evaluate and validate the similarity-based method, a combined regression-geostatistics method[10] was developed for comparison. In order to ensure consistency and comparability, the same environmental covariates were used as the predictors in both of these methods.

Four indices were used to evaluate prediction accuracy, including root mean square error (RMSE), mean absolute error (MAE), Pearson's correlation coefficient (r) and coefficient of determination ($R^2$), which are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{l=1}^{n}(\hat{y}_l - y_l)^2}{n}} \quad (5)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{l=1}^{n}(|\hat{y}_l - y_l|) \quad (6)$$

$$\mathrm{r} = \frac{\sum_{l=1}^{n}(y_l - \overline{y_l})(\hat{y}_l - \overline{\hat{y}_l})}{\sqrt{\sum_{l=1}^{n}(y_l - \overline{y_l})^2}\sqrt{\sum_{l=1}^{n}(\hat{y}_l - \overline{\hat{y}_l})^2}} \quad (7)$$

$$\mathrm{R^2} = 1 - \frac{\sum_{l=1}^{n}(y_l - \hat{y}_l)^2}{\sum_{l=1}^{n}(y_l - \bar{y})^2} \quad (8)$$

Where n is the number of the validation points, $y_l$ is the observed isotope ratio value at location *l*, $\hat{y}_l$ is the predicted isotope ratio value at location *l*. $\overline{y_l}$ and $\overline{\hat{y}_l}$ are the means of $y_l$ and $\hat{y}_l$, respectively. If the RMSE and MAE is close to 0, it means that the model performs well. If r and $R^2$ are close to 1, this also suggests that the model has good performance.

**Acknowledgements**

This work was supported by the Self-design Project from State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products [Grant No. 2010DS700124-ZZ1908]; Crop Varietal Improvement and Insect Pests Control by Nuclear Radiation Project, China; IAEA Coordinated Research Project [Grant No. D52042]; Zhejiang Provincial Public Welfare Technology

Research Program [Grant No. LGJ20C200003] and Special Fund of Discipline Construction for Traceability of Agricultural Product (2021-ZAAS). We also acknowledge Dr. Michael J. Friedel for providing feedback and improving this manuscript.

**References**

1. Ariyama, K., Shinozaki, M. & Kawasaki, A. Determination of the geographic origin of rice by hemometrics with strontium and lead isotope ratios and multielement concentrations. J Agric Food Chem 60, 1628-1634, doi:10.1021/jf204296p (2012).
2. Elert,E., Rice by the numbers: A good grain. Nature 514, 50–51.https://doi.org/10.1038/514S50a (2014).
3. Carriger, S., Vallee, D.,. More crop per drop. Rice Today. 6 (2), 10–13 (2007).
4. Lamers, M., et al. Pesticide Pollution in Surface- and Groundwater by Paddy Rice Cultivation: A Case Study from Northern Vietnam. CleanSoil, Air,Water. 39, 356–361. https://doi.org/10.1002/clen.201000268 (2011).
5. Dong, J.W., Xiao, X.M.,. Evolution of regional to global paddy ricemapping methods: A review. ISPRS J. Photogramm. Remote Sens.119, 214–227. https://doi.org/10.1016/j.isprsjprs.2016.05.010 (2016).
6. Wang, J. et al. Tracing the geographical origin of rice by stable isotopic analyses combined with chemometrics. Food Chem 313, 126093, doi:10.1016/j.foodchem.2019.126093 (2020).
7. Yang, P. et al. Evaluation of sample preparation methods for rice geographic origin classification using laser-induced breakdown spectroscopy. Journal of Cereal Science 80, 111-118, doi:10.1016/j.jcs.2018.01.007 (2018).
8. Suzuki, Y., et al. Geographical origin of polished rice based on multiple element and stable isotope analyses. Food Chem 109, 470-475, doi:10.1016/j.foodchem.2007.12.063 (2008).
9. Kelly, S. et al. The application of isotopic and elemental analysis to determine the geographical origin of premium long grain rice. European Food Research and Technology 214, 72-78, doi:10.1007/s002170100400 (2002).
10. Wu, Y. et al. Geographical origin of cereal grains based on element analyser-stable isotope ratio mass spectrometry (EA-SIRMS). Food Chem 174, 553-557, doi:10.1016/j.foodchem.2014.11.096 (2015).
11. Chen, T. et al. Variation of the light stable isotopes in the superior and inferior grains of rice (Oryza sativa L.) with different geographical origins. Food Chem 209, 95-98, doi:10.1016/j.foodchem.2016.04.029 (2016).
12. Chung, I. M., et al. Authenticity of rice (Oryza sativa L.) geographical origin based on analysis of C, N, O and S stable isotope ratios: a preliminary case report in Korea, China and Philippine. J Sci Food Agric 96, 2433-2439, doi:10.1002/jsfa.7363 (2016).
13. Yuan, Y. et al. Differentiating Organically Farmed Rice from Conventional and Green Rice Harvested from an Experimental Field Trial Using Stable Isotopes and Multi-Element Chemometrics. J Agric Food Chem 66, 2607-2615, doi:10.1021/acs.jafc.7b05422 (2018).
14. Chiocchini, F., et al. Isoscapes of carbon and oxygen stable isotope compositions in tracing authenticity and geographical origin of Italian extra-virgin olive oils. Food Chem 202, 291-301, doi:10.1016/j.foodchem.2016.01.146 (2016).

15. West, J. B., et al. Stable isotopes as one of nature's ecological recorders. Trends EcolEvol 21, 408-414, doi:10.1016/j.tree.2006.04.002 (2006).
16. West, J. B., et al. (Eds.). Isoscapes: Understanding movement, pattern, and process on earth through isotope mapping.Springer: New York. 487 pp (2010).
17. Bowen, G. J. Isoscapes: Spatial Pattern in Isotopic Biogeochemistry. Annual Review of Earth and Planetary Sciences 38, 161-187, doi:10.1146/annurev-earth-040809-152429 (2010).
18. Hatvani, I. G., et al. Geostatistical analysis and isoscape of ice core derived water stable isotope records in an Antarctic macro region. Polar Science 13, 23-32, doi:10.1016/j.polar.2017.04.001 (2017).
19. Cheesman, A. W. &Cernusak, L. A. Isoscapes: a new dimension in community ecology. Tree Physiol 36, 1456-1459, doi:10.1093/treephys/tpw099 (2016).
20. Hellmann, C., et al. Isoscapes resolve species-specific spatial patterns in plant-plant interactions in an invaded Mediterranean dune ecosystem. Tree Physiol 36, 1460-1470, doi:10.1093/treephys/tpw075 (2016).
21. Gori, Y., Stradiotti, A. & Camin, F. Timber isoscapes. A case study in a mountain area in the Italian Alps. PLoS One 13, e0192970, doi:10.1371/journal.pone.0192970 (2018).
22. West, J. B., Ehleringer, J. R. &Cerling, T. E. Geography and vintage predicted by a novel GIS model of wine delta O-18. Journal of Agricultural and Food Chemistry 55, 7075-7083, doi:10.1021/jf071211r (2007).
23. van der Veer, G. in New Analytical Approaches for Verifying the Origin of Food 60-80 (2013).
24. Zhu, A. X., et al. Spatial prediction based on Third Law of Geography. Annals of GIS 24, 225-240, doi:10.1080/19475683.2018.1534890 (2018).
25. Cernusak, L. A. et al. Stable isotopes in leaf water of terrestrial plants. Plant Cell Environ 39, 1087-1102, doi:10.1111/pce.12703 (2016).
26. Brugnoli, E., & Farquhar, G. D. Photosynthetic fractionation of carbon isotopes. In R. C. Leegood, T. D. Sharkey, & S. von Caemmerer (Eds.). Photosynthesis: Physiology and metabolism – advances in photosynthesis. The Netherlands: Kluwer Academic Publishers. Vol. 9, pp. 399–434 (2000).
27. Akamatsu, F., Okuda, M. & Fujii, T. Long-term responses to climate change of the carbon and oxygen stable isotopic compositions and gelatinization temperature of rice. Food Chemistry 315, doi:10.1016/j.foodchem.2020.126239 (2020).
28. Dawson, T. E., et al. Stable isotopes in plant ecology. Annual Review of Ecology and Systematics, 33, 507–559. http://dx.doi.org/10.1146/annurev.ecolsys.33.020602.095451 (2002).
29. Camin, F. et al. Climatic and geographical dependence of the H, C and O stable isotope ratios of Italian wine. Anal Chim Acta 853, 384-390, doi:10.1016/j.aca. 2014.09.049 (2015).
30. Xia, W. et al. Stable isotope and photosynthetic response of tea grown under different temperature and light conditions. Food Chem 368, 130771, doi:10.1016/j.foodchem.2021.130771 (2022).
31. Deng, X. et al. Predictive geographical authentication of green tea with protected designation of origin using a random forest model. Food Control 107, doi:10.1016/j.foodcont.2019.106807 (2020).
32. Zhu, A. X. et al. Construction of membership functions for predictive soil mapping under fuzzy

logic. Geoderma 155, 164-174, doi:10.1016/j.geoderma.2009.05.024 (2010).
33. Liu, F. et al. A similarity-based method for three-dimensional prediction of soil organic matter concentration. Geoderma 263, 254-263, doi:10.1016/j.geoderma.2015.05.013 (2016).
34. Zhu, A. X. et al. Predictive soil mapping with limited sample data. European Journal of Soil Science 66, 535-547, doi:10.1111/ejss.12244 (2015).
35. Yang, L., et al. Mapping soil salinity using a similarity-based prediction approach: A case study in Huanghe River Delta, China. Chinese Geographical Science 25, 283-294, doi:10.1007/s11769-015-0740-7 (2015).
36. Zhu, A. X. et al. A similarity-based approach to sampling absence data for landslide susceptibility mapping using data-driven methods. Catena 183, 104188, doi:10.1016/j.catena.2019.104188 (2019).
37. Du, H., Zhu, A. X. & Wang, Y. Spatial prediction of flea index of transmitting plague based on environmental similarity. Annals of GIS 26, 227-236, doi:10.1080/19475683.2020.1788639 (2020).
38. Dawson, T. E., Mambelli, S., Plamboeck, A. H., Templer, P. H. & Tu, K. P. Stable Isotopes in Plant Ecology. Annual Review of Ecology and Systematics 33, 507-559, doi:10.1146/annurev.ecolsys.33.020602.095451 (2002).
39. Ehleringer, J. R., Cerling, T. E., & Dearing, M. D. (Eds.). A history of atmospheric CO2 and its effect on plants, animals, and ecosystems. New York: Springer Verlag (2005).
40. Wang, Z. et al. Linking growing conditions to stable isotope ratios and elemental compositions of Costa Rican bananas (Musa spp.). Food Res Int 129, 108882, doi:10.1016/j.foodres.2019.108882 (2020).
41. 3Gower, J. C. A General Coefficient of Similarity and Some of Its Properties. Biometrics 27, 857-871 (1971).
42. Zhu, A. X. & Band, L. E. A Knowledge-Based Approach to Data Integration for Soil Mapping. Canadian Journal of Remote Sensing 20, 408-418, doi:10.1080/07038992.1994.10874583 (1994).
43. Liu, J., et al. A trustworthiness indicator to select sample points for the individual predictive soil mapping method (iPSM). Geoderma 373, 114440, doi:10.1016/j.geoderma.2020.114440 (2020).
44. Ran, Y. Land cover products of China. National Tibetan Plateau Data Center, DOI: 10.3972/westdc.007.2013.db. CSTR: 18406.11.westdc.007.2013.db (2013).
45. Liu, Z. et al. Assuring food safety and traceability of polished rice from different production regions in China and Southeast Asia using chemometric models. Food Control 99, 1-10, doi:10.1016/j.foodcont.2018.12.011 (2019).
46. Liu, Z. et al. Long-Term Agricultural Effects on the Authentication Accuracy of Organic, Green, and Conventional Rice Using Isotopic and Elemental Chemometric Analyses. J Agric Food Chem 68, 1213-1225, doi:10.1021/acs.jafc.9b06847 (2020).

# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx