Tracking county-level cooking emissions and their drivers in China from 1990 to 2021 with ensemble machine learning

Zeqi Li[1,2], **Bin Zhao**[1,2], **Shengyue Li**[1,2], **Zhezhe Shi**[1,2], **Dejia Yin**[1,2], **Qingru Wu**[1,2], **Fenfen Zhang**[1,2,3], **Xiao Yun**[4,5], **Guanghan Huang**[6], **Yun Zhu**[7], and **Shuxiao Wang**[1,2]

[1]State Key Joint Laboratory of Environmental Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China
[2]State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, Beijing 100084, China
[3]Department of Environment, Yangtze Delta Region Institute of Tsinghua University, Zhejiang, Jiaxing 314006, China
[4]China Energy Longyuan Environmental Protection Co., Ltd., Beijing 100039, China
[5]National Engineering Research Center of New Energy Power Generation, North China Electric Power University, Beijing 102206, China
[6]Beijing Municipal Research Institute of Eco-Environmental Protection, Beijing 100037, China
[7]Guangdong Provincial Key Laboratory of Atmos. Environ. and Pollution Control, College of Environment and Energy, South China University of Technology, Guangzhou 510006, China

Correspondence: Shuxiao Wang (shxwang@tsinghua.edu.cn)

Abstract. Cooking emissions are a significant source of $PM_{2.5}$, posing considerable public health risks due to their high toxicity and proximity to densely populated areas. Despite their importance, there is currently a lack of an accurate, long-term, high-resolution national cooking emission inventory in China, primarily due to the challenges of obtaining high-quality activity-level data over extended periods at fine spatial scales. Here, we address these limitations by leveraging advanced machine learning techniques to predict activity levels and further estimate emissions.

Specifically, we develop an ensemble model of machine learning algorithms – random forest (RF), eXtreme gradient boosting (XGBoost), multilayer perceptron neural network (MLP), and deep neural networks (DNNs) – to accurately predict cooking activity levels across Chinese counties based on statistical indicators related to population, economy, and the catering industry. The ensemble machine learning model demonstrates exceptional generalization and transferability ($R^2 = 0.892$–0.989), outperforming traditional statistical models and individual machine learning models. Unlike previous inventories that rely on simplistic proxy data such as population for calculation and downscaling, our inventory precisely calculates county-level cooking emissions, providing more accurate emission estimates and spatial distributions. Furthermore, we incorporate critical but previously missing toxic pollutants, such as ultrafine particles (UFPs) and polycyclic aromatic hydrocarbons (PAHs), into the national cooking emission inventory. Therefore, we develop China's first county-level cooking emission inventory, spanning 1990 to 2021, with high spatial resolution and wide pollutant coverage.

According to our inventory, in 2021, China's total cooking emissions of organics in the full volatility range, $PM_{2.5}$, UFPs, and PAHs are 997, 408 kt, 6.50×10^{25} particles, and 15.8 kt, respectively. From 1990 to 2021, emissions of these pollutants increased by over 65 %, and their spatiotemporal trends were affected to varying degrees by external factors, such as population migration, economic development, pollution control policies, and the pandemic in different periods. We further analyze the contribution patterns of key driving factors, such

as urbanization rate, population, and pollution control, to emission changes. Notably, driver analysis reveals that existing control measures are insufficient to curb the rapid growth of emissions, necessitating enhanced controls. Regarding control strategies, our county-level inventory finds that 62.3 % of China's organic emissions are concentrated in 30 % of the counties, which are densely populated and occupy only 14.4 % of the national land area. Therefore, prioritizing control of these areas will be an efficient and targeted strategy. Our research provides crucial data and insights for understanding the impact of cooking emissions on air pollution and health, aiding in policy development. Our long-term, high-resolution emission datasets are publicly available at https://doi.org/10.6084/m9.figshare.26085487 (Li et al., 2025).

1 Introduction

Cooking activities, through the heating and processing of oil and food ingredients, emit large quantities of pollutants, posing significant harm to air quality and human health. Cooking emissions are one of the major sources of organic aerosol (OA, the organic component of $PM_{2.5}$) in urban areas (Lee et al., 2015; Logue et al., 2014; Zhao and Zhao, 2018). Source apportionment results indicate that cooking organic aerosols account for 5 %–37 % of the total OA concentration in various urban atmospheres (Abdullahi et al., 2013; Huang et al., 2021; Mohr et al., 2012). Moreover, cooking emissions contain multiple hazardous components, such as ultrafine particles (UFPs) and polycyclic aromatic hydrocarbons (PAHs), which are linked to health problems including cardiovascular disease, oxidative stress, and lung cancer (Kim et al., 2024; Lin et al., 2022b; Naseri et al., 2024; Xu et al., 2020). Experiments have proved that both gaseous organics and $PM_{2.5}$ emitted from cooking exhibit much more negative biological effects like cytotoxicity compared to ambient $PM_{2.5}$ (Guo et al., 2023). Consequently, cooking emissions can increase $PM_{2.5}$ concentrations and toxicity, thereby exacerbating air pollution and associated disease burdens (Chafe et al., 2014; Wang et al., 2017; Zhang et al., 2024a). Given that cooking activities predominantly occur in densely populated areas, they pose a substantial public health risk. Therefore, long-term high-spatial-resolution emission inventories are critical for assessing the impacts of cooking emissions on human health, as they support exposure analysis studies across different locations and periods.

Chinese cooking emissions exhibit unique characteristics that require special consideration in emission inventory development (Zhao and Zhao, 2018). The widespread use of high-temperature oils, various seasonings, and special techniques like stir-frying generates pollutants with complex chemical compositions (Chen et al., 2018; Zhao and Zhao, 2018). Furthermore, cooking styles vary greatly across different regions in China, which have historically developed into multiple distinct cuisine systems (Li et al., 2023; Lin et al., 2022b). To accurately capture these emission patterns and quantify their impact, the emission estimate needs to incorporate cuisine-specific multi-pollutant emission factors while explicitly accounting for regional variations through spatial resolution representation (Lin et al., 2022b; Zhao and Zhao, 2018).

Internationally, some efforts have been made to develop cooking emission inventories. High-resolution emission datasets have been established for small-scale regions, such as greater Athens in Greece and the Red River Delta in Vietnam, through field surveys and measurements (Fameli et al., 2022; Huy et al., 2021). However, at larger scales (e.g., national or global), cooking sources are often omitted from anthropogenic emission inventories or only roughly estimated using uniform emission factors and simplistic statistics like food supply or meat consumption (Huang et al., 2023; Saha et al., 2024). These methods and data are difficult to apply to China because, as mentioned above, cooking inventories in China require localized emission factors and estimation methods that explicitly consider regional differences.

Domestic inventories also exhibit the characteristic of being "precise at small scales but coarse at large scales", making it difficult to balance accuracy and breadth (Cheng et al., 2022; Jin et al., 2021; Liang et al., 2022; Wang et al., 2018a). These limitations are mainly due to the difficulty of obtaining high-quality data, particularly activity-level data, over large spatiotemporal scales and at fine spatial resolutions. Some studies have collected key data for emission calculations by cuisine-specific emission factor testing, door-to-door surveys of restaurants, and online fume monitoring systems, thereby establishing high-resolution inventories of single years in cities or districts such as Beijing, Shanghai, and Shunde (Lin et al., 2022b; Wang et al., 2018b, a; Yuan et al., 2023). These studies have provided valuable localized basic data for China's cooking emission inventories. However, obtaining accurate cooking activity data (e.g., restaurant numbers) remains challenging at larger temporal and spatial scales. Traditionally, China's national cooking emission inventories either use simplistic statistical data (such as population and catering consumption expenditure) as proxies for activity levels or linearly extrapolate the activity levels of one city to other areas based on these simple statistics (Cheng et al., 2022; Jin et al., 2021; Liang et al., 2022; Wang et al., 2018a). These simplifications and linear assumptions result in high uncertainties and low spatial resolution. Recent studies have more accurately estimated national cooking emis-

sions based on data from digital maps or catering service platforms (Li et al., 2023; Zhang et al., 2024b). However, these inventories are limited to recent years, as they rely on newly developed data platforms.

Apart from lacking accuracy and breadth, another limitation of existing cooking emission inventories is their limited pollutant coverage. Previous studies on cooking emissions primarily focused on $PM_{2.5}$ (whose organic component is primary organic aerosol, POA) and volatile organic compounds (VOCs) (Jin et al., 2021; Wang et al., 2018a, b). However, recent advancements in the framework of organic compounds in the full volatility range (including VOCs, intermediate-volatility organic compounds (IVOCs), semi-volatile organic compounds (SVOCs), and organic compounds with even lower volatility (xLVOCs)) have revealed the previously overlooked significant contributions of I/SVOCs to secondary organic aerosols (SOAs) (Chang et al., 2022; Li et al., 2024b; Zhang et al., 2021). Although our latest study has supplemented the inventory with organics in the full volatility range (Li et al., 2023), the emissions for certain highly toxic pollutants of particular concern emitted from cooking – notably ultrafine particles (UFPs) and polycyclic aromatic hydrocarbons (PAHs) – remain lacking (Chen and Zhao, 2024; Jørgensen et al., 2013; Lachowicz et al., 2023; Lin et al., 2022a). This gap limits our comprehensive assessment of the environmental and health risks associated with cooking emissions.

In recent years, machine learning has been widely applied in atmospheric pollution research due to its powerful capability to process large-scale spatiotemporal datasets and capture complex nonlinear relationships within them (Liu et al., 2023; Prodhan et al., 2022a; Zhang and Zhao, 2024; Zheng et al., 2021). Models such as random forest (RF), eXtreme gradient boosting (XGBoost), and deep neural networks (DNNs) have demonstrated strong performance in predicting pollutant concentration time series and identifying spatial distributions (Chen et al., 2024; Prodhan et al., 2022b; Ren et al., 2022; Wu et al., 2024; Xu et al., 2023). Ensemble machine learning models further achieve better and more stable results by combining predictions from individual base models (Liu et al., 2023; Ren et al., 2022). They can help supplement sparse datasets, serving as an effective alternative for obtaining key data that would otherwise be computationally expensive or inaccessible to collect (Ren et al., 2022; Shi et al., 2024; Xiao et al., 2018). When integrating machine learning with the SHapley Additive exPlanations (SHAP) additivity algorithm, the key factors for the predictive target and their influence patterns can be identified (Hou et al., 2022; Yang et al., 2023a). Most importantly, these approaches hold significant potential to address key challenges in cooking emission inventory development. Where conventional activity data at large spatiotemporal scales are unavailable, ensemble machine learning models can predict long-term, high-resolution activity levels by capturing complex relationships between cooking activities and fundamental socioeconomic indicators. Coupled with SHAP analysis, they can provide insights into how socioeconomic factors influence emission trends. However, such efforts have not yet been made.

In summary, limited by the difficulty of obtaining high-quality activity data, there is currently a lack of an accurate, long-term, high-resolution national cooking emission, and existing inventories remain deficient in their coverage of important toxic pollutants such as PAHs and UFPs. This hinders studies on $PM_{2.5}$ modeling, source apportionment, and health risk analysis. In this study, we use machine learning models to overcome the limitations of data acquisition and driving force analysis, while also expanding the range of pollutants covered in the emission inventory. Specifically, we employ an ensemble of four preferred machine learning algorithms to estimate long-term, high-spatial-resolution cooking activity data. This ensemble model integrates the strengths of the four base models – RF, XGBoost, multilayer perceptron neural network (MLP), and DNN – enabling it to accurately predict cooking activity levels across various Chinese counties based on statistical indicators related to population, the economy, and the catering industry. We validate the model's generalizability and transferability using unseen testing datasets. By further combining advanced emission factors and pollution control data, we estimate the emissions of various pollutants (including organics in the full volatility range, $PM_{2.5}$, UFPs, and PAHs) from commercial, residential, and canteen cooking at the county level from 1990 to 2021. Finally, we apply the one-factor-at-a-time method to analyze the key drivers of national cooking emissions, while using the SHAP algorithm to identify the key influential factors for county-level emissions. This provides essential data and new insights for studies of the impact of cooking emissions on air pollution and human health and helps to formulate targeted emission control policies.

We expect to achieve breakthroughs in spatiotemporal resolution, pollutant coverage, and emission estimation accuracy. This study represents the first long-term (nearly 31 years) high-resolution (county-level) inventory, whereas existing national inventories were mostly limited to single years or recent years at provincial resolution. Also, our study covers key pollutant categories from cooking emissions, including organics in the full volatility range, PAHs, and UFPs, that were not included in other national inventories. In terms of estimation accuracy, we adopt cuisine-specific emission factors, consider dynamically changing purification facility installation proportions (PFIPs) driven by provincial policies, and use precise county-level activity data to calculate emissions more accurately and better reflect regional differences. Finally, this study will provide important data and new perspectives for researching the impacts of cooking emissions on air pollution and human health, facilitating the development of targeted emission control policies.

2 Data and method

The calculation method for emissions of the three sectors of cooking (commercial cooking, residential cooking, and canteen cooking) is based on Li et al. (2023), as shown in Eq. (1):

$$E = A \times [\text{EF} \times y + \text{EF}'(1 - y)], \tag{1}$$

where A represents the activity level, EF and EF$'$ are the controlled and uncontrolled EFs for a certain pollutant, and y is the PFIP.

Figure 1 illustrates the workflow of activity-level modeling, emission estimation, and driver analysis in this study. We first gather historical annual statistical data related to population, the economy, and the catering industry as predictive variables, and collect existing high-resolution cooking activity levels as response variables. All data are standardized to the resolution of county level, ensuring that the sample set used for modeling is rich, diverse, and of high spatial resolution. Then, we integrate four machine learning algorithms – RF, XGBoost, MLP, and DNN – which are selected for their superior predictive performance and complementary strengths, to develop predictive models for cooking activity levels across three sectors: commercial, residential, and canteen cooking. The reliability of the model is validated on unseen testing datasets. The activity levels predicted by the model, combined with emission factors and the PFIPs, can yield historical county-level cooking emissions. Finally, we apply the one-factor-at-a-time method to analyze the key drivers of national cooking emissions, whereas the SHAP algorithm is used to evaluate the relative importance of features for county-level emissions.

2.1 Data acquisition and processing

To obtain long-term, high-resolution national emissions, it is important to acquire nationwide activity-level data that span extended periods and maintain fine spatial resolution (such as county level or at least municipal level). However, this is a highly challenging task, especially before the year 2000, when a significant amount of data was missing. Fortunately, we can leverage the powerful data imputation and predictive capabilities of machine learning to overcome this challenge. Specifically, the activity levels for commercial, residential, and canteen cooking are the annual total fume volume, annual total household edible oil consumption, and the annual total number of meals served in canteens, respectively. We develop predictive models based on machine learning algorithms that only use easily accessible statistical data to estimate these county-level activity levels (as discussed in Sect. 2.2).

We collected 14 statistical indicators related to population, the economy, and the catering industry from 1990 to 2021 for modeling and predicting. The types, sources and initial resolution of all statistical data can be accessed in Table S1. Population-related variables include population, the number of employees in enterprises, and the number of students in primary school and middle school. Economy-related variables encompass urbanization rate, total gross domestic product (GDP), GDP of primary, secondary, and tertiary industries, and per capita disposable income. Variables related to the catering industry include household per capita oil consumption, household per capita meat consumption, the number of chain restaurants, and the number of employees in the catering and accommodation industry. These data, mostly at the county-level resolution, primarily originate from statistical yearbooks (National Bureau of Statistics of China, 2022a, c, b). These long-term datasets are preprocessed to meet the requirements of machine learning by imputing missing values using inverse distance weighting, K-nearest neighbor methods, and higher-order statistical data (Murti et al., 2019; Sree Dhevi, 2014). Given the changes in China's county administrative divisions over the past 31 years (Yu et al., 2018), we trace the renaming, merging, and splitting events of counties according to official government reports, mapping the data for each year to the county administrative system of 2020 (a total of 2848 counties) to ensure continuity across years. Detailed descriptions of the spatial mapping and data processing procedures are provided in Text S1 in the Supplement. Additionally, we standardize the initial resolution of some variables, which may be at the provincial, municipal, or grid level (1 km × 1 km), to the county level by allocating based on population or GDP, taking provincial averages, or using cumulative summation. We also normalized all predictor variables to a range of 0 to 1 to ensure a consistent scale.

Next, we conduct feature selection on 14 predictor variables to reduce dimensionality and minimize multicollinearity (Zhu et al., 2023). Variables were first ranked by RF model importance scores, with those scoring < 5 % considered for potential removal because of low importance (Alduailij et al., 2022; Ye et al., 2022). Each candidate variable was then temporarily excluded, and only those whose removal resulted in an R^2 decline of < 1 % were permanently discarded – ensuring no critical features were lost (Altmann et al., 2010; Zhu et al., 2023). Also, we performed multicollinearity checks using the variance inflation factor (VIF), gradually removing features with higher VIF values until all remaining features were mutually independent (all VIF values of independent variables were below 10) (Daoud, 2017; Hu et al., 2017). By removing irrelevant or redundant features in this way, we can reduce the influence of noise, decrease the risk of overfitting, enhance the model's predictive performance and generalizability, and provide clearer and more meaningful model explanations (Zhu et al., 2023).

For machine learning modeling, the dataset needs to be partitioned into the training dataset and the testing dataset. During the data partitioning, we implement strict data leakage management to ensure that information from the testing dataset is not used during training, thus guaranteeing an accurate model evaluation (Nayak and Ojha, 2020; Zhu et al., 2023). The response variables available for modeling and
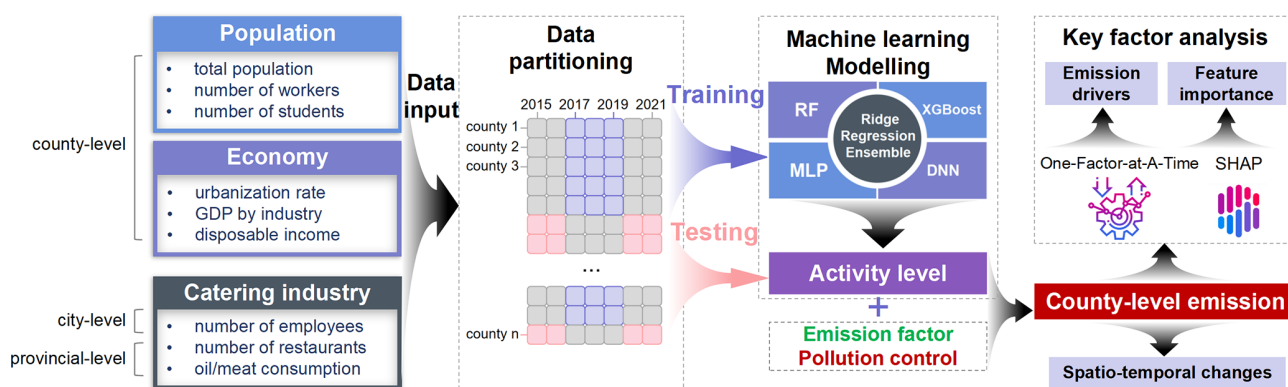
Earth Syst. Sci. Data, 17, 5113–5135, 2025

https://doi.org/10.5194/essd-17-5113-2025

Figure 1. Schematics of the model developed in this study including model development, emission calculation, and key factor analysis.

testing, namely high-resolution cooking activity levels, are limited to the years 2015 to 2021 (Li et al., 2023). This gives us data samples for 7 years, with 2848 counties each year. Given the significant similarity in data for the same county across different years, we bundle data samples from different years for the same county during the data partitioning. As shown in the second column of Fig. 1, we use data from 70 % of the counties from 2017 to 2019 (totaling 5982 samples) for training to establish the underlying relationships between input factors and the prediction target. Additionally, we use data from the remaining 30 % of the counties in the years 2015, 2016, 2020, and 2021 (totaling 3416 samples) as the testing dataset to validate the model. Under this partitioning strategy, data from the same county only appear in either the training dataset or the testing dataset, ensuring that the model can effectively generalize and be tested on unseen datasets, thereby demonstrating the model's transferability across different times and locations (Nayak and Ojha, 2020; Zhu et al., 2023). Notably, the modeling data only cover the period from 2015 to 2021, as earlier data are unavailable. This may introduce some bias when backcasting activity levels for earlier periods. We further validate the backcasted activity levels and evaluate their uncertainties. Modeling and validation of the machine learning model are described in detail in Sect. 2.2.

After obtaining activity levels through the machine learning model, we further collect data for Eq. (1) to calculate cooking emissions. As for the EF, we consider various types of pollutants of concern emitted from cooking activities, including organics in the full volatility range (VOCs, SVOCs, IVOCs, and xLVOCs), $PM_{2.5}$, UFP, and PAHs (encompassing gaseous PAHs, particulate PAHs, and benzo[a]pyrene toxic equivalent quantity (BaP_{eq})). The organic EFs in the full volatility range are sourced from Li et al. (2023). The EFs for $PM_{2.5}$ are calculated as POA / 81.5 % (Li et al., 2023), where POA represents the particulate fraction of organics in the full volatility range. The EFs for UFPs are derived from the literature (Chen et al., 2017, 2018; Géhin et al., 2008; Kim et al., 2024; Zhang et al., 2010). The EFs of gaseous and

particulate PAHs are mainly sourced from simultaneous gas–particle testing in multiple studies (Chen et al., 2007; Feng et al., 2021; Li et al., 2003, 2018; Lin et al., 2022a; Saito et al., 2014; Ye et al., 2013). We consider 16 priority PAHs and 5 non-priority PAHs commonly found in cooking emissions. Their BaP_{eq} values are calculated based on the recommended toxic equivalency factors (TEFs) suggested in the literature to estimate the carcinogenic toxicity of PAH emissions (Larsen and Larsen, 1998; Malcolm et al., 1994; Nisbet and LaGoy, 1992). The molecular information and recommended TEF values for all PAH species considered in this study are listed in Table S2. The specific values and sources of EFs for various pollutants are listed in Table S3.

As for PFIPs, we have established a grading standard for provincial catering emission control stringency and corresponding PFIPs based on field surveys (Li et al., 2023). By collecting provincial-level catering pollution control policies and considering their implementation timelines and transition periods, we can obtain dynamically changing PFIPs driven by provincial-level control policies. In this study, we also applied this method to estimate PFIPs for 1990–2021 (see Tables S4–S5 for the PFIP results over the years).

2.2 Establishment and optimization of ensemble machine learning model

Ensemble methods of machine learning have recently been increasingly applied in the large-scale spatiotemporal estimation of atmospheric pollution (Yang et al., 2023b; Zhu et al., 2022). These methods enhance prediction accuracy and robustness by combining the forecast results from multiple base models and reducing the risk of overfitting. In this study, we establish an ensemble prediction model for cooking activity levels by integrating four machine learning algorithms – RF, XGBoost, MLP, and DNN. These four models are selected because they exhibit superior performance in predicting activity levels (as discussed in Sect. 3.1), and each of them possesses unique strengths, as discussed below.

RF and XGBoost are both ensemble learning algorithms based on decision trees. RF improves accuracy and generalization by combining multiple independent decision trees, making it suitable for handling high-dimensional data (Liu et al., 2023; Segal, 2004). Its advantage lies in the effective reduction of overfitting through random feature selection (Wu et al., 2024). XGBoost, as an efficient gradient-boosting decision tree method, also reduces overfitting by introducing regularization and has a high execution speed, making it suitable for processing large-scale datasets (Chen and Guestrin, 2016). While these tree-based algorithms provide stable predictions and good interpretability, they may have limited extrapolation capabilities (Wang et al., 2023). To address this, we introduce MLP and DNN, two deep learning algorithms, to enhance the model's applicability. MLP, a fundamental deep learning model with a multilayer structure, can capture complex nonlinear trends in data and can infer patterns beyond the training data range, with lower computational requirements compared to other deep learning models (Pinkus, 1999). DNN, on the other hand, captures advanced abstract features in complex data through deeper network structures, offering powerful feature learning and generalization capabilities (Zhang et al., 2016). However, both MLP and DNN may face the challenge of overfitting (Pinkus, 1999; Zhang et al., 2016), which can be mitigated by integrating them with RF and XGBoost.

To combine the advantages of these four models, we use ridge regression as the integrator to build an ensemble machine learning model (McDonald, 2009). Ridge regression is chosen for its ability to balance model complexity and generalization through regularization, which helps prevent overfitting (Ebrahimi et al., 2024; McDonald, 2009). Furthermore, as validated in Text S2 and Table S6, ridge regression demonstrates a favorable balance between performance and computational efficiency when compared to other fusion strategies. Specifically, the predictions from the base models serve as new features to input into the ridge regression model, which then determines how to effectively combine these predictions (Carneiro et al., 2022). This approach allows us to leverage the strengths of each model: the interpretability and stability of RF and XGBoost and the ability of MLP and DNN to capture complex nonlinear patterns. By integrating these models, we aim to achieve a more robust and accurate prediction model that can handle diverse data scenarios (Carneiro et al., 2022).

Due to variations in influencing factors and mechanisms within different cooking emission sectors, we develop an ensemble model for commercial, residential, and canteen cooking, respectively. For each sector's training dataset, models are trained using 10-fold cross-validation to ensure that their predictive capabilities are not influenced by specific data subsets (Santos et al., 2018). Moreover, a grid search is conducted on the hyperparameters of each base machine learning model and the ridge regression model to identify the optimal hyperparameter combination that maximizes overall pre-

dictive performance (Belete and Huchaiah, 2022; Lou et al., 2024).

2.3 Model validation and comparison

After completing the modeling, we apply the models to the unseen testing datasets and evaluate their predictive performance using various statistical metrics. The validation metrics include the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Their calculation formulas are as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\mathrm{Obs}_i - \mathrm{Pred}_i)^2}{\sum_{i=1}^{n}(\mathrm{Obs}_i - \mathrm{MeanObs})^2} \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\mathrm{Obs}_i - \mathrm{Pred}_i)^2}{n}} \tag{3}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\mathrm{Pred}_i - \mathrm{Obs}_i| \tag{4}$$

Obs_i represents the actual values (i.e., the activity levels obtained from accurate calculations), Pred_i refers to the model-predicted activity levels, MeanObs is the average of all Obs_i, and n is the number of samples in the testing datasets.

To demonstrate the superiority of our ensemble model, we also compare its predictive performance with the abovementioned four individual machine learning models and five advanced traditional statistical models, including multiple linear regression, non-negative least-squares regression, generalized linear models with exponential link, Poisson regression, and power function regression (Frome, 1983; Jansson, 1985; Myers and Montgomery, 1997; Slawski and Hein, 2013; Uyanık and Güler, 2013).

2.4 National driver analysis and county-level feature importance analysis of cooking emissions

Based on the model-predicted cooking activity levels, the EFs applicable at all times, and PFIPs that can be extrapolated to any time, we can theoretically estimate the cooking emissions in various scenarios (such as different population conditions, economic circumstances, and pollution control intensities). In this study, we first obtain the emissions of three cooking sectors in each county from 1990 to 2021. Further, we can conduct sensitivity analysis of emissions by adjusting various influencing factors (input features of the ensemble model and PFIPs). Since EFs are static data that do not change across different years, we do not consider their impact. We first pay attention to the national total emissions using the one-factor-at-a-time method (Zhang et al., 2018) to illustrate the sensitivity of each factor to emission variations. We divide the years from 1990 to 2021 into several periods. For a given period, we sequentially adjust the value of a single factor from the initial value at the beginning of the period to the final value at the end of the period. The difference between the emissions before and after the adjustment is con-

sidered to be the contribution of that factor to the change in emissions during that period. This enables us to quantify the contributions of each factor to emissions in different periods.

Notably, due to interaction effects among variables, different adjustment sequences may yield distinct driver decomposition results. To enhance the robustness, we mitigate sequence-dependent bias by incorporating the averaged results from multiple adjustment sequences. Specifically, we first identify the top five variables with the largest individual impacts on emission changes in each period. We then exhaustively examine all possible permutations (120 in total) of these five variables and apply the adjustments accordingly. For variables ranked sixth and below, adjustments are made in descending order of their individual impacts, without considering further permutations, as these variables contribute minimally to emission changes ($< 2\%$ annually), and exhaustive permutations would be computationally prohibitive. Finally, we calculate the average contribution of each factor across all considered sequences. While this improved approach cannot completely decouple the independent contributions of all variables, it significantly reduces order-dependent biases and enhances the reliability of the driver decomposition.

The dominant factors associated with emission changes for counties at different development stages are also worth elucidating, which are crucial for understanding the current and future trends in cooking emissions and for the targeted development of control strategies. We employ the SHAP algorithm (Lundberg and Lee, 2017) to quantify the impact of each factor on cooking emissions in different counties. These factors include features related to population, the economy, and the catering industry that are input to the activity level prediction model, as well as the EF and the PFIP. The SHAP algorithm is based on cooperative game theory (Jiménez-Luna et al., 2020; Lundberg and Lee, 2017). By including or excluding a variable from all possible subsets of the remaining variables, the model is retrained to calculate the difference in predicted values in two scenarios, referred to as SHAP values. The magnitude of SHAP values quantifies the specific contribution of each feature to the model's predictions. A positive value indicates that the feature raises the predicted result relative to the baseline, while a negative value signifies a reduction in the predicted result (Hou et al., 2022; Zhu et al., 2023).

3 Results

3.1 Performance comparison of the models

We first train five traditional statistical models, four individual machine learning models, and our ensemble model using the training dataset. The performance of each model on the training dataset is shown in Table S7. Then, all models are applied to an unseen testing dataset (detailed dataset partitioning is described in Sect. 2.1) to assess their performance

in predicting the activity levels of three cooking sectors. The predictive performance of all models for activity levels of three cooking sectors on the testing dataset is shown in Table 1 and Figs. S1–S3 in the Supplement. To enhance clarity, we scale the units for the three activity levels: the activity levels for commercial, residential, and canteen cooking are respectively represented as annual total fume volume (unit: $10^9\,\mathrm{m}^3$ fume), annual total household edible oil consumption (unit: kt oil), and annual total number of meals served in canteens (10^6 meals). We also present the predictive performance of the best statistical models, the best individual machine learning models, and the ensemble machine learning model for the three cooking sectors in Fig. 2a.

According to Table 1, validation metrics indicate that machine learning models greatly outperform the best traditional statistical models, with the ensemble machine learning model even surpassing the best individual machine learning models. Among statistical models, multiple linear regression has moderate performance, but it is prone to predicting negative values, which do not correspond to real-world cooking activity. Among the other four non-negative predictive models, power function regression performs best for predicting commercial cooking and residential cooking, while nonnegative least-squares regression performs best for predicting canteen cooking. Generalized linear models with exponential links and Poisson regression perform poorly in most cases.

Machine learning models tend to have better predictive capabilities than the traditional statistical models. Among the five machine learning models, we find that ensemble machine learning models consistently perform the best, with R^2 values of 0.892, 0.989, and 0.973 for commercial cooking, residential cooking, and canteen cooking activity levels, respectively. RMSE and MAE metrics of the ensemble models are also relatively low. The superiority of validation metrics implies that the ensemble model can effectively depict the relationship between indicators related to statistic indicators and cooking activity levels. Moreover, the overall performance of individual machine learning models is also satisfactory. Specifically, for commercial cooking and canteen cooking, which are influenced by complex factors, the performance of the two deep learning models is superior, as they are more adept at capturing complex nonlinear relationships. On the other hand, for residential cooking, whose influencing factors are relatively simple and clear, the performance of RF is better than that of deep learning models, possibly because it can effectively prevent overfitting. Finally, the ensemble models can exploit complementary advantages, reduce the uncertainties of single models, and achieve performance maximization.

We also review the predictive performance of all models for the Chinese total activity levels of the three cooking sectors for each year from 2015 to 2021, as shown in Fig. 2b. Although the training dataset was randomly sampled from counties only from 2017 to 2019, the machine learning models (represented by the solid line) demonstrate a robust

Table 1. The values of validation metrics of all models for activity levels of three cooking sectors in the testing dataset.

Model	Commercial cooking			Residential cooking			Canteen cooking		
	R^2	RMSE (10^9 m^3)	MAE (10^9 m^3)	R^2	RMSE (kt)	MAE (kt)	R^2	RMSE (10^6 meals)	MAE (10^6 meals)
Multiple linear regression	0.718	25.618	16.199	0.936	1.078	0.625	0.955	5.697	3.202
Non-negative least-squares regression	0.617	29.857	18.910	0.898	1.368	0.953	0.955	5.750	3.172
Generalized linear models with exponential link	0.625	29.548	17.777	0.348	3.455	2.649	0.496	19.157	14.666
Poisson regression	0.454	35.673	19.737	0.056	4.156	2.947	0.278	22.940	16.294
Power function regression	0.772	23.055	10.599	0.965	0.804	0.406	0.950	6.044	3.085
RF	0.835	19.589	10.173	0.979	0.618	0.155	0.958	5.545	3.109
XGBoost	0.807	21.224	11.582	0.971	0.726	0.277	0.958	5.561	3.037
MLP	0.856	18.316	9.867	0.972	0.714	0.185	0.970	4.675	1.750
DNN	0.866	17.644	8.355	0.970	0.738	0.231	0.969	4.764	2.360
Ensemble machine learning model	0.892	15.834	7.968	0.989	0.455	0.109	0.973	4.447	1.832

ability for generalization and extrapolation. They accurately capture the Chinese total activity level trends of the modeling years (2017–2019) and extend to historical years (2015–2016) and future years (2020–2021), whereas traditional statistical models (represented by the dashed lines) often fail to accurately reproduce the changes in total activity levels.

While the model demonstrates good performance for near-term extrapolation, greater uncertainty may exist when back-casting to earlier periods, which include more underdeveloped counties. To evaluate this, we conduct sensitivity analyses by training on the top 70 % GDP-ranked counties and testing on the bottom 30 %. For commercial catering (the most complex case, as shown in Fig. S4), the ensemble models' test-set R^2 remained robust ($R^2 = 0.719$), outperforming the best statistical models ($R^2 = 0.523$). For real-world backcasting of historical data, the 2015–2021 training data already include some less-developed regions that can represent early-stage conditions, mitigating extreme extrapolation risks. Additionally, we further validate the historical trends of predicted activity data based on the limited available historical data (Fig. S5). From 1990 to 2021, the growth rate of commercial cooking activity levels was intermediate between population growth and tertiary GDP. The temporal evolution resembles that of the chain restaurant number (slow early growth followed by acceleration), though the chain restaurant number is more stable as it excludes small independent restaurants. We also incorporate chain restaurant revenue data (available since 2004), which corroborate the fluctuations in our predictions including the post-2015 rapid growth driven by food delivery platforms and the 2020 pandemic-driven decline (Maimaiti et al., 2018; Zhao et al., 2021). Therefore, while temporal extrapolation may introduce biases (uncertainties are quantified in Sect. 3.3), our multi-pronged validation demonstrates the reasonableness and optimality of our backcasted estimates.

3.2 Long-term county-level cooking emissions

After verifying the reliability and superiority of the ensemble model, we utilize it to predict precise county-level activity data at a broad spatial and temporal scale and further obtain county-level cooking emissions in China from 1990 to 2021. Each county's annual emissions inventory for organics (hereafter representing the organic compounds in the full volatile range), PM$_{2.5}$, UFPs, and PAHs is available at the repository (https://doi.org/10.6084/m9.figshare.26085487) (Li et al., 2025). Considering that organics have significant impacts on atmospheric pollution, particularly OA pollution, and that various pollutants share the same activity levels, leading to similar spatial and temporal distributions, we primarily focus on organic compounds in the following discussion.

Figure 3 provides high-resolution spatial distribution maps of cooking organic emissions in China from 1990 to 2021, while Fig. S6 provides the sector-specific spatial distributions of cooking organic emissions for a representative year (2021). We also provide a map of the Chinese provinces (Fig. S7) for reference to the location of the emissions mentioned below. In 1990, cooking organic emissions were mainly distributed in densely populated areas such as the North China Plain (including Beijing, Tianjin, Hebei, Henan, and Shandong), the middle–lower Yangtze Plain (including Hubei, Hunan, Anhui, Jiangxi, Jiangsu, and Zhejiang), and Sichuan Basin (including Sichuan and Chongqing). Also, emission hotspots were often observed in the core urban areas of provincial capitals. Over time, the national total organic emissions have generally increased, and high-emission areas have expanded. By 2021, many counties in eastern China, especially along the southeast coast, exhibited extensive high emissions. The North China Plain region, the Yangtze River Delta, the Pearl River Delta, and the Sichuan–Chongqin region became the four key emission zones, contributing 20.2 %, 19.9 %, 8.63 %, and 7.98 %, respectively, of the nation's total cooking organic emissions.
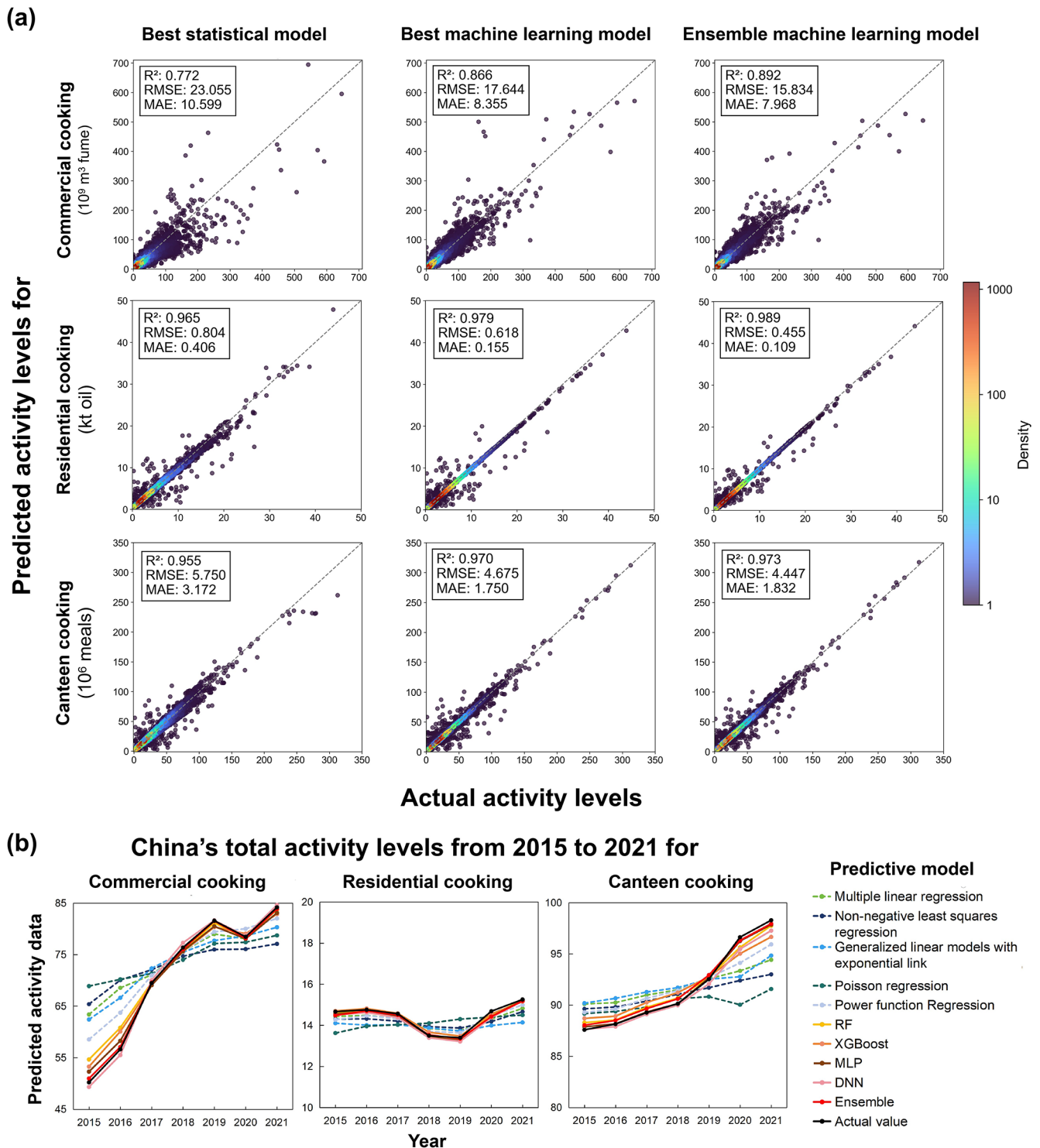
Earth Syst. Sci. Data, 17, 5113–5135, 2025

https://doi.org/10.5194/essd-17-5113-2025

(a)

Figure 2. Comparison of statistical models, individual machine learning models, and ensemble machine learning models: **(a)** scatter plots comparing the actual and predicted values of the best statistical model, best individual machine learning model, and ensemble machine learning model for the activity levels in three cooking source sectors in China, with each point representing the activity level in a county from the testing dataset. **(b)** Predicted and actual values of Chinese total activity levels of the three cooking sectors for each year from 2015 to 2021. The black line represents the actual values. Lines of other colors represent model predictions, where the solid lines are for machine learning model predictions, and dashed lines are for predictions from traditional statistical models.
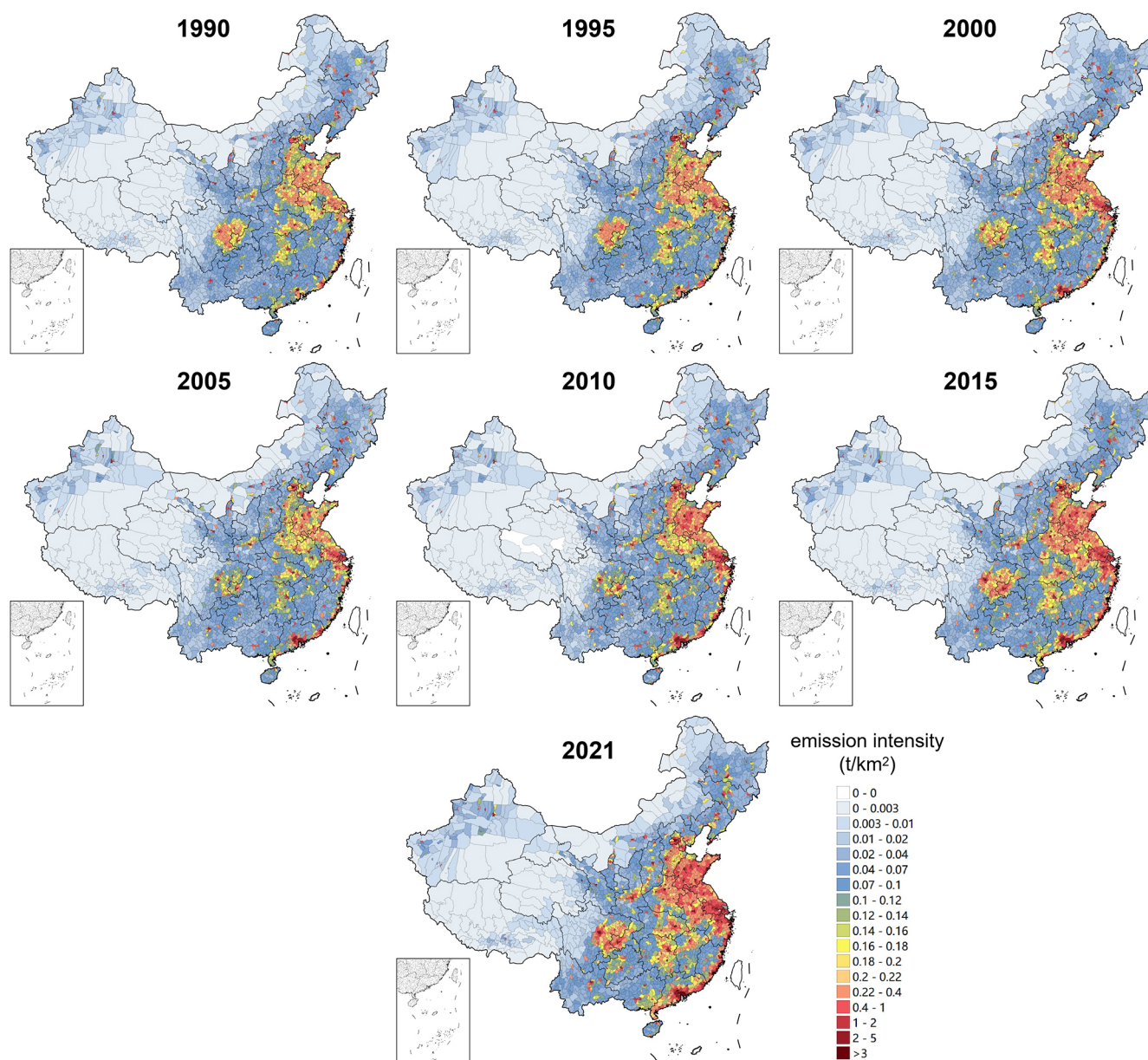
Figure 3. The spatial distribution of nationwide county-level cooking organic emission intensity from 1990 to 2021.

In summary, cooking emissions are concentrated in densely populated and economically developed areas. For example, in 2021, there was a strong correlation between county population size and cooking emissions, with an R^2 of 0.873 for the emissions and population of 2848 counties. Notably, the top 30 % of counties by population (as shown in Fig. S8a) accounted for 62.3 % of the total national cooking organic emissions. These counties cover only 14.5 % of China's total land area but support 59.9 % of the country's population. This finding indicates that, when formulating control strategies, these densely populated counties should be prioritized to enhance pollution control efficiency and effectively reduce the health risks associated with cook-

ing emissions. From 1990 to 2021, the proportion of total national emissions contributed by the top 30 % of counties by population increased from 49.6 % to 62.3 %, suggesting that cooking emissions in densely populated counties have grown faster than in other areas, necessitating stricter pollution control measures. Additionally, cooking emissions are also correlated with local GDP, although this correlation is weaker than with population, with an R^2 of 0.563 for emissions and GDP across all counties in 2021. The top 30 % of counties by GDP (as shown in Fig. S8b) accounted for 55.9 % of the total national cooking organic emissions. We further analyze the spatiotemporal trends of emission distribution and their underlying socioeconomic drivers in Sect. 4.1.

Fortunately, in these densely populated and economically developed areas (Fig. S8), where emissions are typically high, our county-level emissions inventory achieves a very high spatial resolution. Compared to traditional provincial inventories, our fine-grained inventory is more capable of accurately studying the impact of cooking emissions on air pollution and human health. Specifically, our inventory may update the understanding of $PM_{2.5}$ sources. Combining the full-volatility organic emissions inventory (excluding the cooking source) developed by Zheng et al. (2023), we find that cooking emissions are significant sources of I/SVOC emissions in densely populated counties. In 2019, for counties within the top 30 % of population density (as shown in Fig. S8c), cooking emissions can account for an average of 20.1 % of IVOCs and 38.5 % of SVOCs emitted from all anthropogenic sources, and the maximum contribution of cooking emissions to total IVOC and SVOC emissions in these counties even reached 52.9 % and 88.4 %, respectively. Given the high formation potential for SOA of I/SVOCs emitted from cooking (Yu et al., 2022), the contribution of cooking organic emissions to $PM_{2.5}$ and their hazards for human health could be substantial. However, if considering only national or provincial emissions, the contributions of cooking emissions to the total IVOC emissions and total SVOC emissions are both less than 16 %, potentially leading to an underestimation of the importance of the cooking source.

3.3 Trends of national total cooking emissions

Figure 4 illustrates the long-term trend of national cooking emissions of organic compounds in the full volatility range from 1990 to 2021. The total cooking organic emissions in China exhibit an overall increasing trend, rising from 517 (272–828, 95 % confidence level) kt yr^{-1} in 1990 to 997 (530–1590) kt in 2021. The uncertainty ranges are determined through Monte Carlo simulations referencing previous studies (Chang et al., 2022; Li, 2017), incorporating cumulative biases introduced by extrapolating historical emissions using limited training data (see Text S3 for details). Notably, there were slight decreases in total organic emissions after 2001 and after 2013, attributed to the implementation of crucial control policies. In 2001, the issuance of the Emission Standards of Catering Oil Fume (GB 18483-2001) (State Environmental Protection Administration of China, 2001) marked the first significant attention of the Chinese government to cooking emission control. It imposes requirements on the concentration of oily fumes emitted by restaurants and the removal efficiency of the purification facilities, which has contributed to the reduction of emissions (State Environmental Protection Administration of China, 2001). Furthermore, the release of the Action Plan for the Prevention and Control of Air Pollutants in 2013 pushed provinces to comprehensively strengthen air pollution control (CPGPRC, 2013), leading to a corresponding enhancement of the catering industry's regulation in many regions. Additionally, the down-

turn observed in the 2020 emission was brought about by the lockdown measures implemented due to the COVID-19 pandemic.

As for source apportionment, the cooking organic emissions mainly come from commercial cooking and residential cooking. Commercial cooking emissions show an overall upward trend, with some slight fluctuations due to high sensitivity to external factors such as pollution control policies and epidemic lockdowns. Commercial cooking emissions increased from 241 kt in 1990 to 622 kt in 2021, and its share has correspondingly increased from 46.7 % to 62.3 %. Residential cooking emissions show an overall slow upward trend, with its share ranging between 28.3 % and 37.2 %. In contrast, canteen cooking emissions show an overall stable or slightly declining trend. This is possible because they mainly come from staff and student canteens, where the number of staff and students and their meal frequencies are relatively stable. However, with pollution control measures becoming stricter, this has led to a reduction in total canteen cooking emissions.

Furthermore, we also present emissions of $PM_{2.5}$, UFPs, and PAHs (including gaseous PAHs, particulate PAHs, and BaP_{eq}) from the three cooking sectors in China from 1990 to 2021, as shown in Fig. S9. The trends and source apportionment of $PM_{2.5}$ emissions are similar to those of organic emissions. The total $PM_{2.5}$ emissions increased from 215 kt in 1990 to 408 kt in 2021, representing a growth of 90.7 %. Commercial cooking is the most significant emission source, accounting for 39.3 %–57.7 %, followed by residential cooking (34.8 %–44.3 %). The total UFP emissions increased from 3.93×10^{25} particles in 1990 to 6.50×10^{25} particles in 2021, with an increase of 66.0 %. Commercial emissions have consistently been the largest source, maintaining a share of over 71 %.

The total PAH emissions increased from 6.76 kt in 1990 to 15.8 kt in 2021, representing a growth of 134 %. The BaP_{eq} emissions rose from 0.359 kt in 1990 to 0.853 kt in 2021, with an increase of 137 %. Additionally, the emissions of the 16 priority PAHs increased from 6.20 kt in 1990 to 14.5 kt in 2021. After supplementing the emissions inventory of the 16 priority PAHs in China (excluding cooking sources) by Wang et al. (2021), we find that cooking emissions accounted for 11.0 % of the total anthropogenic emissions of priority PAHs in China in 2017, and the share may be even larger in urban areas. Among these priority PAHs, naphthalene has the highest emissions share (46.8 %), followed by acenaphthylene (11.7 %) and phenanthrene (10.5 %). As for toxicity, dibenz[*a,h*]anthracene has the highest BaP_{eq} emissions share (42.8 %), followed by benzo[*a*]pyrene (36.1 %). Notably, high-molecular-weight PAHs (containing five- to seven-ringed PAHs) accounted for only 8.2 % of the emissions but contributed 85.3 % of the BaP_{eq} emissions due to their high toxicity. Over the 31 years, gaseous and particulate PAHs accounted for an average of 78.6 % and 21.4 % of the total PAH emissions, respectively. Commercial cooking
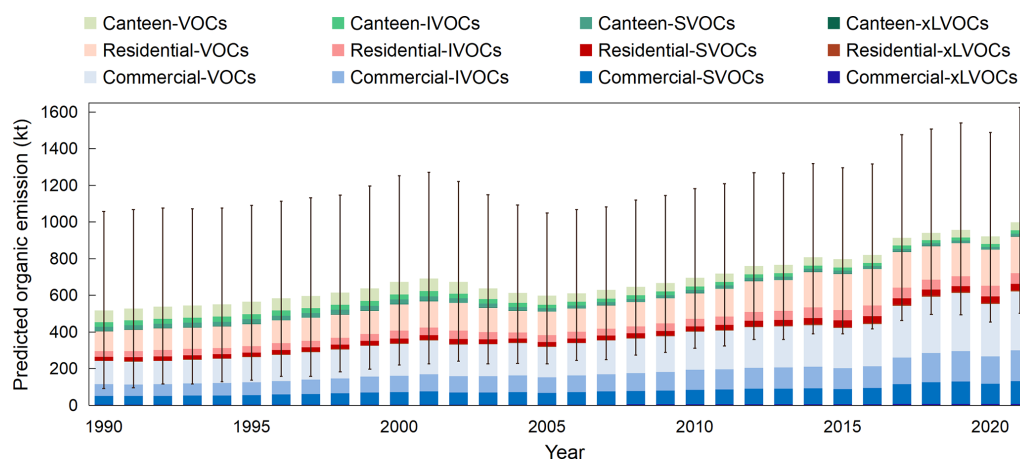
Figure 4. Organic emissions in the four volatility ranges from the three cooking sectors from 1990 to 2021 in China. The blue, red, and green bars represent the organic emissions from commercial cooking, residential cooking, and canteen cooking. Within each color group, the four different shades represent organic compounds of different volatility ranges. The error bars represent the uncertainty range at the 95 % confidence level.

remained the primary emission source, contributing 74.6 %–83.2 % of the national PAH emissions.

3.4 Comparison with other studies

We compare our cooking emission inventory with other China national cooking emission inventories (Cheng et al., 2022; Jin et al., 2021; Liang et al., 2022; Wang et al., 2018a; Zhang et al., 2024b). Most previous inventories only include pollutants such as VOCs and $PM_{2.5}$ (or organic carbon (OC), a component of $PM_{2.5}$) and provide emissions for only a single year. We first compare the national total emissions for the corresponding years and pollutants with theirs, and the results are presented in Table S8. Many previous inventories underestimated emissions due to the omission of emission sources (e.g., residential cooking) or the use of simple proxy data (e.g., population, meat consumption) (Cheng et al., 2022; Jin et al., 2021; Liang et al., 2022; Wang et al., 2018a), so their total emissions are much lower than ours. The latest studies (Zhang et al., 2024b), which used data from a service platform of Chinese catering enterprises, yielded national total VOC emissions relatively close to those of our inventory, supporting the accuracy of our emission calculations. In contrast, our inventory covers a longer time range (1990–2021), comprehensive cooking sources (including commercial, household, and canteen cooking), and a wider range of pollutants (not limited to VOCs and $PM_{2.5}$, but also including PAHs, UFP, etc.), which was difficult to achieve in previous studies.

Furthermore, our inventory demonstrates superior accuracy in spatial distribution. Unlike previous studies, this study precisely calculates emissions at the county level rather than first estimating provincial-level inventories and then downscaling them to the county level (or further to the grid level) using proxy data such as population. We compare our

inventory with the aforementioned latest inventory based on data from the catering service platform (Zhang et al., 2024b), which calculated the provincial emission inventory and then allocated it to the county levels based on population. We select the county-level emissions in Guangdong in 2020 as a case study for comparison, as Guangdong is a province with high cooking emissions, a large population, and a developed economy. In terms of total emissions, the Guangdong provincial emissions from this study and Zhang's inventory are 63.2 and 58.6 kt, respectively (Zhang et al., 2024b), showing close agreement. Figure 5 illustrates the emission intensity across all counties in Guangdong from the two inventories. The key difference between the two is that the emissions in our study are more concentrated in economically developed regions such as the Pearl River Delta, while the emission intensity in non-coastal areas is lower. This discrepancy arises because allocating provincial inventories to the county level based on population distribution may not fully reflect real-world conditions. In fact, some residential areas may have high population density, but dining activities are often more concentrated in commercial districts (Lin et al., 2022b). As discussed in Sect. 3.2, although the correlation between population and emissions is high at the county level ($R^2 = 0.873$), it is not a perfect match. In contrast, our methodology employs an effective machine learning model trained on advanced point-source cooking emission inventories (Li et al., 2023), with predictive variables related to population, the economy, and the catering industry. This method effectively captures the spatial distribution of comprehensive cooking activities, including information on the catering industry, residential cooking, and other factors considered in the previous advanced inventory (Li et al., 2023), thereby enabling an accurate representation of the spatial distribution of county-level cooking emissions.
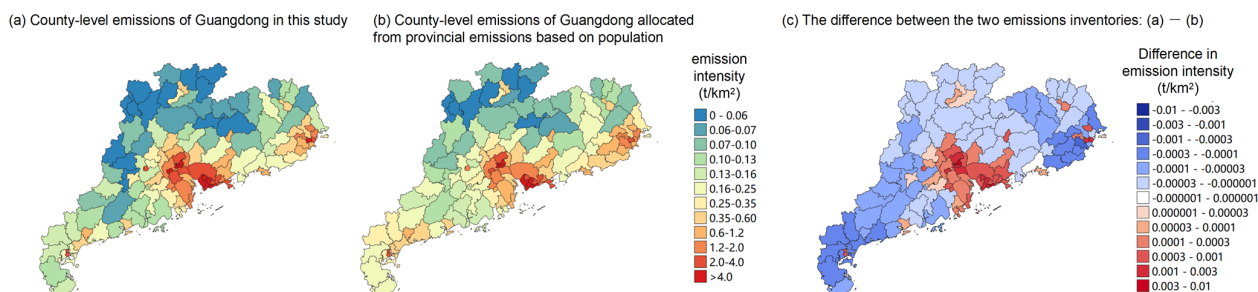
Figure 5. A comparison of **(a)** county-level emissions in this study, **(b)** county-level emissions allocated from provincial emissions based on population in Guangdong in 2020, and **(c)** the difference between the two emissions inventories.

4 Discussion

4.1 Spatiotemporal trends of county-level cooking emissions

To explore the spatiotemporal variation trends of cooking emissions in China, we obtain the changes in county-level organic emissions every 5 or 6 years through differencing, as illustrated in Fig. 6, where red indicates an increase in emissions during a particular period and blue represents a decrease. Our modeling explicitly incorporates key socioeconomic and policy variables such as population, urbanization rate, and PFIP and captures their relationships with county-level cooking emissions. Therefore, the spatiotemporal variation trends of emissions can be well explained by these underlying socioeconomic and policy drivers.

Before 2000, the observed changes in emissions could primarily be explained by population and economic factors, since China had few policies for cooking emission control (Zhao, 2004). From 1990 to 1995, emissions across most counties generally increased because of economic development, but emissions in a few counties decreased, probably due to population migration. The most significant emission growth was concentrated in Beijing, Shanghai, and the Pearl River Delta region – areas undergoing rapid economic development and urban expansion at the time (Démurger et al., 2002; Gaubatz, 2004). Conversely, emission reductions were typically observed in adjacent areas, suggesting population redistribution toward these emerging economic centers (Fan, 2005). The migration-driven spatial redistribution of emissions became more pronounced between 1995–2000 (Fan, 2005). For example, Guangdong's emissions became concentrated in the Pearl River Delta, while Zhejiang and Fujian's emissions clustered in the Yangtze River Delta and coastal regions. Emissions from eastern Sichuan have also shifted towards Chengdu (the provincial capital of Sichuan) and Chongqing. The migration was probably because these areas became focal points of economic reform during this period, attracting a large population influx (Fang et al., 2009), such as Chongqing being designated a directly controlled municipality in 1997 (Hong, 2004).

From 2000 to 2005, emissions declined in most parts of the country due to the implementation of China's first nationwide pollution control policy targeting cooking emissions in 2001 (see Sect. 3.3 for details) (State Environmental Protection Administration of China, 2001). However, Guangdong, Zhejiang, and Beijing maintained emission growth during this period, attributable to their exceptional economic expansion and continued population inflow (Kong, 2022; Zhu, 2012). From 2005 to 2010, cooking emissions in most counties in eastern China increased rapidly, likely because the emissions increase driven by rapid economic development outweighed the reductions from pollution control measures (Fleisher et al., 2010). In contrast, emissions in the slower-developing western regions decreased during this period (Fleisher et al., 2010).

During 2010–2021, emissions increased significantly nationwide except in provinces with stringent control policies (Beijing Environmental Protection Bureau, 2018; Feng et al., 2019; Liaoning Provincial Government, 2017; Shanxi Provincial Government, 2017). Coastal regions experienced the most notable emission increases, reflecting their faster economic growth and urbanization rates compared to inland regions (Shi, 2020). This rapid development also led to increased demand for cooking activities and a thriving catering industry due to population growth. Notably, Sichuan also exhibits significant emission increases, likely attributable to its special local cuisine that has attracted tourists nationwide and fostered a thriving catering industry (Li, 2017; Tian and Shen, 2024).

Additionally, we specifically examine the impact of the COVID-19 pandemic on cooking emissions from 2019 to 2021. In 2020, lockdown measures were implemented across China to control the spread of the pandemic (Chang et al., 2023). As shown in Fig. 6g, cooking emissions in many regions decreased in 2020. For example, Beijing, the Yangtze River Delta, and the Pearl River Delta saw significant reductions in cooking emissions, likely because these areas originally had thriving catering industries that were heavily restricted by lockdown policies in 2020 (Lan et al., 2018; Li et al., 2021; Yuan et al., 2024), leading to a substantial decrease in commercial cooking emissions. Conversely, emis-
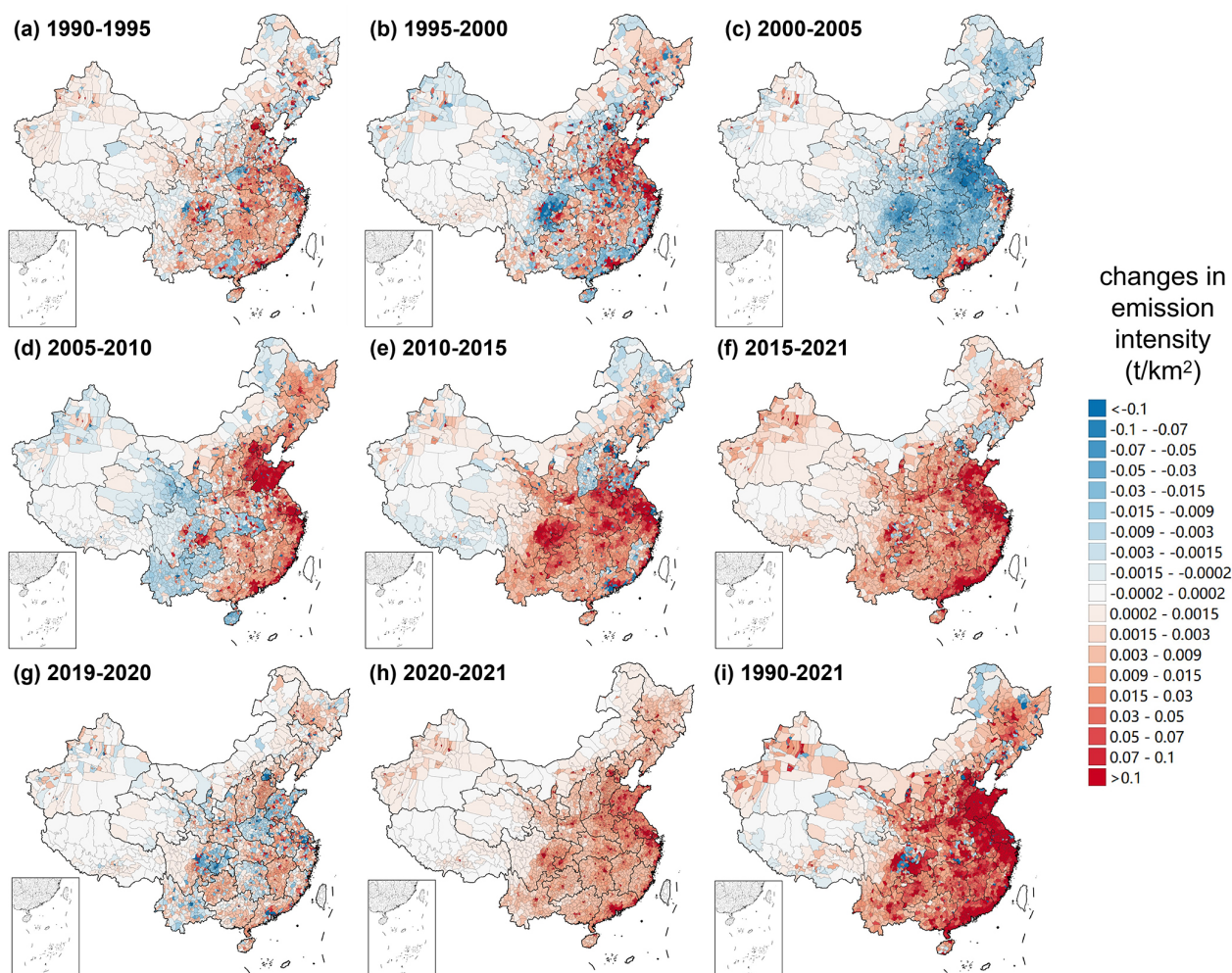
Figure 6. Changes in organic emission intensity in each county during different periods.

sions increased in many other regions, likely because lockdown policies forced people to stay at home, shifting cooking and dining from centralized locations like canteens and restaurants to more dispersed cooking and dining at home (Yang et al., 2021), thereby increasing overall cooking emissions. In 2021, as lockdown policies were gradually relaxed and the catering industry began to recover, overall cooking emissions rebounded nationwide (Li et al., 2021).

The observations above indicate that our emission calculation methodology can effectively capture the influences of pivotal external factors affecting emissions. In the 1990s, changes in emissions across counties were primarily influenced by economic growth rates and population migration. After 2000, variations in emissions were likely influenced by the promotion of pollution control measures and the development of the catering industry. Overall, cooking emissions have increased in the vast majority of the country over the last three decades (1990–2021) as shown in Fig. 6i, with particularly significant increases in the eastern region. Only a

few counties have seen a reduction in emissions, typically coinciding with population changes.

4.2 National emission drivers and key influencing factors for county-level emissions

Based on sensitivity simulation, we find significant differences in the driving factors of China's cooking organic emissions during different periods. The decomposition of emission change drivers for each period is shown in Fig. 7. From 1990 to 2000, emission levels grew slowly, mainly driven by the increasing population and urbanization rate, which contributed 50.3 % and 23.4 % to the emission growth, respectively. From 2001 to 2005, while population growth and urbanization also promoted an increase in emissions, the implementation of emission standards in 2001 significantly strengthened pollution control measures (State Environmental Protection Administration of China, 2001), leading to a considerable reduction (-21 %) in cooking emissions. From 2005 to 2015, while pollution emission standards continued

to be enforced, the lack of new regulatory policies targeting cooking sources limited further pollution control effectiveness, resulting in an average annual reduction of only about 1 % (Gao, 2020). Meanwhile, the rise in tertiary GDP and urbanization rates, marking rapid economic development, prompted a rapid increase in cooking emissions. Between 2005 and 2015, the rise in tertiary GDP and urbanization rates contributed 33.1 % and 29.5 % to the growth in emissions, respectively.

Since 2015, the increase in the number of chain restaurants has been the main driver for cooking emissions, possibly attributed to the prosperity of the catering industry brought about by online food delivery services (Maimaiti et al., 2018; Zhao et al., 2021). From 2015 to 2017, the number of users of online food delivery surged from 114 million to 343 million, and this figure continues to climb (Maimaiti et al., 2018). The tertiary GDP, urbanization rate, and population also contribute to the growth of cooking emissions. Meanwhile, stricter pollution control measures led to a more pronounced reduction in emissions (achieving an 8.4 % decrease from 2015 to 2019). However, this effect remained relatively limited compared to the rapid growth in total emissions (+30.0 % during the same period). This suggests that existing regulations were insufficient to address the growing emissions from the catering industry, highlighting the need for updated and more stringent policies specifically aimed at controlling cooking emissions. Overall, the primary driving factors of cooking organic emissions in the early (1990–2001), middle (2001–2015), and recent (2015–2021) periods are population growth, the rise in tertiary GDP and urbanization rates, and the increase in the number of chain restaurants.

We also explore the impact of the pandemic on the total national cooking emissions. From 2019 to 2020, factors such as the number of chain restaurants, population, and tertiary GDP were negatively affected by the pandemic, leading to a decrease in cooking emissions. However, some factors, including the urbanization rate and household cooking oil consumption, contributed to an increase in emissions. Despite the pandemic, China's urbanization rate rose from 60.6 % in 2019 to 63.9 % in 2020. This could be attributed to the Chinese government's efforts towards achieving the goal of *a moderately prosperous society in all respects before 2021* (Li, 2023), which involved continued urban development and infrastructure improvements. Additionally, the increase in household cooking oil consumption likely drove up emissions because lockdowns led to more people cooking at home rather than dining out (Yang et al., 2021). In 2021, as the economy recovered and the catering industry rebuilt, many factors (including the number of chain restaurants, population, and tertiary GDP) began to again lead the way toward increased cooking emissions (Li et al., 2021).

Additionally, we employ counterfactual analysis to quantitatively assess the emission reduction effects of control policy interventions. Specifically, we examine the two core policies introduced in Sect. 3.3: Policy-2001 and Policy-2013

(CPGPRC, 2013; State Environmental Protection Administration of China, 2001). While we collected provincial policies to determine province-specific PFIPs, these provincial regulations were largely driven by these two national policies, with variations in provincial response timing and enforcement stringency. Therefore, we sequentially exclude these two policies in our model, simulating counterfactual scenarios where provinces did not further strengthen pollution control measures after the policies were not implemented. The results are shown in Fig. S10. Without Policy-2013, 2021 emissions would have been 13.1 % higher. Without both policies, the increase would reach 49.4 %, highlighting the significant long-term impact of Policy-2001 in curbing emissions. The more modest effect of Policy-2013 may stem from its focus on comprehensive air pollution control, with catering sources being only a minor component. Moreover, the 2013 policy and subsequent provincial policies primarily emphasized increasing the installation rates of catering fume purification devices without stringent requirements for ensuring the removal efficiency of installed equipment (CPGPRC, 2013). Due to equipment aging and inadequate supervision (Li et al., 2023), the average removal efficiency remained low, resulting in no significant emission reductions despite higher installation rates.

Furthermore, we also pay attention to the key influential factors of cooking emissions of various counties at different development stages, applying the SHAP algorithm for the quantitative analysis. Figure S11 presents an overview of the SHAP values for each factor influencing emissions of the three cooking emission sectors, with the y axis sorted from high to low based on the impact of each factor on emissions. The influencing factors of commercial cooking emissions are the most complex. Urbanization rate (UR), population (POP), and EFs are the top three factors that have the greatest impact on commercial cooking emissions of counties, with increasing values leading to emissions growth. Additionally, PFIP, the tertiary GDP (GDP3), per capita household edible oil consumption (HOC), the number of chain restaurants (NCR), and per capita disposable income (DI) all affect commercial cooking emissions to some extent. Additionally, residential cooking emissions are mainly influenced by population and per capita household edible oil consumption. Canteen cooking emissions are mainly affected by population, PFIP, and the population of employees in enterprises (PEE).

We further analyze the marginal effects of each influencing factor on the cooking organic emissions – that is, how emission values (indicated by SHAP values) vary with the values of individual influencing factors. Taking commercial cooking emissions as an example, a plot of the partial dependence of SHAP values on the main influencing factors is shown in Fig. 8. For the urbanization rate, the relationship between SHAP values and the urbanization rate forms an S-shaped curve. This means that the sensitivity of commercial cooking emissions to the urbanization rate is relatively high
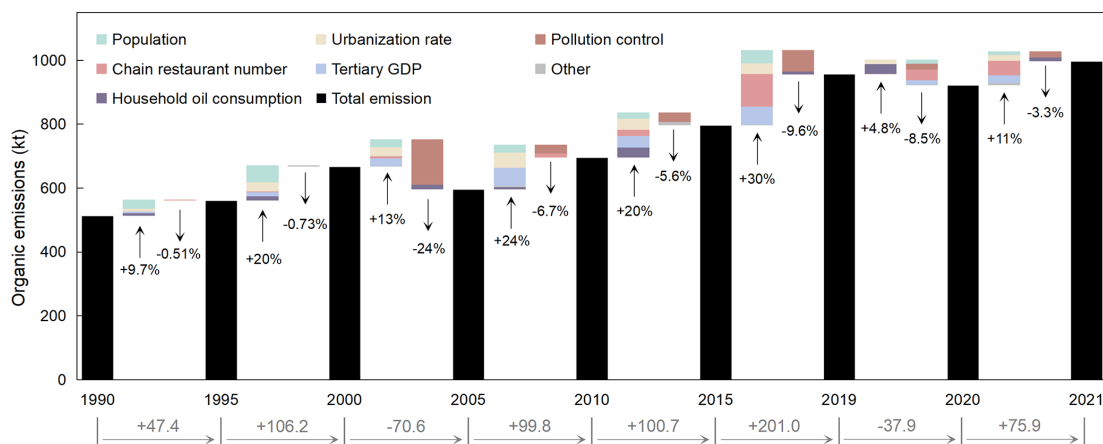
Figure 7. The contribution of various driving factors to the changes in national cooking organic emissions across different periods.
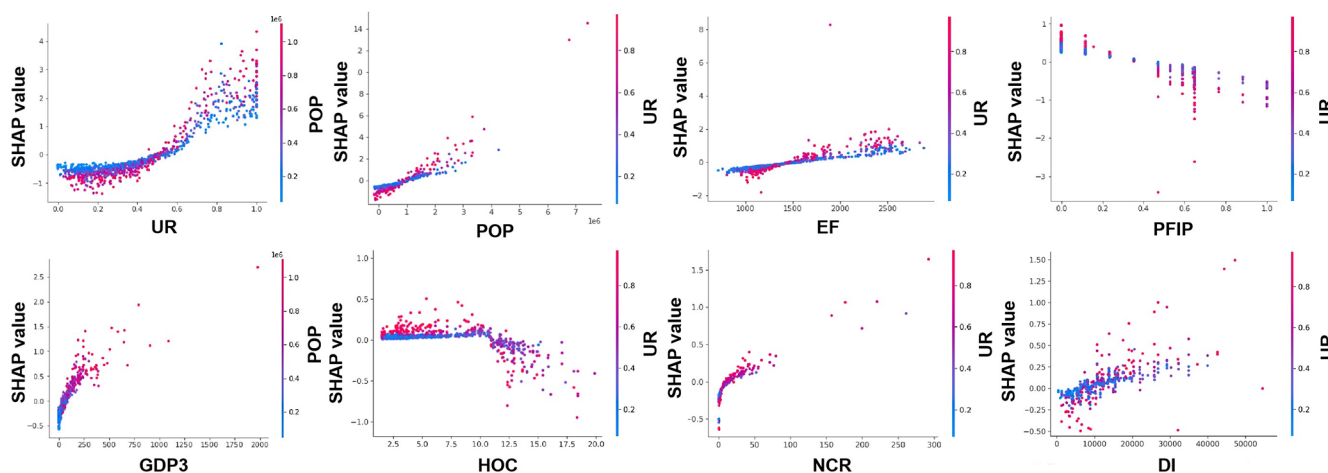

Figure 8. Plot of the partial dependence of SHAP values on the main influencing factors of commercial cooking organic emissions in Chinese counties.

when the urbanization rate is at the medium level (45 %–75 %). Additionally, the SHAP values are approximately linearly correlated with the local population and EFs, while the emissions are negatively correlated with the PFIP value. The relationship between the tertiary GDP and the number of chain restaurants and SHAP values approximates a logarithmic growth curve, where growth is rapid at lower feature values and slows down as the feature values increase. The relationship between HOC and commercial cooking emissions is very intricate. When the HOC value is low, its increase signifies an improvement in people's living standards starting from a low level, which in turn leads to a corresponding increase in commercial cooking emissions. As the HOC value reaches a certain level, further increases indicate an increase in the frequency of residential cooking that competes with commercial cooking, resulting in a decrease in commercial cooking emissions. Finally, an overall increase in per capita disposable income will lead to an increase in commercial cooking emissions, as this can be explained by people hav-

ing more funds for dining out. The relationship between residential cooking emissions and its main influencing factors (population and HOC), as well as the relationship between canteen cooking emissions and its main influencing factors (POP and PFIP), is very similar to the relationship between commercial cooking emissions and these variables.

5   Data availability

The county-level cooking emission inventory in China from 1990 to 2021 is publicly available at https://doi.org/10.6084/m9.figshare.26085487 (Li et al., 2025). This dataset provides comprehensive emissions data at the county level, covering all 2848 counties in mainland China based on the 2020 administrative divisions, and includes annual emissions for every year from 1990 to 2021. The emissions are categorized by subsectors, including commercial cooking, residential cooking, and canteen cooking, and by pollutants, including organics

across the full volatility range (VOCs, SVOCs, IVOCs, and xLVOCs), $PM_{2.5}$, UFPs, and PAHs. The types of emission pollutants related to PAHs include gaseous PAHs, particulate PAHs, and BaP_{eq}. Also, the main text and the Supplement (Tables S2–S5) provide detailed listings of emission factors, PFIPs, PAH TEF, and other parameters used for calculating emissions. Additionally, the input data for the machine learning models, such as population, economic, and catering-related statistical indicators, are sourced from the Chinese County Statistical Yearbook, China Urban Statistical Yearbook, and China Market Statistics Yearbook, with a full description provided in Table S1 (National Bureau of Statistics of China, 2022a, c, b).

6 Conclusions and implications

In this study, leveraging machine learning to overcome the challenges of obtaining activity data, we establish China's first county-level cooking emission inventory, with a temporal scale extending back to 1990. Unlike previous inventories that relied on proxy data such as population for calculation and downscaling, our inventory employs a powerful ensemble machine learning model to capture the complex relationships between county-level cooking activities and factors involving population, economics, and the catering industry. This method enables direct calculation of emissions at the county level, resulting in spatial distributions that better reflect real-world conditions. Moreover, our method can sensitively identify the impact of external factors, such as the COVID-19 pandemic and the rise of food delivery services, on cooking emissions. Based on this accurate, high-resolution, and long-term inventory, we have updated the scientific understanding of the spatiotemporal trends and driving forces of cooking emissions.

Given that cooking is a significant source of $PM_{2.5}$ (Yuan et al., 2023), our long-term, high-spatial-resolution cooking emission inventory provides essential data for accurately simulating $PM_{2.5}$ concentrations and conducting precise source apportionments at large spatiotemporal scale. Furthermore, for the first time, we incorporate UFPs and PAHs into the national cooking emission inventory, filling a gap in studies on the health impact of cooking emissions. Previous studies on the health impacts of cooking emissions primarily focused on indoor environments (Chen et al., 2018; Zhang et al., 2023; Zhao and Zhao, 2018). However, pollutants emitted into the outdoor atmosphere from cooking may also have significant health risks due to the proximity of cooking emission sources to the human living environment. Our accurate, high-resolution cooking inventory, combined with the inclusion of highly toxic pollutants, provides critical but previously missing data for assessing exposure risks to cooking-related pollutants in outdoor environments. This enables a comprehensive understanding of the health impacts of cooking emissions by integrating both indoor and outdoor exposure assessments.

Our identification of the spatiotemporal patterns and driving factors of national cooking emissions also provides valuable insights for targeted policy formulation. With the significant reduction in emissions from sectors such as industry and energy, the critical impact of cooking emissions is becoming increasingly prominent and may become a major source in the future (Zhao and Zhao, 2018). However, our results indicate that existing control measures are insufficient to curb the rapid growth of cooking emissions, necessitating the development of updated and more effective control strategies. Given that cooking is a fundamental human need, reducing emissions by restricting cooking activities or altering dietary habits is not feasible. A more viable approach is to enhance end-of-pipe treatment. However, cooking emission sources are numerous and widespread, making comprehensive control efforts highly labor-intensive. Fortunately, our high-resolution inventory reveals a strong spatial coupling between high emission intensity and population density. Specifically, 30 % of counties – covering only 14.5 % of the national land area – contribute over 60 % of cooking-related organic emissions while housing 60 % of the population. These areas face substantial population exposure risks, and prioritizing stricter controls there could effectively mitigate health impacts. Additionally, the limited effectiveness of recent pollution control policies may stem from their sole focus on the installation rate of purification facilities while neglecting actual purification efficiency. Therefore, future efforts should establish stricter regulations or standards for purification efficiency and strengthen enforcement. Residential emissions have consistently accounted for ∼ 30 % of total cooking emissions, yet targeted policies and specialized purification facilities remain scarce. Potential solutions include developing compact purifiers for household kitchens and implementing exhaust purification systems for residential chimneys. Finally, our fine-scale emission inventory can serve as a key input for air quality modeling and control strategy optimization models, enabling further exploration of differentiated mitigation strategies that consider costs, health benefits, and local resource capacity.

Additionally, the methodology adopted in this study also offers a reference for the long-term and accurate estimation of emissions from other sources and other regions. We innovatively use counties as the basic unit to estimate emissions, which not only provides the machine learning model with rich and wide-span county samples at different development stages, enhancing the model's performance, but also ensures a high spatial resolution. The data used for machine learning modeling are also readily available, significantly reducing the difficulty of activity-level acquisition. Similar to cooking emissions, emissions from domestic combustion, for example, can be estimated using statistical indicators such as temperature, per capita disposable income, urbanization rate, and energy consumption. In other regions, this methodology also shows potential in estimating high-resolution emissions through machine learning models and

localized datasets. This contributes to more comprehensive and accurate research on air pollution.

We also acknowledge some limitations of our study. For county-level emissions, while SHAP provides interpretable insights into feature importance, it cannot infer causality to identify the underlying driving factors. Future work could employ causal inference techniques such as counterfactual prediction and difference-in-differences analysis (Dong et al., 2022; Li et al., 2024a) to more accurately assess the actual impacts of policy interventions and socioeconomic factors on emission trends, thereby providing more robust evidence to support localized emission reduction policies.

Author contributions. ZL, SW, BZ, and SL designed the study. ZL developed the emission inventory. SL, ZS, XY, and GH provided key data for the calculation of the emission inventory. DY, QW, FZ, XY, and YZ helped to improve the emission inventory. ZL wrote the original draft; all the coauthors revised the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Publisher's remark: please note that Figs. 3 and 6 as well as the key figure contain disputed territories.

Review statement. This paper was edited by Yuqiang Zhang and reviewed by two anonymous referees.

References

Abdullahi, K. L., Delgado-Saborit, J. M., and Harrison, R. M.: Emissions and indoor concentrations of particulate matter and its specific chemical components from cooking: A review, Atmos. Environ., 71, 260–294, https://doi.org/10.1016/j.atmosenv.2013.01.061, 2013.

Alduailij, M., Khan, Q. W., Tahir, M., Sardaraz, M., Alduailij, M., and Malik, F.: Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method, Symmetry, 14, 1095, https://doi.org/10.3390/sym14061095, 2022.

Altmann, A., Toloşi, L., Sander, O., and Lengauer, T.: Permutation importance: a corrected feature importance measure, Bioinformatics, 26, 1340–1347, https://doi.org/10.1093/bioinformatics/btq134, 2010.

Beijing Environmental Protection Bureau: Emission standards of air pollutants for catering industry, Beijing Environmental Protection Bureau, China Environmental Science Press, DB 11/1488-2018, http://sthjj.beijing.gov.cn/eportal/fileDir/bjhrb/resource/cms/2018/01/2018012214300944367.pdf (last access: 11 February 2025), 2018.

Belete, D. M. and Huchaiah, M. D.: Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results, International Journal of Computers and Applications, 44, 875–886, https://doi.org/10.1080/1206212X.2021.1974663, 2022.

Carneiro, T. C., Rocha, P. A. C., Carvalho, P. C. M., and Fernández-Ramírez, L. M.: Ridge regression ensemble of machine learning models applied to solar and wind forecasting in Brazil and Spain, Appl. Energ., 314, 118936, https://doi.org/10.1016/j.apenergy.2022.118936, 2022.

Chafe, Z. A., Brauer, M., Klimont, Z., Van Dingenen, R., Mehta, S., Rao, S., Riahi, K., Dentener, F., and Smith, K. R.: Household Cooking with Solid Fuels Contributes to Ambient $PM_{2.5}$ Air Pollution and the Burden of Disease, Environ. Health Persp., 122, 1314–1320, https://doi.org/10.1289/ehp.1206340, 2014.

Chang, X., Zhao, B., Zheng, H., Wang, S., Cai, S., Guo, F., Gui, P., Huang, G., Wu, D., Han, L., Xing, J., Man, H., Hu, R., Liang, C., Xu, Q., Qiu, X., Ding, D., Liu, K., Han, R., Robinson, A. L., and Donahue, N. M.: Full-volatility emission framework corrects missing and underestimated secondary organic aerosol sources, One Earth, 5, 403–412, https://doi.org/10.1016/j.oneear.2022.03.015, 2022.

Chang, X., Zheng, H., Zhao, B., Yan, C., Jiang, Y., Hu, R., Song, S., Dong, Z., Li, S., Li, Z., Zhu, Y., Shi, H., Jiang, Z., Xing, J., and Wang, S.: Drivers of High Concentrations of Secondary Organic Aerosols in Northern China during the COVID-19 Lockdowns, Environ. Sci. Technol., 57, 5521–5531, https://doi.org/10.1021/acs.est.2c06914, 2023.

Chen, C. and Zhao, B.: Indoor Emissions Contributed the Majority of Ultrafine Particles in Chinese Urban Residences, Environ. Sci. Technol., 58, 8444–8456, https://doi.org/10.1021/acs.est.4c00556, 2024.

Chen, C., Zhao, Y., Zhang, Y., and Zhao, B.: Source strength of ultrafine and fine particle due to Chinese cooking, Procedia Engineer., 205, 2231–2237, https://doi.org/10.1016/j.proeng.2017.10.062, 2017.

Earth Syst. Sci. Data, 17, 5113–5135, 2025

https://doi.org/10.5194/essd-17-5113-2025

Chen, C., Zhao, Y., and Zhao, B.: Emission Rates of Multiple Air Pollutants Generated from Chinese Residential Cooking, Environ. Sci. Technol., 52, 1081–1087, https://doi.org/10.1021/acs.est.7b05600, 2018.

Chen, D., Gu, X., Guo, H., Cheng, T., Yang, J., Zhan, Y., and Fu, Q.: Spatiotemporally continuous $PM_{2.5}$ dataset in the Mekong River Basin from 2015 to 2022 using a stacking model, Sci. Total Environ., 914, 169801, https://doi.org/10.1016/j.scitotenv.2023.169801, 2024.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 13–17 August 2016, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Chen, Y., Ho, K. F., Ho, S. S. H., Ho, W. K., Lee, S. C., Yu, J. Z., and Sit, E. H. L.: Gaseous and particulatepolycyclicaromatichydrocarbons (PAHs) emissions from commercial restaurants in Hong Kong, J. Environ. Monitor., 9, 1402–1409, https://doi.org/10.1039/B710259C, 2007.

Cheng, Y., Kong, S., Yao, L., Zheng, H., Wu, J., Yan, Q., Zheng, S., Hu, Y., Niu, Z., Yan, Y., Shen, Z., Shen, G., Liu, D., Wang, S., and Qi, S.: Multiyear emissions of carbonaceous aerosols from cooking, fireworks, sacrificial incense, joss paper burning, and barbecue as well as their key driving forces in China, Earth Syst. Sci. Data, 14, 4757–4775, https://doi.org/10.5194/essd-14-4757-2022, 2022.

CPGPRC (The Central People's Government of the People's Republic of China): Action plan for the prevention and control of air pollutants, General Office of the State Council, http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm (last access: 11 February 2025), 2013.

Daoud, J. I.: Multicollinearity and Regression Analysis, J. Phys. Conf. Ser., 949, 012009, https://doi.org/10.1088/1742-6596/949/1/012009, 2017.

Démurger, S., Sachs, J. D., Woo, W. T., Bao, S., Chang, G., and Mellinger, A.: Geography, Economic Policy, and Regional Development in China, Asian Econ. Pap., 1, 146–197, https://doi.org/10.1162/153535102320264512, 2002.

Dong, F., Li, Y., Qin, C., Zhang, X., Chen, Y., Zhao, X., and Wang, C.: Information infrastructure and greenhouse gas emission performance in urban China: A difference-in-differences analysis, J. Environ. Manage., 316, 115252, https://doi.org/10.1016/j.jenvman.2022.115252, 2022.

Ebrahimi, S. H. S., Majidzadeh, K., and Gharehchopogh, F. S.: A hybrid principal label space transformation-based ridge regression and decision tree for multi-label classification, Evolving Systems, 15, 2441–2477, https://doi.org/10.1007/s12530-024-09618-0, 2024.

Fameli, K.-M., Kladakis, A., and Assimakopoulos, V. D.: Inventory of Commercial Cooking Activities and Emissions in a Typical Urban Area in Greece, Atmosphere, 13, 792, https://doi.org/10.3390/atmos13050792, 2022.

Fan, C. C.: Interprovincial Migration, Population Redistribution, and Regional Development in China: 1990 and 2000 Census Comparisons, Prof. Geogr., 57, 295–311, https://doi.org/10.1111/j.0033-0124.2005.00479.x, 2005.

Fang, C., Yang, D., and Meiyan, W.: Migration and labor mobility in China, Human Development Research Paper (HDRP) Series, 09, https://mpra.ub.uni-muenchen.de/19187/ (last access: 11 February 2025), 2009

Feng, S., Shen, X., Hao, X., Cao, X., Li, X., Yao, X., Shi, Y., Lv, T., and Yao, Z.: Polycyclic and nitro-polycyclic aromatic hydrocarbon pollution characteristics and carcinogenic risk assessment of indoor kitchen air during cooking periods in rural households in North China, Environ. Sci. Pollut. R., 28, 11498–11508, https://doi.org/10.1007/s11356-020-11316-8, 2021.

Feng, Y., Ning, M., Lei, Y., Sun, Y., Liu, W., and Wang, J.: Defending blue sky in China: Effectiveness of the "Air Pollution Prevention and Control Action Plan" on air quality improvements from 2013 to 2017, J. Environ. Manage., 252, 109603, https://doi.org/10.1016/j.jenvman.2019.109603, 2019.

Fleisher, B., Li, H., and Zhao, M. Q.: Human capital, economic growth, and regional inequality in China, J. Dev. Econ., 92, 215–231, https://doi.org/10.1016/j.jdeveco.2009.01.010, 2010.

Frome, E. L.: The Analysis of Rates Using Poisson Regression Models, Biometrics, 39, 665–674, https://doi.org/10.2307/2531094, 1983.

Gao, J.: Study on legislation of fume pollution prevention and control in catering industry in China, Hunan Normal University, https://doi.org/10.27137/d.cnki.ghusu.2020.000475, 2020.

Gaubatz, P.: Globalization and the development of new central business districts in Beijing, Shanghai and Guangzhou, in: Restructuring the Chinese city, Routledge, 87–108, https://doi.org/10.4324/9780203414460, 2004.

Géhin, E., Ramalho, O., and Kirchner, S.: Size distribution and emission rate measurement of fine and ultrafine particle from indoor human activities, Atmos. Environ., 42, 8341–8352, https://doi.org/10.1016/j.atmosenv.2008.07.021, 2008.

Guo, Z., Chen, X., Wu, D., Huo, Y., Cheng, A., Liu, Y., Li, Q., and Chen, J.: Higher Toxicity of Gaseous Organics Relative to Particulate Matters Emitted from Typical Cooking Processes, Environ. Sci. Technol., 57, 17022–17031, https://doi.org/10.1021/acs.est.3c05425, 2023.

Hong, L.: Chongqing: Opportunities and Risks, China Quart., 178, 448–466, https://doi.org/10.1017/S0305741004000256, 2004.

Hou, L., Dai, Q., Song, C., Liu, B., Guo, F., Dai, T., Li, L., Liu, B., Bi, X., Zhang, Y., and Feng, Y.: Revealing Drivers of Haze Pollution by Explainable Machine Learning, Environ. Sci. Tech. Let., 9, 112–119, https://doi.org/10.1021/acs.estlett.1c00865, 2022.

Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., and Liu, Y.: Estimating $PM_{2.5}$ Concentrations in the Conterminous United States Using the Random Forest Approach, Environ. Sci. Technol., 51, 6936–6944, https://doi.org/10.1021/acs.est.7b01210, 2017.

Huang, D. D., Zhu, S., An, J., Wang, Q., Qiao, L., Zhou, M., He, X., Ma, Y., Sun, Y., Huang, C., Yu, J. Z., and Zhang, Q.: Comparative Assessment of Cooking Emission Contributions to Urban Organic Aerosol Using Online Molecular Tracers and Aerosol Mass Spectrometry Measurements, Environ. Sci. Technol., 55, 14526–14535, https://doi.org/10.1021/acs.est.1c03280, 2021.

Huang, L., Zhao, B., Wang, S., Chang, X., Klimont, Z., Huang, G., Zheng, H., and Hao, J.: Global Anthropogenic Emissions of Full-Volatility Organic Compounds, Environ. Sci. Technol., 57, 16435–16445, https://doi.org/10.1021/acs.est.3c04106, 2023.

Huy, L. N., Oanh, N. T. K., Phuc, N. H., and Nhung, C. P.: Survey-based inventory for atmospheric emissions from residential com-

https://doi.org/10.5194/essd-17-5113-2025

Earth Syst. Sci. Data, 17, 5113–5135, 2025

bustion in Vietnam, Environ. Sci. Pollut. R., 28, 10678–10695, https://doi.org/10.1007/s11356-020-11067-6, 2021.

Jansson, M.: A Comparison of Detransformed Logarithmic Regressions and Power Function Regressions, Geogr. Ann. A, 67, 61–70, https://doi.org/10.1080/04353676.1985.11880130, 1985.

Jiménez-Luna, J., Grisoni, F., and Schneider, G.: Drug discovery with explainable artificial intelligence, Nat. Mach. Intell., 2, 573–584, https://doi.org/10.1038/s42256-020-00236-4, 2020.

Jin, W., Zhi, G., Zhang, Y., Wang, L., Guo, S., Zhang, Y., Xue, Z., Zhang, X., Du, J., Zhang, H., Ren, Y., Xu, P., Ma, J., Zhao, W., Wang, L., and Fu, R.: Toward a national emission inventory for the catering industry in China, Sci. Total Environ., 754, 142184, https://doi.org/10.1016/j.scitotenv.2020.142184, 2021.

Jørgensen, R. B., Strandberg, B., Sjaastad, A. K., Johansen, A., and Svendsen, K.: Simulated Restaurant Cook Exposure to Emissions of PAHs, Mutagenic Aldehydes, and Particles from Frying Bacon, J. Occup. Environ. Hyg., 10, 122–131, https://doi.org/10.1080/15459624.2012.755864, 2013.

Kim, S., Machesky, J., Gentner, D. R., and Presto, A. A.: Real-world observations of reduced nitrogen and ultrafine particles in commercial cooking organic aerosol emissions, Atmos. Chem. Phys., 24, 1281–1298, https://doi.org/10.5194/acp-24-1281-2024, 2024.

Kong, Z.: Statistical Analysis of the Differential Economic Development of China's Provincial Economies, master, Southwest University, https://doi.org/10.27684/d.cnki.gxndx.2022.001077, 2022.

Lachowicz, J. I., Milia, S., Jaremko, M., Oddone, E., Cannizzaro, E., Cirrincione, L., Malta, G., Campagna, M., and Lecca, L. I.: Cooking Particulate Matter: A Systematic Review on Nanoparticle Exposure in the Indoor Cooking Environment, Atmosphere, 14, 12, https://doi.org/10.3390/atmos14010012, 2023.

Lan, T., Yu, M., Xu, Z., and Wu, Y.: Temporal and Spatial Variation Characteristics of Catering Facilities Based on POI Data: A Case Study within 5th Ring Road in Beijing, Procedia Comput. Sci., 131, 1260–1268, https://doi.org/10.1016/j.procs.2018.04.343, 2018.

Larsen, J. C. and Larsen, P. B.: Chemical carcinogens, in: Issues in Environmental Science and Technology, Air Pollution and Health, The Royal Society of Chemistry, Cambridge, UK, 33–56, https://doi.org/10.1039/9781847550095-00033, 1998.

Lee, B. P., Li, Y. J., Yu, J. Z., Louie, P. K. K., and Chan, C. K.: Characteristics of submicron particulate matter at the urban roadside in downtown Hong Kong – Overview of 4 months of continuous high-resolution aerosol mass spectrometer measurements, J. Geophys. Res.-Atmos., 120, 7040–7058, https://doi.org/10.1002/2015JD023311, 2015.

Li, B., Zhong, Y., Zhang, T., and Hua, N.: Transcending the COVID-19 crisis: Business resilience and innovation of the restaurant industry in China, Journal of Hospitality and Tourism Management, 49, 44–53, https://doi.org/10.1016/j.jhtm.2021.08.024, 2021.

Li, C., Liu, W., and Yang, H.: Deep causal inference for understanding the impact of meteorological variations on traffic, Transport. Res. C-Emer., 165, 104744, https://doi.org/10.1016/j.trc.2024.104744, 2024a.

Li, N.: Quantitative Uncertainty Analysis and Verification of Emission Inventory in Guangdong Province 2012, Master thesis, South China University of Technology, China,

https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201802&filename=1017733919.nh (last access: 11 February 2025), 2017.

Li, Z., Zhao, B., Li, S., Shi, Z., Yin, D., Wu, Q., Zhang, F., Yun, X., Huang, G., Zhu, Y., and Wang, S.: County-level Cooking Emission inventory in China from 1990 to 2021, figshare [data set], https://doi.org/10.6084/m9.figshare.26085487, 2025.

Li, C.-T., Lin, Y.-C., Lee, W.-J., and Tsai, P.-J.: Emission of polycyclic aromatic hydrocarbons and their carcinogenic potencies from cooking sources to the urban atmosphere, Environ. Health Persp., 111, 483–487, https://doi.org/10.1289/ehp.5518, 2003.

Li, X.: Advantages and Countermeasures of the Development of China's Food Tourism Industry, 2017 4th International Conference on Education, Management and Computing Technology (ICEMCT 2017), Hangzhou, China, 15–16 April 2017, 1118–1122, https://doi.org/10.2991/icemct-17.2017.244, 2017.

Li, Y.: Research on the Historical Experience of the Communist Party of China in Building a Moderately Prosperous Society in All Respects, Lanzhou Jiaotong University, https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202401&filename=1023116483.nh&uniplatform=OVERSEA&v=dUWnesjuZMK7xnFjuN_3sBDaY5--jvuWbzEXnYT-W3T-9Wj2I_V5_QYTQ8d5q_MQ (last access: 11 February 2025), 2023.

Li, Y.-C., Qiu, J.-Q., Shu, M., Ho, S. S. H., Cao, J.-J., Wang, G.-H., Wang, X.-X., and Zhao, X.-Q.: Characteristics of polycyclic aromatic hydrocarbons in $PM_{2.5}$ emitted from different cooking activities in China, Environ. Sci. Pollut. R., 25, 4750–4760, https://doi.org/10.1007/s11356-017-0603-0, 2018.

Li, Z., Wang, S., Li, S., Wang, X., Huang, G., Chang, X., Huang, L., Liang, C., Zhu, Y., Zheng, H., Song, Q., Wu, Q., Zhang, F., and Zhao, B.: High-resolution emission inventory of full-volatility organic compounds from cooking in China during 2015–2021, Earth Syst. Sci. Data, 15, 5017–5037, https://doi.org/10.5194/essd-15-5017-2023, 2023.

Li, Z., Zhao, B., Yin, D., Wang, S., Qiao, X., Jiang, J., Li, Y., Shen, J., He, Y., Chang, X., Li, X., Liu, Y., Li, Y., Liu, C., Qi, X., Chen, L., Chi, X., Jiang, Y., Li, Y., Wu, J., Nie, W., and Ding, A.: Modeling the Formation of Organic Compounds across Full Volatility Ranges and Their Contribution to Nanoparticle Growth in a Polluted Atmosphere, Environ. Sci. Technol., 58, 1223–1235, https://doi.org/10.1021/acs.est.3c06708, 2024b.

Li, Z., Zhao, B., Li, S., Shi, Z., Yin, D., Wu, Q., Zhang, F., Yun, X., Huang, G., Zhu, Y., and Wang, S.: County-level Cooking Emission inventory in China from 1990 to 2021, figshare [data set], https://doi.org/10.6084/m9.figshare.26085487, 2025.

Liang, X., Chen, L., Liu, M., Lu, Q., Lu, H., Gao, B., Zhao, W., Sun, X., Xu, J., and Ye, D.: Carbonyls from commercial, canteen and residential cooking activities as crucial components of VOC emissions in China, Sci. Total Environ., 846, 157317, https://doi.org/10.1016/j.scitotenv.2022.157317, 2022.

Liaoning Provincial Government: Air Pollution Prevention and Control Regulations for Liaoning Province, Department of Ecology and Environment, https://sthj.ln.gov.cn/sthj/zfxxgk/fdzdgknr/lzyj/dffg/20240319141181365214/index.shtml (last access: 11 February 2025), 2017.

Lin, C., Huang, R.-J., Duan, J., Zhong, H., and Xu, W.: Polycyclic aromatic hydrocarbons from cook-

ing emissions, Sci. Total Environ., 818, 151700, https://doi.org/10.1016/j.scitotenv.2021.151700, 2022a.

Lin, P., Gao, J., Xu, Y., Schauer, J. J., Wang, J., He, W., and Nie, L.: Enhanced commercial cooking inventories from the city scale through normalized emission factor dataset and big data, Environ. Pollut., 315, 120320, https://doi.org/10.1016/j.envpol.2022.120320, 2022b.

Liu, R., Ma, Z., Gasparrini, A., de la Cruz, A., Bi, J., and Chen, K.: Integrating Augmented In Situ Measurements and a Spatiotemporal Machine Learning Model To Back Extrapolate Historical Particulate Matter Pollution over the United Kingdom: 1980–2019, Environ. Sci. Technol., 57, 21605–21615, https://doi.org/10.1021/acs.est.3c05424, 2023.

Logue, J. M., Klepeis, N. E., Lobscheid, A. B., and Singer, B. C.: Pollutant Exposures from Natural Gas Cooking Burners: A Simulation-Based Assessment for Southern California, Environ. Health Persp., 122, 43–50, https://doi.org/10.1289/ehp.1306673, 2014.

Lou, P., Wu, T., Yin, G., Chen, J., Zhu, X., Wu, X., Li, R., and Yang, S.: A novel framework for multiple thermokarst hazards risk assessment and controlling environmental factors analysis on the Qinghai-Tibet Plateau, CATENA, 246, 108367, https://doi.org/10.1016/j.catena.2024.108367, 2024.

Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, arXiv [preprint], https://doi.org/10.48550/arXiv.1705.07874, 25 November 2017.

Maimaiti, M., Zhao, X., Jia, M., Ru, Y., and Zhu, S.: How we eat determines what we become: opportunities and challenges brought by food delivery industry in a changing world in China, Eur. J. Clin. Nutr., 72, 1282–1286, https://doi.org/10.1038/s41430-018-0191-1, 2018.

Malcolm, H. M., Dobson, S., and Great Britain HM Inspectorate of Pollution: The Calculation of the Environmental Assessment Level (EAL) for Atmospheric PAHs Using Relative Potencies, Her Majesty's Inspectorate of Pollution, London, UK, 34–46, https://search.worldcat.org/de/title/calculation-of-the-environmental-assessment-level-eal-for-atmospheric-pahs-using-relative-potencies/oclc/60235224 (last access: 11 February 2025), 1994.

McDonald, G. C.: Ridge regression, WIREs Computational Statistics, 1, 93–100, https://doi.org/10.1002/wics.14, 2009.

Mohr, C., DeCarlo, P. F., Heringa, M. F., Chirico, R., Slowik, J. G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J. L., Crippa, M., Zimmermann, R., Baltensperger, U., and Prévôt, A. S. H.: Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data, Atmos. Chem. Phys., 12, 1649–1665, https://doi.org/10.5194/acp-12-1649-2012, 2012.

Murti, D. M. P., Pujianto, U., Wibawa, A. P., and Akbar, M. I.: K-Nearest Neighbor (K-NN) based Missing Data Imputation, in: 2019 5th International Conference on Science in Information Technology (ICSITech), 2019 5th International Conference on Science in Information Technology (ICSITech), Yogyakarta, Indonesia, 23–24 October 2019, 83–88, https://doi.org/10.1109/ICSITech46713.2019.8987530, 2019.

Myers, R. H. and Montgomery, D. C.: A Tutorial on Generalized Linear Models, J. Qual. Technol., 29, 274–291, https://doi.org/10.1080/00224065.1997.11979769, 1997.

Naseri, M., Sultanbekovna, A. A., Malekipirbazari, M., Kenzhegaliyeva, E., Buonanno, G., Stabile, L., Hopke, P. K., Cassee, F., Crape, B., Sabanov, S., Zhumambayeva, S., Ozturk, F., Tadi, M. J., Torkmahalleh, M. A., and Shah, D.: Human exposure to aerosol from indoor gas stove cooking and the resulting cardiovascular system responses, Toxicology Reports, 13, 101716, https://doi.org/10.1016/j.toxrep.2024.101716, 2024.

National Bureau of Statistics of China: China Statistical Yearbook, China Statistics Press, Beijing, ISBN 978-7-5037-9950-1, https://oversea.cnki.net/KNavi/YearbookDetail?pcode=CYFD&pykm=YINFN&bh=&uniplatform=OVERSEA&language=en (last access: 11 February 2025), 2022a.

National Bureau of Statistics of China: China City Statistical Yearbook, China Statistics Press, Beijing, ISBN 978-7-5230-0630-6, https://oversea.cnki.net/KNavi/YearbookDetail?pcode=CYFD&pykm=YZGCA&bh=&uniplatform=OVERSEA&language=en (last access: 11 February 2025), 2022b.

National Bureau of Statistics of China: China Statistical Yearbook (Township), China Statistics Press, Beijing, ISBN 978-7-5230-0632-0, https://oversea.cnki.net/KNavi/YearbookDetail?pcode=CYFD&pykm=YXSKU&bh=&uniplatform=OVERSEA&language=en (last access: 11 February 2025), 2022c.

Nayak, S. K. and Ojha, A. C.: Data Leakage Detection and Prevention: Review and Research Directions, in: Machine Learning and Information Processing, Singapore, 203–212, https://doi.org/10.1007/978-981-15-1884-3_19, 2020.

Nisbet, I. C. T. and LaGoy, P. K: Toxic equivalency factors (TEFs) for polycyclic aromatic hydrocarbons (PAHs), Regul. Toxicol. Pharmacol., 16, 290–300, https://doi.org/10.1016/0273-2300(92)90009-X, 1992.

Pinkus, A.: Approximation theory of the MLP model in neural networks, Acta Numer., 8, 143–195, https://doi.org/10.1017/S0962492900002919, 1999.

Prodhan, F. A., Zhang, J., Hasan, S. S., Pangali Sharma, T. P., and Mohana, H. P.: A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions, Environ. Modell. Softw., 149, 105327, https://doi.org/10.1016/j.envsoft.2022.105327, 2022a.

Prodhan, F. A., Zhang, J., Pangali Sharma, T. P., Nanzad, L., Zhang, D., Seka, A. M., Ahmed, N., Hasan, S. S., Hoque, M. Z., and Mohana, H. P.: Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach, Sci. Total Environ., 807, 151029, https://doi.org/10.1016/j.scitotenv.2021.151029, 2022b.

Ren, X., Mi, Z., Cai, T., Nolte, C. G., and Georgopoulos, P. G.: Flexible Bayesian Ensemble Machine Learning Framework for Predicting Local Ozone Concentrations, Environ. Sci. Technol., 56, 3871–3883, https://doi.org/10.1021/acs.est.1c04076, 2022.

Saha, P. K., Presto, A. A., Hankey, S., Marshall, J. D., and Robinson, A. L.: Cooking emissions are a major source of racial-ethnic air pollution exposure disparities in the United States, Environ. Res. Lett., 19, 014084, https://doi.org/10.1088/1748-9326/ad1721, 2024.

Saito, E., Tanaka, N., Miyazaki, A., and Tsuzaki, M.: Concentration and particle size distribution of polycyclic aromatic hydro-

carbons formed by thermal cooking, Food Chem., 153, 285–291, https://doi.org/10.1016/j.foodchem.2013.12.055, 2014.

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., and Santos, J.: Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches [Research Frontier], IEEE Comput. Intell. M., 13, 59–76, https://doi.org/10.1109/MCI.2018.2866730, 2018.

Segal, M. R.: Machine Learning Benchmarks and Random Forest Regression, Center for Bioinformatics & Molecular Biostatistics, https://escholarship.org/uc/item/35x3v9t4 (last access: 11 February 2025), 2004.

Shanxi Provincial Government: Air Pollution Prevention and Control Regulations for Shanxi Province, https://www.shaanxi.gov.cn/zfxxgk/zcwjk/dfxfg/202402/t20240226_2320563.html (last access: 11 February 2025), 2017.

Shi, L.: Changes of Industrial Structure and Economic Growth in Coastal Regions of China: A Threshold Panel Model Based Study, J. Coast. Res., 107, 278–282, https://doi.org/10.2112/JCR-SI107-068.1, 2020.

Shi, S., Chen, R., Wang, P., Zhang, H., Kan, H., and Meng, X.: An Ensemble Machine Learning Model to Enhance Extrapolation Ability of Predicting Coarse Particulate Matter with High Resolutions in China, Environ. Sci. Technol., 58, 19325–19337, https://doi.org/10.1021/acs.est.4c08610, 2024.

Slawski, M. and Hein, M.: Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization, Electron. J. Stat., 7, 3004–3056, https://doi.org/10.1214/13-EJS868, 2013.

Sree Dhevi, A. T.: Imputing missing values using Inverse Distance Weighted Interpolation for time series data, in: 2014 Sixth International Conference on Advanced Computing (ICoAC), 2014 Sixth International Conference on Advanced Computing (ICoAC), Chennai, India, 17–19 December 2014, 255–259, https://doi.org/10.1109/ICoAC.2014.7229721, 2014.

MEE (Ministry of Ecological Environment): Emission standards of catering oil fume, Ministry of Ecological Environment, China Environmental Science Press, GB 18483-2001, https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/dqhjbh/dqgdwrywrwpfbz/200201/t20020101_67405.htm (last access: 11 February 2025), 2001.

Tian, L. and Shen, X.: Spatial patterns and their influencing factors for China's catering industry, Humanit. Soc. Sci. Commun., 11, 1–12, https://doi.org/10.1057/s41599-024-03580-7, 2024.

Uyanık, G. K. and Güler, N.: A Study on Multiple Linear Regression Analysis, Procd. Soc. Behv., 106, 234–240, https://doi.org/10.1016/j.sbspro.2013.12.027, 2013.

Wang, H., Xiang, Z., Wang, L., Jing, S., Lou, S., Tao, S., Liu, J., Yu, M., Li, L., Lin, L., Chen, Y., Wiedensohler, A., and Chen, C.: Emissions of volatile organic compounds (VOCs) from cooking and their speciation: A case study for Shanghai with implications for China, Sci. Total Environ., 621, 1300–1309, https://doi.org/10.1016/j.scitotenv.2017.10.098, 2018a.

Wang, H., Jing, S., Lou, S., Tao, S., Qiao, L., Li, L., Huang, C., Lin, L., and Cheng, C.: Estimation of Fine Particle ($PM_{2.5}$) Emission Inventory from Cooking: Case Study for Shanghai, Environm. Sci., 39, 1971–1977, https://doi.org/10.13227/j.hjkx.201708228, 2018b.

Wang, H., Yang, J., Chen, G., Ren, C., and Zhang, J.: Machine learning applications on air temperature prediction in the urban canopy layer: A critical review of 2011–2022, Urban Climate, 49, 101499, https://doi.org/10.1016/j.uclim.2023.101499, 2023.

Wang, L., Xiang, Z., Stevanovic, S., Ristovski, Z., Salimi, F., Gao, J., Wang, H., and Li, L.: Role of Chinese cooking emissions on ambient air quality and human health, Sci. Total Environ., 589, 173–181, https://doi.org/10.1016/j.scitotenv.2017.02.124, 2017.

Wang, T., Li, B., Liao, H., Li, Y.: Spatiotemporal distribution of atmospheric polycyclic aromatic hydrocarbon emissions during 2013–2017 in mainland China, Sci. Total Environ., 789, 148003, https://doi.org/10.1016/j.scitotenv.2021.148003, 2021.

Wu, H., Wang, L., Ling, X., Cui, L., Sun, R., and Jiang, N.: Spatiotemporal reconstruction of global ocean surface pCO_2 based on optimized random forest, Sci. Total Environ., 912, 169209, https://doi.org/10.1016/j.scitotenv.2023.169209, 2024.

Xiao, Q., Chang, H. H., Geng, G., and Liu, Y.: An Ensemble Machine-Learning Model To Predict Historical $PM_{2.5}$ Concentrations in China from Satellite Data, Environ. Sci. Technol., 52, 13260–13269, https://doi.org/10.1021/acs.est.8b02917, 2018.

Xu, H., Ta, W., Yang, L., Feng, R., He, K., Shen, Z., Meng, Z., Zhang, N., Li, Y., Zhang, Y., Lu, J., Li, X., Qu, L., Ho, S. S. H., and Cao, J.: Characterizations of $PM_{2.5}$-bound organic compounds and associated potential cancer risks on cooking emissions from dominated types of commercial restaurants in northwestern China, Chemosphere, 261, 127758, https://doi.org/10.1016/j.chemosphere.2020.127758, 2020.

Xu, H., Yu, H., Xu, B., Wang, Z., Wang, F., Wei, Y., Liang, W., Liu, J., Liang, D., Feng, Y., and Shi, G.: Machine learning coupled structure mining method visualizes the impact of multiple drivers on ambient ozone, Commun. Earth Environ., 4, 1–10, https://doi.org/10.1038/s43247-023-00932-0, 2023.

Yang, C., Dong, H., Chen, Y., Wang, Y., Fan, X., Tham, Y. J., Chen, G., Xu, L., Lin, Z., Li, M., Hong, Y., and Chen, J.: Machine Learning Reveals the Parameters Affecting the Gaseous Sulfuric Acid Distribution in a Coastal City: Model Construction and Interpretation, Environ. Sci. Tech. Let., 10, 1045–1051, https://doi.org/10.1021/acs.estlett.3c00170, 2023a.

Yang, C., Dong, H., Chen, Y., Xu, L., Chen, G., Fan, X., Wang, Y., Tham, Y. J., Lin, Z., Li, M., Hong, Y., and Chen, J.: New Insights on the Formation of Nucleation Mode Particles in a Coastal City Based on a Machine Learning Approach, Environ. Sci. Technol., 58, 1187–1198, https://doi.org/10.1021/acs.est.3c07042, 2023b.

Yang, G., Lin, X., Fang, A., and Zhu, H.: Eating Habits and Lifestyles during the Initial Stage of the COVID-19 Lockdown in China: A Cross-Sectional Study, Nutrients, 13, 970, https://doi.org/10.3390/nu13030970, 2021.

Ye, S., Zhang, B., Fu, H., Tian, N., Shang, H., Chen, X., and Wu, S.: Emission of Fine Particles and fine particle-bound polycyclic aromatic hydrocarbons from simulated cooking fumes, Journal of Xiamen University (Natural Science), 52, 824–829, https://www.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDHIS2&filename=XDZK201306018&uniplatform=OVERSEA&v=Ij6nPaxg7WbNUg0SHE7RYOtv-ItAjPoWIA QpZXyZOwYFQqMnejcvZwlueusjucvF (last access: 11 February 2025), 2013.

Ye, X., Wang, X., and Zhang, L.: Diagnosing the Model Bias in Simulating Daily Surface Ozone Variability Using a Machine Learning Method: The Effects of Dry Deposition and Cloud Optical Depth, Environ. Sci. Technol., 56, 16665–16675, https://doi.org/10.1021/acs.est.2c05712, 2022.

Yu, H., Deng, Y., and Xu, S.: Evolutionary Pattern and Effect of Administrative Division Adjustment During Urbanization of China: Empirical Analysis on Multiple Scales, Chinese Geogr. Sci., 28, 758–772, https://doi.org/10.1007/s11769-018-0990-2, 2018.

Yu, Y., Guo, S., Wang, H., Shen, R., Zhu, W., Tan, R., Song, K., Zhang, Z., Li, S., Chen, Y., and Hu, M.: Importance of Semivolatile/Intermediate-Volatility Organic Compounds to Secondary Organic Aerosol Formation from Chinese Domestic Cooking Emissions, Environ. Sci. Tech. Let., 9, 507–512, https://doi.org/10.1021/acs.estlett.2c00207, 2022.

Yuan, X., Chen, B., He, X., Zhang, G., and Zhou, C.: Spatial Differentiation and Influencing Factors of Tertiary Industry in the Pearl River Delta Urban Agglomeration, Land, 13, 172, https://doi.org/10.3390/land13020172, 2024.

Yuan, Y., Zhu, Y., Lin, C.-J., Wang, S., Xie, Y., Li, H., Xing, J., Zhao, B., Zhang, M., and You, Z.: Impact of commercial cooking on urban $PM_{2.5}$ and O_3 with online data-assisted emission inventory, Sci. Total Environ., 873, 162256, https://doi.org/10.1016/j.scitotenv.2023.162256, 2023.

Zhang, J. and Zhao, X.: Using POI and multisource satellite datasets for mainland China's population spatialization and spatiotemporal changes based on regional heterogeneity, Sci. Total Environ., 912, 169499, https://doi.org/10.1016/j.scitotenv.2023.169499, 2024.

Zhang, J., Zheng, Y., Qi, D., Li, R., and Yi, X.: DNN-based prediction model for spatio-temporal data, in: Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL'16: 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, New York, United States, 31 October–3 November 2016, 92, https://doi.org/10.1145/2996913.2997016, 2016.

Zhang, J., Duan, W., Cheng, S., and Wang, C.: A comprehensive evaluation of the atmospheric impacts and health risks of cooking fumes from different cuisines, Atmos. Environ., 338, 120837, https://doi.org/10.1016/j.atmosenv.2024.120837, 2024a.

Zhang, J., Duan, W., Cheng, S., and Wang, C.: A high-resolution $(0.1° \times 0.1°)$ emission inventory for the catering industry based on VOCs and $PM_{2.5}$ emission characteristics of Chinese multi-cuisines, Atmos. Environ., 319, 120314, https://doi.org/10.1016/j.atmosenv.2023.120314, 2024b.

Zhang, L., Nan, Z., Yu, W., Zhao, Y., and Xu, Y.: Comparison of baseline period choices for separating climate and land use/land cover change impacts on watershed hydrology using distributed hydrological models, Sci. Total Environ., 622–623, 1016–1028, https://doi.org/10.1016/j.scitotenv.2017.12.055, 2018.

Zhang, Q., Gangupomu, R. H., Ramirez, D., and Zhu, Y.: Measurement of Ultrafine Particles and Other Air Pollutants Emitted by Cooking Activities, Int. J. Env. Res. Pub. He., 7, 1744–1759, https://doi.org/10.3390/ijerph7041744, 2010.

Zhang, W., Bai, Z., Shi, L., Son, J. H., Li, L., Wang, L., and Chen, J.: Investigating aldehyde and ketone compounds produced from indoor cooking emissions and assessing their health risk to human beings, J. Environ. Sci., 127, 389–398, https://doi.org/10.1016/j.jes.2022.05.033, 2023.

Zhang, Z., Zhu, W., Hu, M., Wang, H., Chen, Z., Shen, R., Yu, Y., Tan, R., and Guo, S.: Secondary Organic Aerosol from Typical Chinese Domestic Cooking Emissions, Environ. Sci. Tech. Let., 8, 24–31, https://doi.org/10.1021/acs.estlett.0c00754, 2021.

Zhao, L.: Legal Control of Soot Nuisance in Urban Catering Trade, Master thesis, Wuhan University, https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD9904&filename=2004113402.nh (last access: 11 February 2025), 2004.

Zhao, X., Lin, W., Cen, S., Zhu, H., Duan, M., Li, W., and Zhu, S.: The online-to-offline (O2O) food delivery industry and its recent development in China, Eur. J. Clin. Nutr., 75, 232–237, https://doi.org/10.1038/s41430-020-00842-w, 2021.

Zhao, Y. and Zhao, B.: Emissions of air pollutants from Chinese cooking: A literature review, Build. Simul.-China, 11, 977–995, https://doi.org/10.1007/s12273-018-0456-6, 2018.

Zheng, L., Lin, R., Wang, X., and Chen, W.: The Development and Application of Machine Learning in Atmos. Environ. Studies, Remote Sensing, 13, 4839, https://doi.org/10.3390/rs13234839, 2021.

Zheng, H., Chang, X., Wang, S., Li, S., Yin, D., Zhao, B., Huang, G., Huang, L., Jiang, Y., Dong, Z., He, Y., Huang, C., and Xing, J.: Trends of Full-Volatility Organic Emissions in China from 2005 to 2019 and Their Organic Aerosol Formation Potentials, Environ. Sci. Technol. Lett., 10, 137–144, https://doi.org/10.1021/acs.estlett.2c00944, 2023.

Zhu, J.-J., Yang, M., and Ren, Z. J.: Machine Learning in Environmental Research: Common Pitfalls and Best Practices, Environ. Sci. Technol., 57, 17671–17689, https://doi.org/10.1021/acs.est.3c00026, 2023.

Zhu, Q., Laughner, J. L., and Cohen, R. C.: Combining Machine Learning and Satellite Observations to Predict Spatial and Temporal Variation of near Surface OH in North American Cities, Environ. Sci. Technol., 56, 7362–7371, https://doi.org/10.1021/acs.est.1c05636, 2022.

Zhu, X.: Understanding China's Growth: Past, Present, and Future, J. Econ. Perspect., 26, 103–124, https://doi.org/10.1257/jep.26.4.103, 2012.