Developing a novel hybrid model for the estimation of surface 8-h ozone (O$_3$) across the remote Tibetan Plateau during 2005-2018

Rui Li$^a$, Yilong Zhao$^a$, Ya Meng$^a$, Wenhui Zhou$^a$, Ziyu Zhang$^a$, Hongbo Fu$^{a,c}$*

$^a$Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention, Department of Environmental Science & Engineering, Institute of Atmospheric Sciences, Fudan University, Shanghai, 200433, P.R. China

$^b$Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science and Technology, Nanjing 210044, P.R. China

$^c$Shanghai Institute of Pollution Control and Ecological Security, Shanghai 200092, P.R. China

Corresponding author

fuhb@fudan.edu.cn

Abstract

We developed a two-stage model named random forest-generalized additive model (RF-GAM) based on satellite data, meteorological factors, and other geographical covariates to predict the surface 8-h O$_3$ concentration across the remote Tibetan Plateau. The 10-fold cross-validation result suggested that RF-GAM showed the excellent performance with the highest $R^2$ value (0.76) and lowest root mean square error (RMSE) (14.41 μg/m$^3$) compared with other seven machine learning models. The predictive performance of RF-GAM model showed significantly seasonal discrepancy with the highest $R^2$ value observed in summer (0.74), followed by winter (0.69) and autumn (0.67), and the lowest one in spring (0.64). Additionally, the unlearning ground-observed O$_3$ data collected from open websites were applied to test the transferring ability of the novel model, and confirmed...
that the model was robust to predict the surface 8-h O3 concentration during other periods ($R^2 = 0.67$, RMSE = 25.68 μg/m$^3$). RF-GAM was then used to predict the daily 8-h O3 level over Tibetan Plateau during 2005-2018 for the first time. It was found that the estimated O3 concentration displayed a slow increase from $64.74 \pm 8.30 \mu g/m^3$ to $66.45 \pm 8.67 \mu g/m^3$ from 2005 to 2015, whereas it decreased from the peak to $65.87 \pm 8.52 \mu g/m^3$ during 2015-2018. Besides, the estimated 8-h O3 concentrations exhibited notably spatial variation with the highest values in some cities of North Tibetan Plateau such as Huangnan (73.48 ± 4.53 μg/m$^3$) and Hainan (72.24 ± 5.34 μg/m$^3$), followed by the cities in the central region including Lhasa (65.99 ± 7.24 μg/m$^3$) and Shigatse (65.15 ± 6.14 μg/m$^3$), and the lowest one in a city of Southeast Tibetan Plateau named Aba (55.17 ± 12.77 μg/m$^3$). Based on the 8-h O3 critical value (100 μg/m$^3$) scheduled by World Health Organization (WHO), we further estimated the annually mean nonattainment days over Tibetan Plateau. It should be noted that most of the cities in Tibetan Plateau shared with the excellent air quality, while several cities (e.g., Huangnan, Haidong, and Guoluo) still suffered from more than 40 nonattainment days each year, which should be paid more attention to alleviate local O3 pollution. The result shown herein confirms the novel hybrid model improves the prediction accuracy and can be applied to assess the potential health risk, particularly in the remote regions with sparse monitoring sites.

**Keywords:** Surface O3 level; satellite data; random forest; generalized additive model; Tibetan Plateau

### 1. Introduction

Along with the rapid economic development and urbanization, the anthropogenic emissions of nitrogen oxides (NOx) and volatile organic compounds (VOCs) displayed high-speed growth. The chemical reaction between NOx and VOCs in the presence of sunlight was beneficial to the ambient...
ozone (O\textsubscript{3}) formation (Wang et al., 2019; Wang et al., 2017). As a strong oxidant, ambient O\textsubscript{3} could play a negative role on human health through aggravating the cardiovascular and respiratory function (Ghude et al., 2016; Marco, 2017; Yin et al., 2017a). Apart from the effect on human health, O\textsubscript{3} also posed a great threat on vegetation growth (Emerson, 2017; Feng et al., 2015; Qian et al., 2018; Feng et al., 2019). Moreover, the tropospheric O\textsubscript{3} can perturb the radiative energy budget of the earth-atmosphere system as the third most important greenhouse gas next to carbon dioxide (CO\textsubscript{2}) and methane (CH\textsubscript{4}), thereby changing the global climate (Bornman et al., 2019; Fu et al., 2019; Wang et al., 2019). Recently, the particulate matter less than 2.5 μm (PM\textsubscript{2.5}) concentration showed the persistent decrease, while the O\textsubscript{3} issue has been increasingly prominent in China (Li et al., 2017b; Li et al., 2019b). Therefore, it was critical to accurately reveal the spatiotemporal variation of O\textsubscript{3} pollution and assess its health risk in China.

A growing body of studies began to investigate the spatiotemporal variation of O\textsubscript{3} level worldwide. Wang et al. (2014b) demonstrated that the 8-h O\textsubscript{3} concentration in nearly all of the provincial cities experienced the remarkable increase during 2013-2014. Following the work, Li et al. (2017) reported that the annually mean O\textsubscript{3} concentration over China increased by 9.18% during 2014-2016. In other Asian countries except China, Vellingiri et al. (2015) performed long-term observation and found that the O\textsubscript{3} concentration in Seoul, South Korea displayed gradual increase in the past decades. In the Southeast United States, Li et al. (2018) observed that the surface O\textsubscript{3} concentration also displayed the gradual decrease in the recent ten years. Although the number of ground-level monitoring sites have been increasing globally, the limited monitoring sites still cannot accurately reflect the fine-scale O\textsubscript{3} pollution status because each site shows small spatial representativeness (0.25-16.25 km\textsuperscript{2}) (Shi et al., 2018). Furthermore, the number of monitoring
sites in many countries (e.g., China and the United States) displays uneven distribution characteristic at the spatial scale. In China, most of these sites focus on North China Plain (NCP) and Yangtze River Delta (YRD), while West China extremely lacks of the ground-level O$_3$ data, which often increases the uncertainty of health assessment. Therefore, many studies used various models to estimate the O$_3$ concentration without monitoring sites. Chemical transport models (CTMs) were often considered as the typical methods to predict the surface O$_3$ level. Zhang et al. (2011) employed the Geos-Chem model to simulate the surface O$_3$ concentration over the United States, suggesting that the model could capture the spatiotemporal variation of surface O$_3$ concentration at a large spatial scale. Later on, Wang et al. (2016) developed a hybrid model named land use regression (LUR) coupled with CTMs to predict the surface O$_3$ concentration in the Los Angeles Basin, California. In recent years, these methods were also applied to estimate the surface O$_3$ level over China. Liu et al. (2018) used Community Multiscale Air Quality (CMAQ) model to simulate the nationwide O$_3$ concentration over China in 2015. Nonetheless, the high-resolution O$_3$ prediction using CTMs might be widely deviated from the measured value owing to the imperfect knowledge about the chemical mechanism and the higher uncertainty of emission inventory. Moreover, the continuous emission data of NO$_x$ and VOCs were not always open access, which restricted the long-term estimation of surface O$_3$ concentration using CTMs.

Fortunately, the daily satellite data enables the fine-scale estimation of O$_3$ level at a regional scale due to broad spatial coverage and high temporal resolution (McPeters et al., 2015). Shen et al. (2019) confirmed that satellite retrieved O$_3$ column amount can accurately reflect the spatiotemporal distribution of surface O$_3$ level. Therefore, some studies tried to use traditional
statistical models coupled with high-resolution satellite data to estimate the ambient O$_3$ level. Fioletov et al. (2002) used the satellite measurement to investigate the global distribution of O$_3$ concentration based on simple linear model. Recently, Kim et al. (2018) employed the integrated empirical geographic regression method to predict the long-term (1979-2015) variation of ambient O$_3$ concentration over United States based on O$_3$ column amount data. Although the statistical modelling of ambient O$_3$ concentration is widespread all around the world, most of these traditional statistical modelling only utilized the linear model to predict the ambient O$_3$ concentration, which generally decreased the prediction performance because the nonlinearity and high-order interactions between O$_3$ and predictors cannot be managed by a simple linear model.

As an extension of traditional statistical model, machine learning methods have been widely applied to estimate the pollutant level in recent years because of their excellent predictive performances. Among these machine learning algorithms, decision tree models such as random forest (RF) and extreme gradient boosting (XGBoost) strike a perfect balance between prediction performance and computing cost. Furthermore, decision tree models can obtain the contribution of each predictor to air pollutant, which was beneficial to the parameter adaption and model optimization. Chen et al. (2018b) firstly employed RF model to simulate the PM$_{2.5}$ level in China since 2005. Following this work, we recently used the XGBoost model to estimate the 8-h O$_3$ concentration in Hainan Island for the first time and captured the moderate predictive performance ($R^2 = 0.59$) (Li et al., 2020). While decision tree model shows many advantages in predicting pollutant level, the spatiotemporal autocorrelation of pollutant concentration is not concerned by these studies. Li et al. (2019a) confirmed that the prediction
error by decision tree model varied greatly with space and time. Thus, it is imperative to incorporate the spatiotemporal variables into the original model to further improve the performance. To resolve the defects of decision tree models, Zhan et al. (2018) developed a hybrid model named RF-spatiotemporal Kriging (STK) to predict the O$_3$ concentration over China and achieved the better performance (Overall R$^2 = 0.69$, Southwest China R$^2 = 0.66$). Unfortunately, RF-STK model still showed some weaknesses in predicting O$_3$ concentration. First of all, the predictive performance of the STK model was strongly dependent on the number of monitoring sites and their spatial density. The model often showed worse predictive performance in the region with sparse monitoring sites (Gao et al., 2016). Moreover, the ensemble model cannot simulate the O$_3$ level during the periods without ground-measured data. In contrast, generalized additive model (GAM) not only considers the time autocorrelation of O$_3$ concentration, but also shows the better extrapolation ability (Chen et al., 2018a; Ma et al., 2015). Thus, the ensemble model of RF and GAM is proposed to predict the spatiotemporal variation of surface 8-h O$_3$ concentration.

Tibetan Plateau, the highest plateau around the world, shows the higher surface solar radiation compared with the region outside the plateau. It was well documented that high solar radiation is beneficial to generate large amount of OH radical, resulting in the O$_3$ formation via the reaction of VOC and OH radical (Ou et al., 2015). While the total O$_3$ column amount in Tibetan Plateau displayed the slight decrease since 1990s, the convergent airflow formed by subtropical anticyclones could bring ozone-rich air surrounding the plateau to the low atmosphere (Lin et al., 2008), thereby leading to the higher surface O$_3$ concentration over the plateau. Most studies focused on the stratosphere-troposphere transport of O$_3$ in Tibetan Plateau,
whereas limited effort was spared to investigate ground-level O$_3$ level over this region. To date, only several studies concerned about the spatiotemporal variation of surface O$_3$ concentration in this region (Chen et al., 2019; Shen et al., 2014; Yin et al., 2017b). Furthermore, some of these field-observation studies only used the limited monitoring sites to reveal the spatiotemporal variation of O$_3$ concentration, while they cannot clarify the real O$_3$ status in many regions without monitoring sites (e.g., Northern part of Tibetan Plateau). Apart from these field measurements, Liu et al. (2018) (R = 0.60) and Zhan et al. (2018) ($R^2 = 0.66$) used CTM and machine learning model to simulate the surface O$_3$ concentration over China in 2015, respectively. Both of these studies included the predicted O$_3$ level in Tibetan Plateau. Although they have finished the pioneering work, the predictive performances of both studies were not very excellent. Therefore, it was imperative to develop a higher quality model to enhance the modelling accuracy.

Here, we developed a new hybrid method (RF-GAM) model integrating satellite data, meteorological factors, and geographical variables to simulate the gridded 8-h O$_3$ concentration over Tibetan Plateau for the first time. Based on the estimated surface O$_3$ concentration, we clarified the long-term variation (2005-2018) of surface O$_3$ concentration and quantified the key factors for the annual trend. Filling the gap of statistical estimation 8-h O$_3$ level in a remote region, this study provides useful datasets for epidemiological studies and air quality management.

2. Materials and methods

2.1 Study area

Tibetan Plateau is located in Southwest China ranging from 26.00 to 39.58°N and from...
73.33 to 104.78°E. Tibetan Plateau is surrounded by Taklamakan Desert to north, Sichuan Basin to southeast. The land area of Tibetan Plateau reaches 2.50 million km² (Chan et al., 2006). Based on the air circulation pattern, Tibetan Plateau can be roughly classified into the monsoon-influenced region and the westerly-wind influenced region (Wang et al., 2014a). The annually mean air temperature in most regions are below 0°C. The annually mean rainfall amount in Tibetan Plateau ranges from 50 to 2000 mm. The terrain conditions are complex and the higher altitude focused on the central region. Tibetan Plateau is generally treated as the remote region lack of anthropogenic activity and most of the residents focus on southeast and south part of Tibetan Plateau. Tibetan Plateau is consisted of 19 prefecture-level cities and their names and corresponding geographical locations are shown in Fig. 1 and Fig. S1.

2.2 Data preparation

2.2.1 Ground-level 8-h O₃ concentration

The daily 8-h O₃ data in 37 monitoring sites over Tibetan Plateau from May 13th, 2014 to December 31th, 2018 were collected from the national air quality monitoring network. The O₃ levels in all of these sites were determined using an ultraviolet-spectrophotometry method. The highest 8-h moving average O₃ concentration each day was calculated as the daily 8-h O₃ level after data quality assurance. The data quality of all the monitoring sites were assured on the basis of the HJ 630-2011 specifications. The data with no more than two consecutive hourly measurement missing in each day was treated as the valid data.

2.2.2 Satellite-retrieved O₃ column amount

The O₃ column amount (DU) during 2005-2018 were downloaded from the Ozone Monitoring Instrument-O₃ (OMI-O₃) level-3 data with a 0.25° spatial resolution from the website of National...
The OMI-O₃ product shows global coverage and traverses the earth once a day. The O₃ column amount with cloud radiance fraction > 0.5, terrain reflectivity > 30%, and solar zenith angles > 85° should be removed. In addition, the cross-track pixels significantly influenced by row anomaly should be deleted.

2.2.3 Meteorological data and geographical covariates

The daily meteorological data were obtained from ERA-Interim datasets with 0.125° resolution. These meteorological data were consisted of 2 meter dewpoint temperature (d2m), 2 meter temperature (t2m), 10 meter U wind component (u10), 10 meter V wind component (v10), boundary layer height (blh), sunshine duration (sund), surface pressure (sp), and total precipitation (tp). The 30 m-resolution elevation data (DEM) was downloaded from China Resource and Environmental Science Data Center (CRESDC). The data of gross domestic production (GDP) and population density with 1 km resolution were also extracted from CRESDC. Population density and GDP in 2005, 2010, and 2015 were integrated into the model to predict the surface 8-h O₃ concentration over Tibetan Plateau because these data were available each five years. Additionally, the land use data of 30 m resolution (e.g., waters, grassland, urban, forest) were also extracted from CRESDC. At last, the latitude, longitude, and time were also incorporated into the model.

All of the explanatory variables collected were resampled to 0.25° × 0.25° grids to predict the O₃ level. The original meteorological data with 0.125° resolution were resampled to 0.25° grid. The land use area, elevation, GDP and population density in each grid were calculated using spatial clipping. Lastly, all of the predictors were integrated into an intact table to train the model.

2.3 Model development and assessment

The RF-GAM model was regarded as the hybrid model of RF and GAM. The RF-GAM model
is a two-stage model that the prediction error estimated by the RF model was then simulated by GAM. The prediction results of RF and GAM were summed as the final result of RF-GAM model (Fig. 2). The detailed equation is as follows:

\[ Z(s,t) = P(s,t) + E(s,t) \quad (1) \]

where \( Z(s,t) \) is the estimated 8-h \( \text{O}_3 \) level at the location \( s \) and time \( t \); \( P(s,t) \) represents the 8-h \( \text{O}_3 \) concentration predicted by the RF model; \( E(s,t) \) denotes the prediction error by GAM.

In the RF model, a large number of decision trees were planted based on the bootstrap sampling method. At each node of the decision tree, a random sample of all predictors was applied to determine the best split among them. Following the procedure, a simple majority vote was employed to predict the 8-h \( \text{O}_3 \) level. The RF model avoided priori linear assumption of \( \text{O}_3 \) concentration and predictors, which was often not in good agreement with actual state. The RF model has two key parameters including \( n_{\text{tree}} \) (the number of trees grown) and \( m_{\text{try}} \) (the number of explanatory variables sampled for splitting at each node). The prediction performance of the RF model was strongly dependent on the two parameters. The optimal \( n_{\text{tree}} \) and \( m_{\text{try}} \) were determined based on the least out-of-bag (OOB) errors. Besides, the backward variable selection method was performed on the RF submodel to achieve the better performance. At each step of the predictor selection, the variable with the least important value was excluded from the next step. This one-variable-at-a-time exclusion method was repeated until only two explanatory variables remained in the submodel. Finally, all of the selected variables except the area of waters were integrated into the model to achieve the best prediction performance. The detailed RF model is as follows:

\[ \text{O}_3 = \text{O}_3 \text{ column} + \text{Elevation} + \text{Agr} + \text{Urban} + \text{Forest} + \text{GDP} + \text{Grassland} + \text{Population} + \text{Prec} + T + \text{WS} + P + tsun + RH \quad (2) \]
where \( O_3 \) denotes the observed 8-h \( O_3 \) level in the monitoring site; the \( O_3 \) column represents the \( O_3 \) column amount in the corresponding grid; Elevation denotes the corresponding elevation of the site; Agr, Urban, Forest, Grassland are the agricultural land, urban land, forest land, and the grassland, respectively. Population represents the population density in the corresponding site. Prec, T, WS, P, tsun, and RH are precipitation, air temperature, wind speed, air pressure, sunshine duration, and relative humidity, respectively. Additionally, other five models including RF, generalized regression neutral network (GRNN), backward propagation neural network (BPNN), Elman neural network (ElmanNN), and extreme learning machine (ELM) also used the backward variable selection method. The \( R^2 \) value was treated as an important parameter to add or reduce the variable. The variable should be removed when the \( R^2 \) value of the submodel showed the remarkable decrease with the integration of this variable. Lastly, the optimal variable group was applied to establish the submodel.

Following the RF submodel, the prediction error estimated by the RF submodel was further modelled by the GAM. GAM could reflect the time autocorrelation of predictive error of RF model, and thus the ensemble model of RF and GAM might decrease the modelling error of one-stage model. All of the variables were incorporated into the models to establish the second-stage model, and the backward variable selection was also used to determine the optimal variable group.

The 10-fold cross-validation (CV) technique was employed to evaluate the predictive performances for all of the machine learning models. All of the training data set were randomly classified into 10 subsets uniformly. In each round of validation, nine subsets were used to train and the remaining subset was applied to test the model performance. The process was repeated 10 times until every subset has been tested. Some statistical indicators including \( R^2 \), Root Mean Square Error
(RMSE), Mean Prediction Error (MPE) and the slope were calculated to assess the model performance. The optimal model with the best performance was used to estimate the 8-h O₃ concentration in the past decades.

3. Results and discussion

3.1 The validation of model performance

Figure 3 shows the density scatterplots of the fitting and 10-fold cross-validation results for eight machine learning models for China. The 10-fold cross-validation $R^2$ values followed the order of RF-GAM ($R^2 = 0.76$) > RF-STK ($R^2 = 0.63$) > RF ($R^2 = 0.55$) > GRNN ($R^2 = 0.53$) > BPNN ($R^2 = 0.50$) > XGBoost ($R^2 = 0.48$) > ElmanNN ($R^2 = 0.47$) > ELM ($R^2 = 0.32$). The RMSE values of RF-GAM, RF-STK, RF, GRNN, XGBoost, BPNN, ElmanNN, and ELM were 14.41, 17.79, 19.13, 19.41, 20.73, 20.61, and 23.36 μg/m³, respectively. MPE showed the similar characteristic with RMSE in the order of RF-GAM (10.97 μg/m³) < RF-STK (13.48 μg/m³) < RF (14.71 μg/m³) < GRNN (14.89 μg/m³) < BPNN (15.43 μg/m³) < ElmanNN (15.75 μg/m³) < XGBoost (15.80 μg/m³) < ELM (18.23 μg/m³). Besides, the slope of the RF-GAM model was closer to 1 compared with other models. It was well documented that the RF model generally showed the better performance than other models because this method did not need to define complex relationships between the explanatory variables and the O₃ concentration (e.g., linear or nonlinear). Furthermore, the variable importance indicators calculated by the RF model can help user to distinguish the key variables from noise ones and make full use of the strength of each predictor to assure the model robustness.

Although BPNN, GRNN, XGBoost, ElmanNN, and ELM have been widely applied to estimate the air pollutant concentrations (Chen et al., 2018c; Zang et al., 2018; Zhu et al., 2019), these methods suffered from some weaknesses in predicting the pollutant level. For instance, both of BPNN and...
ElmanNN models could capture the locally optimal solution when the training subsets were integrated into the final model, which decreased the predictive performance of the model (Wang et al., 2015). Moreover, BPNN generally showed slow training speed, especially with the huge training subsets (Li and Park, 2009; Wang et al., 2015). ELM often consumed more computing resource and experienced the over-fitting issue due to the increase of sampling size (Huang et al., 2015; Shao et al., 2015). GRNN method advanced the training speed compared with BPNN model and avoided the locally optimal solution during the modelling process (Zang et al., 2019), whereas the predictive performance is still worse than that of RF model. XGBoost was often considered to be robust in predicting air pollutant level (Li et al., 2020), while the model did not display the excellent performance in the present study. It might be attributable to that the sampling size in the present study was not enough because the model generally showed the better performance with big samples. Moreover, we found that the two-stage model was superior to the one-way model in the predictive performance. The encouraging result suggested that the relationship between the predictors and the 8-h $O_3$ concentration varied with space and time. The two-stage model used the GAM method to further adjust the prediction error of the RF model, and considered the spatiotemporal correlation of predictor error in Tibetan Plateau. Although the STK model incorporated space and time into the model simultaneously, the RF-GAM model outperformed the RF-STK model. It was assumed that the STK model showed the higher uncertainty in predicting the $O_3$ concentration in the region with scarce sampling sites (Gao et al., 2016; Li et al., 2017a). Overall, the ensemble RF-GAM model showed the significant improvement in predictive performance.

The performances of the RF-GAM model in four seasons were also assessed by 10-fold cross-validation (Tab. 1). The predictive performance of the RF-GAM model showed significantly
seasonal difference with the highest $R^2$ value observed in summer (0.74), followed by winter (0.69) and autumn (0.67), and the lowest one in spring (0.64). However, both of RMSE and MPE displayed different seasonal characteristics with the $R^2$ value. Both of RMSE and MPE for RF-GAM followed the order of spring ($15.32$ and $11.94 \, \mu g/m^3$) > summer ($15.13$ and $11.75 \, \mu g/m^3$) > winter ($14.58$ and $11.44 \, \mu g/m^3$) > autumn ($13.23$ and $10.52 \, \mu g/m^3$). The lowest $R^2$ value in spring might be caused by multiple $O_3$ sources and complicate $O_3$ formation mechanisms. The large estimation errors (e.g., RMSE and MPE) in spring and summer were attributable to the high 8-h $O_3$ concentration in these seasons, while the low prediction error observed in autumn was contributed by the low $O_3$ level.

Apart from the seasonal variation, we also investigated the spatial variability of the predictive accuracy for RF-GAM model. Tibetan Plateau was classified into five provinces and then the predictive performance of RF-GAM model in each province was calculated. Among the five provinces, Gansu province displayed the highest $R^2$ value (0.74), followed by Sichuan province (0.71), Qinghai province (0.70), Tibet autonomous region (0.69), and Yunnan province (0.54) (Tab. 2). The result shown herein was not in agreement with the previous studies by Geng et al. (2018), who confirmed that the predictive performance of machine learning model was positively associated with the sampling size. It was assumed that the spatial distribution of the sampling sites in Tibet was uneven and the sampling density is low, though Tibet possessed the maximum monitoring sites compared with other provinces. The prediction error (RMSE and MPE) did not exhibit the same characteristics with the $R^2$ value. The higher RMSE and MPE focused on Tibet autonomous region ($14.81$ and $11.24 \, \mu g/m^3$) and Qinghai province ($14.83$ and $11.33 \, \mu g/m^3$) due to the higher values of $blh$ and $sund$. The lowest values of RMSE and MPE could be observed in Yunnan province, which was contributed by the higher rainfall amount.
Although 10-fold cross-validation verified that the RF-GAM model showed the better predictive performance in estimating the surface 8-h O\textsubscript{3} concentration, the test method cannot validate the transferring ability of the final model. The monitoring site in Tibetan Plateau before May, 2014 is very limited, and only the daily 8-h O\textsubscript{3} data in Lhasa from the open website (https://www.aqistudy.cn/historydata/) was available to compare with the simulated data. As depicted in Fig. 4, the R\textsuperscript{2} value of unlearning 8-h O\textsubscript{3} level against predicted 8-h O\textsubscript{3} concentration reached 0.67, which was slightly lower than that of the 10-fold cross-validation R\textsuperscript{2} value. Overall, the extrapolation ability of the RF-GAM model is satisfactory, and thus it was supposed that the model could be applied to estimate the O\textsubscript{3} concentration in other years. Both of RMSE and MPE for the unlearning 8-h O\textsubscript{3} level against the predicted 8-h O\textsubscript{3} concentration were significantly higher than those of the 10-fold cross-validation. It was supposed that Lhasa showed the higher surface 8-h O\textsubscript{3} concentration over Tibetan Plateau.

3.2 Variable importance

The results of variable importance for key variables are depicted in Fig. 5. In the final RF-GAM model, it was found that time was the dominant factor for the 8-h O\textsubscript{3} concentration in Tibetan Plateau, indicating that the ambient O\textsubscript{3} concentration displayed significantly temporal correlation. Following the time, meteorological factors served as the main factors for the O\textsubscript{3} pollution in the remote region. The sum of sund, sp, d2m, t2m, and tp occupied 34.43\% of the overall variable importance. Among others, sund was considered to be the most important meteorological factors for the O\textsubscript{3} pollution. It was assumed that strong solar radiation and long duration of sunshine favored the photochemical generation of ambient O\textsubscript{3} (Malik and Tauler, 2015; Stähle et al., 2018). Tan et al. (2018) demonstrated that the chemical reaction between NO\textsubscript{x} and VOCs was strongly dependent on the
sunlight. Besides, the atmospheric pressure (sp) was also treated as a major driver for the O₃ pollution over Tibetan Plateau. Santurtún et al. (2015) have demonstrated that sp was closely linked to the atmospheric circulation and synoptic scale meteorological pattern, which could influence the long-range transport of ambient O₃. Apart from sun and sp, d2m and t2m played significant role on the O₃ pollution, which was in consistent with many previous studies (Zhan et al., 2018). Zhan et al. (2018) observed that cold temperature was not favorable to the O₃ formation. d2m can affect the surface O₃ pollution through two aspects. On the one hand, RH affected heterogeneous reactions of O₃ and particles (e.g., soot, mineral) (He et al., 2017; He and Zhang, 2019). On the other hand, high RH could increase the soil moisture and evaporation, and thus the water-stressed plants tended to emit more biogenic isoprene, thereby promoting the elevation of O₃ concentration (Zhang and Wang, 2016). It should be noted that the effect of precipitation on O₃ pollution was relatively weaker than those of other meteorological factors. Zhan et al. (2018) also found the similar result and believed that rain scavenging served as the key pathway for the O₃ removal only when O₃ pollution was very serious. The power of O₃ column amount on surface O₃ concentration seemed to be lower than those of most meteorological factors, suggesting that vertical transport of ambient O₃ was complex. Although socioeconomic factors and land use types were not dominant factors for the O₃ pollution in Tibetan Plateau, they still cannot be ignored in the present study because the predictive performance would worsen if these variables were excluded from the model. It was widely acknowledged that the emissions of NOₓ and VOCs focused on the developed urban areas with high population density especially in the remote plateau (Zhang et al., 2007; Zheng et al., 2017). Compared with the urban land, the grassland played more important role on the O₃ pollution in Tibetan Plateau. It was thus supposed that the grassland was widely distributed on Tibetan Plateau,
which could release a large amount of biogenic volatile organic compounds (BVOCs) (Fang et al., 2015). It was well known that photochemical reaction of BVOCs and NO\textsubscript{x} in the presence of sunlight was beneficial to the O\textsubscript{3} formation (Calfapietra et al., 2013). Furthermore, Fang et al. (2015) confirmed that the BVOC emission in Tibetan Plateau displayed a remarkable increase in the wet seasons.

3.3 The spatial distribution of estimated 8-h O\textsubscript{3} concentration over Tibetan Plateau

Figure 6 depicts the spatial distribution of the 8-h O\textsubscript{3} level estimated by the novel RF-GAM model. The spatial distribution pattern modelled by the RF-GAM model showed the similar characteristic with the result simulated by previous studies except North Tibetan Plateau (Liu et al., 2018). The estimated 8-h O\textsubscript{3} concentration displayed the highest value in some cities of North Tibetan Plateau such as Huangnan (73.48±4.53 μg/m\textsuperscript{3}) and Hainan (72.24±5.34 μg/m\textsuperscript{3}), followed by the cities in the central region including Lhasa (65.99±7.24 μg/m\textsuperscript{3}) and Shigatse (65.15±6.14 μg/m\textsuperscript{3}), and the lowest one in some cities of Southeast Tibetan Plateau such as Aba (55.17±12.77 μg/m\textsuperscript{3}). The spatial pattern of 8-h O\textsubscript{3} concentration is highly consistent with the result predicted by Liu et al. (2018) using CMAQ model, while it is not in agreement with the result estimated by Zhan et al. (2018) using RF-STK model. The difference of the present study and Zhan et al. (2018) focuses on the North Tibetan Plateau, which lacks of monitoring site and remains the higher uncertainty.

Firstly, it might be contributed by the weakness of RF-STK mentioned above. Moreover, Zhan et al. (2018) only used the ground-level measured data in 2015 to establish the model and the data in new sites since 2015 were not incorporated into the model, which could increase the model uncertainty (Zhan et al., 2018). As shown in Fig. 6, most of the cities in Qinghai province (e.g., Huangnan, Hainan, and Guoluo) generally showed the higher 8-h O\textsubscript{3} concentration over Tibetan Plateau, which
was in a good agreement with the spatial distribution of O₃ column amount (Fig. S2). Besides, some
cities in Tibet such as Shigatse and Lhasa also showed the higher 8-h O₃ levels. It was supposed that
the precursor (e.g., NOₓ and VOCs) emissions in these regions were significantly higher than those
in other cities of Tibetan Plateau (Fig. S3). Zhang et al. (2007) used the satellite data to observe that
the higher VOCs and NOₓ emission focused on the residential area with high population density in
the remote Tibetan Plateau. Apart from the effect of anthropogenic emission, the meteorological
conditions could be also the important factors for the 8-h O₃ concentration. As shown in Fig. S4-
S10, the higher blh and sp in the Northeast Tibetan Plateau were beneficial to the O₃ formation
through the reaction of VOC and OH radical, leading to the higher 8-h O₃ concentration in these
cities (Ou et al., 2015). In addition, the lower tp occurred in North Tibetan Plateau and Northeast
Tibetan Plateau, both of which were unfavorable to the ambient O₃ removal (Yoo et al., 2014). In
contrast, the higher tp observed in the Southeast Tibetan Plateau resulted in the slight O₃ pollution.

3.4 The temporal variation of the simulated 8-h O₃ concentration over Tibetan Plateau

The annually mean estimated 8-h O₃ concentration in Tibetan Plateau displayed the slow
increase from 64.74 ± 8.30 μg/m³ to 66.45 ± 8.67 μg/m³ 2005 through 2015, whereas it decreased
from the peak to 65.87 ± 8.52 μg/m³ during 2015-2018 (Fig. 7). Based on the Mann-Kendall method
(Fig. 8a), it was found that the surface O₃ concentration exhibited the slight increase as the whole,
while the increase degree was not significant (p > 0.05). Besides, it should be noted that the O₃
concentrations in various regions showed different increase speed. As depicted in Fig. 8b, we found
that the 8-h O₃ concentrations in North, West, and East Tibetan Plateau displayed significant
increase trend by the speed of 1-3 μg/m³ during 2005-2018. The middle region of Tibetan Plateau
showed the moderate increase trend by the speed of 0-1 μg/m³. However, the 8-h O₃ concentration
in Shigatse and Sannan even displayed the decrease trend 2005 through 2018.

Besides, the 8-h O$_3$ concentration in Tibetan Plateau displayed significantly seasonal discrepancy. The estimated 8-h O$_3$ level in Tibetan Plateau followed the order of spring (75.00±8.56 μg/m$^3$) > summer (71.05±11.13 μg/m$^3$) > winter (56.39±7.42 μg/m$^3$) > autumn (56.13±8.27 μg/m$^3$) (Fig. 9 and Tab. 3). The 8-h O$_3$ concentrations in most of prefecture-level cities showed the similarly seasonal characteristics with the overall seasonal variation in Tibetan Plateau. Based on the result summarized in Tab. S1, it was found that the key precursors of ambient O$_3$ (e.g., VOCs, NO$_x$) generally displayed the higher emissions in winter compared with other seasons. However, the seasonal distribution of ambient O$_3$ concentration was not in accordance with the precursor emissions, suggesting that the meteorological factors might play more important roles on ambient O$_3$ concentration. It was well known that the higher air temperature in spring and summer was closely related to the low sp and high sund, both of which were beneficial to the O$_3$ formation (Sitnov et al., 2017). Although summer showed the highest air temperature and the longest sunshine duration, the higher rainfall amount in summer decreased the ambient O$_3$ concentration via wet deposition (Li et al., 2017b; Li et al., 2019b). Moreover, the highest blh occurred in spring, which was favorable to the strong stratosphere-troposphere exchange process in Tibetan Plateau (Skerlak et al., 2014). Therefore, the 8-h O$_3$ concentration in summer and winter were relatively lower than that in spring. Nonetheless, the 8-h O$_3$ levels in Diqing, Sannan, and Nyingchi displayed the highest values in spring (56.38±7.87, 73.90±5.97, and 73.22±2.77 μg/m$^3$), followed by winter (45.88±7.05, 61.71±4.32, and 62.24±3.63 μg/m$^3$) and summer (44.35±5.90, 61.00±5.86, and 59.60±2.33 μg/m$^3$), and the lowest ones in autumn (37.45±5.76, 54.70±3.13, and 53.84±2.06 μg/m$^3$). The lower O$_3$ level in summer than winter was mainly attributable to the higher precipitation observed in the summer
of these cities (Fig. S9). In addition, it should be noted that the NO\textsubscript{x} and VOCs emissions of South Tibetan Plateau (e.g., Sannan) exhibited the higher values in winter compared with other seasons.

3.5 The nonattainment days over Tibetan Plateau during 2005-2018

The annually mean nonattainment days in the 19 prefecture-level cities over Tibetan Plateau are summarized in Tab. 2. 100 µg/m\textsuperscript{3} was regarded as the critical value for the 8-h \textsuperscript{O}_3 level by World Health Organization (WHO). The nonattainment days denoted total days with the 8-h \textsuperscript{O}_3 concentration higher than 100 µg/m\textsuperscript{3}. Although the annually mean 8-h \textsuperscript{O}_3 concentrations in all of the cities over Tibetan Plateau did not exceed the critical value, not all of the regions experienced excellent air quality in the long period (2005-2018). Some cities of Qinghai province including Huangnan, Haidong, and Guoluo suffered from 45, 40, and 40 nonattainment days each year (Fig. 10 and Tab. 4). Besides, some cities in the South Tibetan Plateau such as Shigatse and Sannan also experienced more than 40 nonattainment days each year, suggesting that Tibetan Plateau was still faced of the risk for \textsuperscript{O}_3 pollution. Fortunately, some remote cities such as Ali, Ngari, and Qamdo did not experience the excessive \textsuperscript{O}_3 pollution all the time, which was ascribed to the low precursor emissions and appropriate meteorological conditions. It should be noted that the nonattainment days in the region with high \textsuperscript{O}_3 concentration showed the significantly seasonal difference, whereas the seasonal difference was not remarkable in the city with low \textsuperscript{O}_3 pollution. As shown in Tab. 2, it should be noted that nearly all of the nonattainment days could be detected in spring and summer, which was in good agreement with the \textsuperscript{O}_3 levels in different seasons, indicating that the \textsuperscript{O}_3 pollution issue should be paid more attention in spring and summer.

4. Summary and implication

In the present study, we developed a novel hybrid model (RF-GAM) based on multiple
explanatory variables to estimate the surface 8-h O$_3$ concentration across the remote Tibetan Plateau. The 10-fold cross-validation method demonstrated that RF-GAM achieved excellent performance with the highest $R^2$ value (0.76) and lowest root mean square error (RMSE) (14.41 $\mu$g/m$^3$) compared with other model including RF-STK, RF, BPNN, XGBoost, GRNN, ElmanNN, and ELM models. Moreover, the unlearning ground-measured O$_3$ data validated that the RF-GAM model showed the better extrapolation performance ($R^2$=0.67, RMSE=25.68 $\mu$g/m$^3$). The result of variable importance suggested that time, sund, and sp were key factors for the surface 8-h O$_3$ concentration over Tibetan Plateau. Based on the RF-GAM model, we found that the estimated 8-h O$_3$ concentration exhibited notably spatial variation with the highest value in some cities of North Tibetan Plateau such as Huangnan (73.48±4.53 $\mu$g/m$^3$) and Hainan (72.24±5.34 $\mu$g/m$^3$) and the lowest one in some cities of Southeast Tibetan Plateau such as Aba (55.17±12.77 $\mu$g/m$^3$). Besides, we also found that the O$_3$ level displayed a slow increase from 64.74±8.30 $\mu$g/m$^3$ to 66.45±8.67 $\mu$g/m$^3$ 2005 through 2015, while the O$_3$ concentration decreased to 65.87±8.52 $\mu$g/m$^3$ in 2018. The estimated 8-h O$_3$ level in Tibetan Plateau showed the significantly seasonal discrepancy with the order of spring (75.00±8.56 $\mu$g/m$^3$) > summer (71.05±11.13 $\mu$g/m$^3$) > winter (56.39±7.42 $\mu$g/m$^3$) > autumn (56.13±8.27 $\mu$g/m$^3$). Based on the critical value set by WHO, most of the cities in Tibetan Plateau shared with the excellent air quality, while several cities (e.g., Huangnan, Haidong, and Guoluo) still suffered from more than 40 nonattainment days each year. The RF-GAM model for O$_3$ estimation has several limitations. First of all, the O$_3$ estimation in North Tibetan Plateau might show some uncertainties because the ground-level monitoring site is very scarce, and thus we cannot validate the reliability of predicted value in the region without monitoring site. Secondly, our approach did not include data on emission inventory, or traffic count.
because the continuous emissions of NO\textsubscript{x} and VOCs were not open access. At last, we only focused on the temporal variation of surface O\textsubscript{3} concentration in recent ten years, and the short-term O\textsubscript{3} data cannot reflect the response of O\textsubscript{3} pollution to climate change. In the future work, we should combine more explanatory variables such as long-term NO\textsubscript{x} and VOCs emissions to retrieve the surface O\textsubscript{3} level over Tibetan Plateau in the past decades.

**Author contributions**

This study was conceived by Rui Li and Hongbo Fu. Statistical modelling was performed by Rui Li, Yilong Zhao, Ya Meng, Wenhui Zhou and Ziyu Zhang. Rui Li drafted the paper.

**Acknowledgements**

This work was supported by National Natural Science Foundation of China (Nos. 91744205, 21777025, 21577022, 21177026).

**Reference**


Chen, G., Knibbs, L.D., Zhang, W., Li, S., Cao, W., Guo, J., Ren, H., Wang, B., Wang, H., Williams, G.:
Estimating spatiotemporal distribution of PM$_1$ concentrations in China with satellite remote sensing, meteorology, and land use information. Environ. Pollut. 233, 1086-1094, 


Li, R., Cui, L., Meng, Y., Zhao, Y., Fu, H.: Satellite-based prediction of daily SO2 exposure across China using a high-quality random forest-spatiotemporal Kriging (RF-STK) model for health risk


Shi, X., Zhao, C., Jiang, J.H., Wang, C., Yang, X., Yung, Y.L.: Spatial Representativeness of PM\( _{2.5} \)


Figure and table captions

**Fig. 1** The geographical locations and annually mean 8-h O$_3$ concentrations in the ground-observed sites (red dots) over Tibetan Plateau during 2014-2018. The elevation data are collected from geographical and spatial data cloud at a 30-m spatial resolution.

**Fig. 2** The workflow for predicting the spatiotemporal distributions of 8-h O$_3$ levels.

**Fig. 3** Density scatterplots of model fitting and cross-validation result at a daily level. (a), (b), (c), (d), (e), (f), (g), and (h) represent RF-GAM, RF-STK, RF, GRNN, XGBoost, BPNN, ElmanNN, and ELM models, respectively. The red dotted line denotes the fitting linear regression line. The full names of MPE and RMSE are mean prediction error (µg/m$^3$) and root mean squared prediction error (µg/m$^3$), respectively.

**Fig. 4** The transferring ability validation of RF-GAM method based on the measured daily 8-h O$_3$ concentration during December 2013-May 2014.

**Fig. 5** The variable importance of predictors in the final RF-GAM model.

**Fig. 6** The mean value of estimated 8-h O$_3$ concentration during 2005-2018 over Tibetan Plateau.

**Fig. 7** The inter-annual variation of predicted 8-h O$_3$ level (µg/m$^3$) from 2005 to 2018 across Tibetan Plateau.

**Fig. 8** The trend analysis of predicted 8-h O$_3$ concentration. (a) and (b) represent the result of Mann-Kendall method and discrepancy of estimated O$_3$ level during 2005-2018 across Tibetan Plateau.

**Fig. 9** The seasonal variability of estimated 8-h O$_3$ level across Tibetan Plateau. (a), (b), (c), and (d) represent the predicted 8-h O$_3$ concentrations in spring, summer, autumn, and winter, respectively.

**Fig. 10** The spatial distributions of nonattainment days in Tibetan Plateau during 2005-2018.

**Tab. 1** The $R^2$ values, RMSE, and MPE of RF-GAM in four seasons over Tibetan Plateau.
Tab. 2 The $R^2$ values, RMSE, and MPE of RF-GAM in different provinces over Tibetan Plateau.

Tab. 3 The estimated 8-h $O_3$ concentration in 19 prefecture-level cities over Tibetan Plateau during four seasons including spring, summer, autumn, and winter.

Tab. 4 The mean nonattainment days (8-h $O_3$ level $>100 \mu g/m^3$) in 19 prefecture-level cities over Tibetan Plateau each year.
Fig. 1
Fig. 2
Fig. 3
Fig. 4

\[ Y = 0.73X + 4.89 \]

\[ R^2 = 0.67 \]

\[ \text{RMSE} = 25.68 \]

\[ \text{MPE} = 22.84 \]
Fig. 5
Fig. 6

Average 8-h $O_3$ during 2005-2018
Fig. 7
Fig. 8

(a) Mann-Kendall method

(b) Surface O₃ trend
Fig. 9
Fig. 10
### Tab. 1

<table>
<thead>
<tr>
<th></th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.64</td>
<td>0.74</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>RMSE</td>
<td>15.32</td>
<td>15.13</td>
<td>13.23</td>
<td>14.58</td>
</tr>
<tr>
<td>MPE</td>
<td>11.94</td>
<td>11.75</td>
<td>10.52</td>
<td>11.44</td>
</tr>
</tbody>
</table>
Tab. 2

<table>
<thead>
<tr>
<th></th>
<th>Tibet</th>
<th>Qinghai</th>
<th>Gansu</th>
<th>Sichuan</th>
<th>Yunnan</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.69</td>
<td>0.70</td>
<td>0.74</td>
<td>0.71</td>
<td>0.54</td>
</tr>
<tr>
<td>RMSE</td>
<td>14.81</td>
<td>14.83</td>
<td>13.65</td>
<td>13.23</td>
<td>12.49</td>
</tr>
<tr>
<td>MPE</td>
<td>11.24</td>
<td>11.33</td>
<td>10.88</td>
<td>10.08</td>
<td>10.20</td>
</tr>
<tr>
<td>Province</td>
<td>Spring</td>
<td>Summer</td>
<td>Autumn</td>
<td>Winter</td>
<td>Annual</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>Aba</td>
<td>65.61±14.30</td>
<td>59.46±14.32</td>
<td>45.55±12.03</td>
<td>47.95±10.55</td>
<td>55.17±12.77</td>
</tr>
<tr>
<td>Ngari</td>
<td>71.34±3.12</td>
<td>70.10±3.57</td>
<td>53.14±3.67</td>
<td>51.84±3.69</td>
<td>62.21±3.34</td>
</tr>
<tr>
<td>Qamdo</td>
<td>72.52±4.29</td>
<td>62.74±5.79</td>
<td>52.06±4.01</td>
<td>55.42±3.09</td>
<td>61.10±3.93</td>
</tr>
<tr>
<td>Diqing</td>
<td>56.38±7.87</td>
<td>44.35±5.90</td>
<td>37.45±5.76</td>
<td>45.88±7.05</td>
<td>46.22±6.51</td>
</tr>
<tr>
<td>Gannan</td>
<td>76.77±9.73</td>
<td>73.27±10.67</td>
<td>54.74±8.33</td>
<td>54.72±6.95</td>
<td>65.60±8.91</td>
</tr>
<tr>
<td>Gnzzi</td>
<td>69.38±10.99</td>
<td>61.45±11.58</td>
<td>48.49±8.79</td>
<td>50.94±6.62</td>
<td>58.06±9.48</td>
</tr>
<tr>
<td>Guoluo</td>
<td>80.12±5.12</td>
<td>76.13±5.83</td>
<td>58.86±5.71</td>
<td>57.38±4.66</td>
<td>68.77±5.25</td>
</tr>
<tr>
<td>Haibei</td>
<td>78.18±10.21</td>
<td>78.84±10.31</td>
<td>60.90±9.69</td>
<td>57.48±9.78</td>
<td>69.47±9.99</td>
</tr>
<tr>
<td>Haidong</td>
<td>74.20±10.34</td>
<td>73.70±9.12</td>
<td>53.61±8.11</td>
<td>51.02±9.60</td>
<td>63.84±9.21</td>
</tr>
<tr>
<td>Haixian</td>
<td>83.01±5.36</td>
<td>82.27±5.72</td>
<td>61.57±5.39</td>
<td>58.96±5.44</td>
<td>72.24±5.34</td>
</tr>
<tr>
<td>Haixi</td>
<td>79.39±6.88</td>
<td>79.48±7.79</td>
<td>60.78±7.48</td>
<td>57.71±6.99</td>
<td>69.99±7.24</td>
</tr>
<tr>
<td>Huangnan</td>
<td>85.21±4.98</td>
<td>83.01±4.66</td>
<td>61.95±4.18</td>
<td>60.62±4.49</td>
<td>73.48±4.53</td>
</tr>
<tr>
<td>Lhasa</td>
<td>80.08±9.63</td>
<td>70.13±8.42</td>
<td>55.86±5.78</td>
<td>55.85±5.19</td>
<td>65.99±7.24</td>
</tr>
<tr>
<td>Nagqu</td>
<td>74.59±5.13</td>
<td>70.46±6.69</td>
<td>54.60±5.16</td>
<td>53.53±4.83</td>
<td>63.83±5.23</td>
</tr>
<tr>
<td>Shigatse</td>
<td>77.31±8.62</td>
<td>69.66±7.69</td>
<td>55.93±4.58</td>
<td>55.57±4.72</td>
<td>65.15±6.14</td>
</tr>
<tr>
<td>Sannan</td>
<td>73.90±5.97</td>
<td>61.00±5.86</td>
<td>54.70±3.13</td>
<td>61.71±4.32</td>
<td>63.04±4.00</td>
</tr>
<tr>
<td>Xining</td>
<td>77.43±10.27</td>
<td>77.84±9.44</td>
<td>58.19±9.29</td>
<td>54.72±10.04</td>
<td>67.77±9.70</td>
</tr>
<tr>
<td>Yushu</td>
<td>77.35±5.55</td>
<td>73.34±6.37</td>
<td>56.12±5.53</td>
<td>55.02±5.01</td>
<td>66.05±5.50</td>
</tr>
<tr>
<td>Nyingchi</td>
<td>73.22±2.77</td>
<td>59.60±2.33</td>
<td>53.84±2.06</td>
<td>62.24±3.63</td>
<td>62.40±2.20</td>
</tr>
</tbody>
</table>
Tab. 4

<table>
<thead>
<tr>
<th></th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aba</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ngari</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Qamdo</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Diqing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gan nan</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ganzi</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Guoluo</td>
<td>19</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Halbei</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Haidong</td>
<td>22</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Hainan</td>
<td>14</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>Haixi</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Huangnan</td>
<td>23</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>Lhasa</td>
<td>12</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Nagqu</td>
<td>24</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>Shigatse</td>
<td>28</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>41</td>
</tr>
<tr>
<td>Sannan</td>
<td>33</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>Xining</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Yushu</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nyingchi</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>