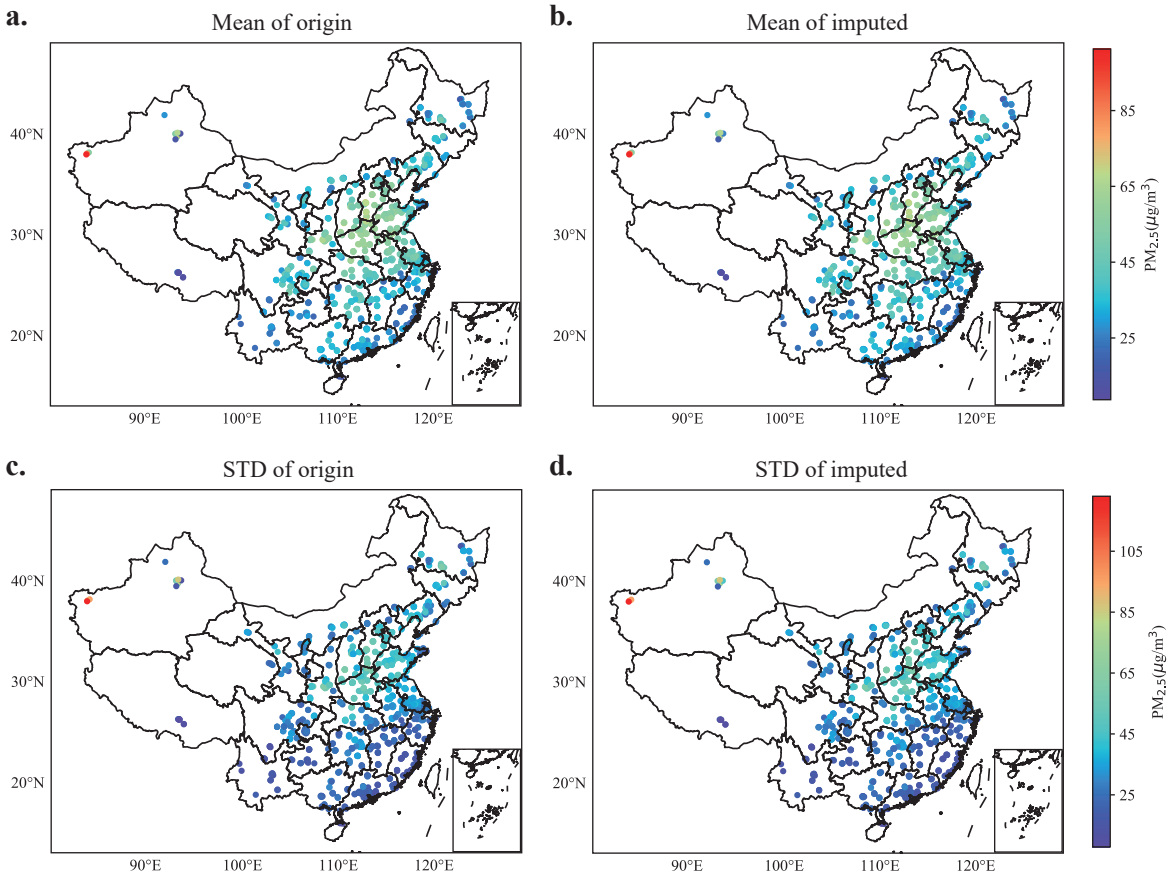

Figure S1. Spatial distribution of original and interpolated time series as for mean and standard deviation. Panel a and c are the mean and STD of time series of whole original $PM_{2.5}$ used in this work. Panel b and d are the same with a and c but with imputation. There is no significant change in colors between the distribution of original and imputed time series which implies the interpolation method is reliable.

1

Algorithm 1 KNN interpolation based on IDW

Initialization: read input site E
Calculate its' distance D_s with surrounding sites and construct distance matrix
if $D_s < 0.8$ radius **then**
    Get all proper sites and count the amount F
    if $F_s > 4$ **then**
        Randomly select 4 sites
    else if $F_s < 2$ **then**
        Drop E
    else
        Select all proper sites
    end if
end if
while exist missing values **do**
    IDW
end while

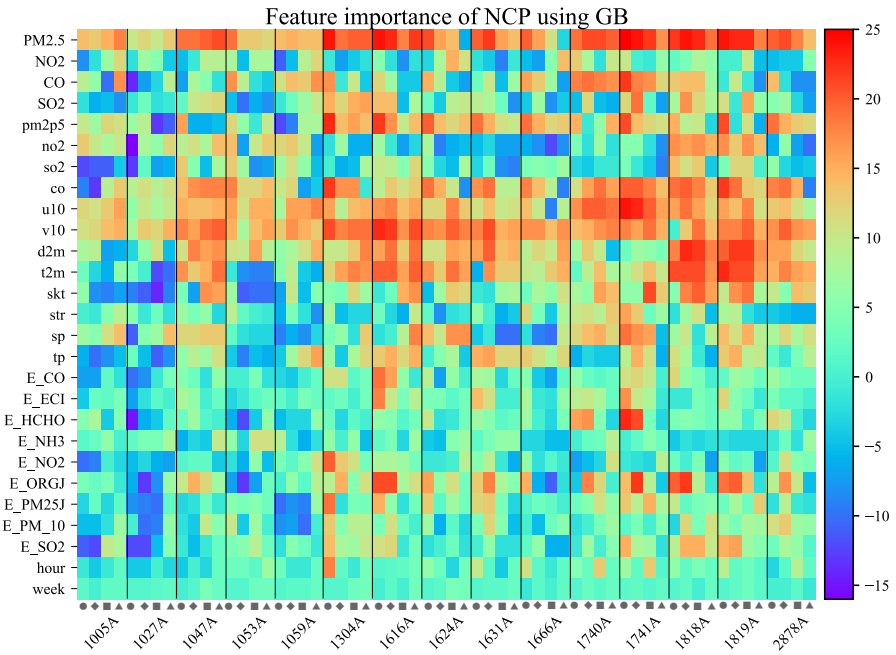

Figure S2. Heatmap of all empirical features with random 15 monitoring stations in NCP and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by GB. The warmer the color tone on the whole rows, the more important the corresponding feature.

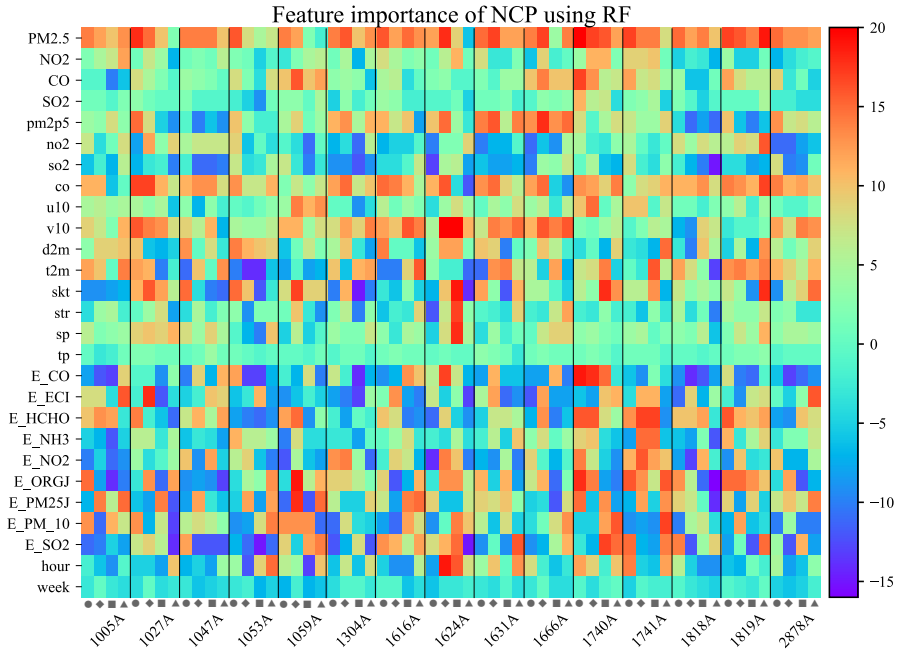

Figure S3. Heatmap of all empirical features with random 15 monitoring stations in NCP and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by RF. The warmer the color tone on the whole rows, the more important the corresponding feature.

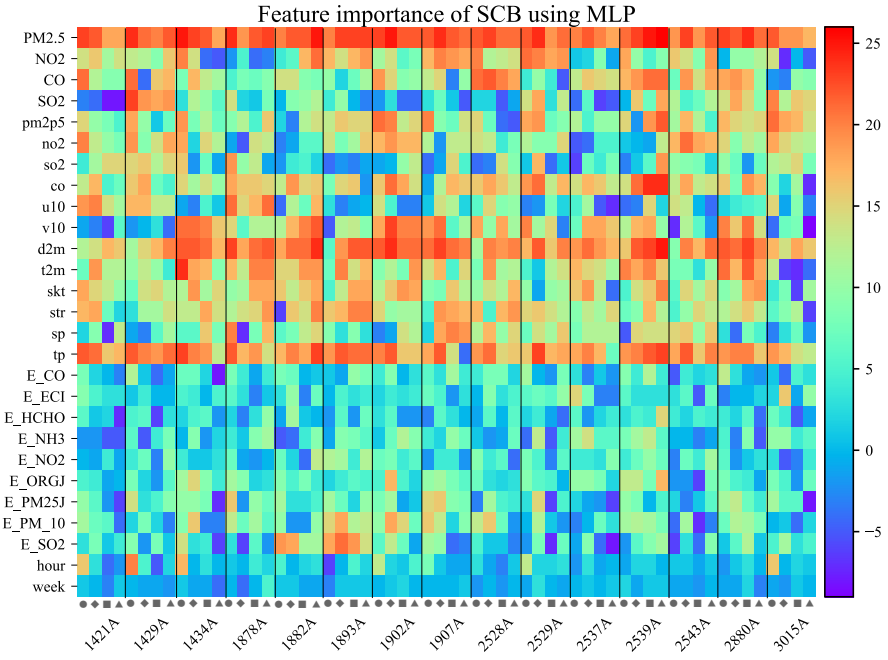

Figure S4. Heatmap of all empirical features with random 15 monitoring stations in SCB and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by MLP. The warmer the color tone on the whole rows, the more important the corresponding feature.

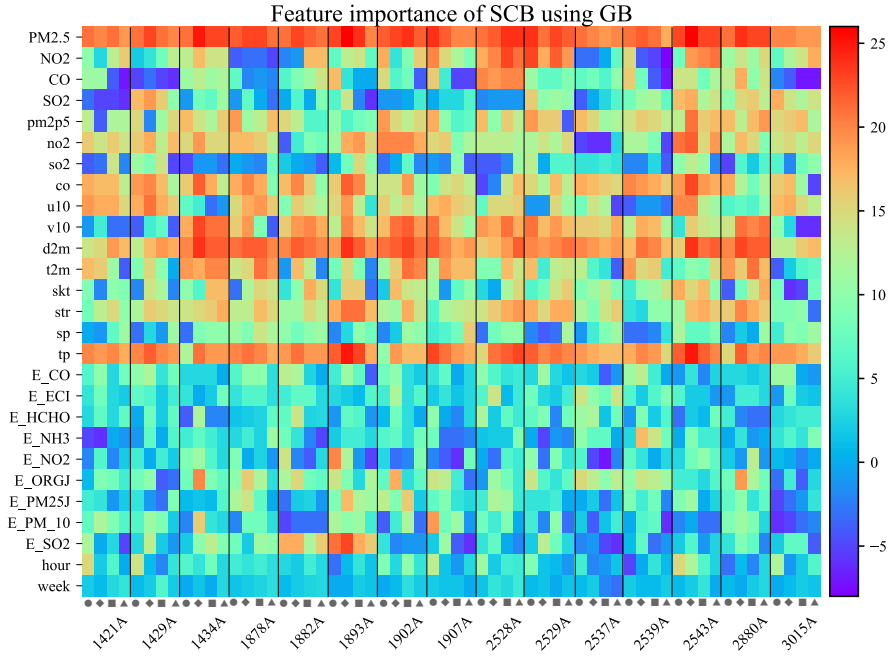

Figure S5. Heatmap of all empirical features with random 15 monitoring stations in SCB and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by GB. The warmer the color tone on the whole rows, the more important the corresponding feature.

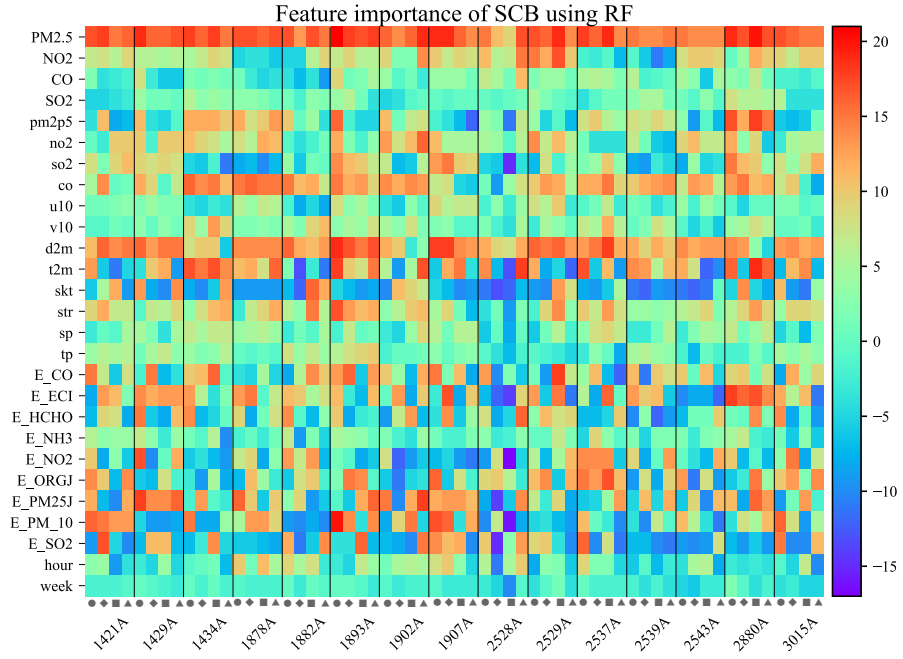

Figure S6. Heatmap of all empirical features with random 15 monitoring stations in SCB and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by RF. The warmer the color tone on the whole rows, the more important the corresponding feature.

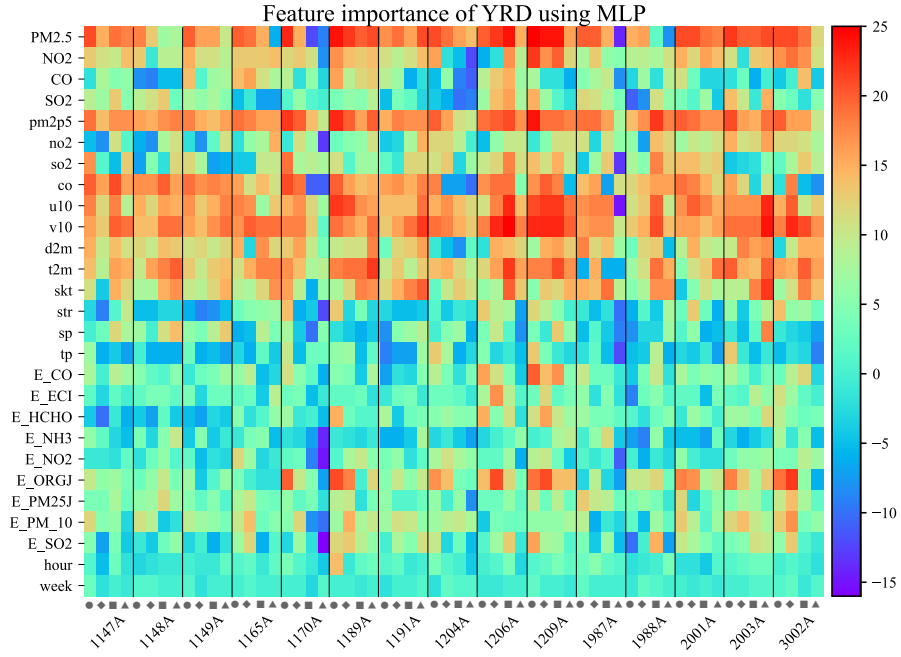

Figure S7. Heatmap of all empirical features with random 15 monitoring stations in YRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by MLP. The warmer the color tone on the whole rows, the more important the corresponding feature.

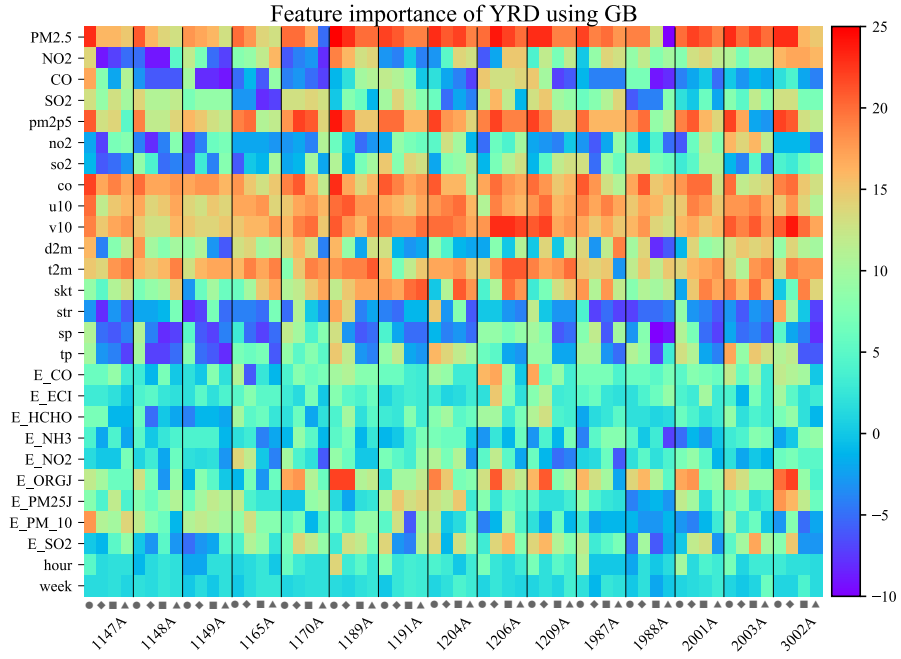

Figure S8. Heatmap of all empirical features with random 15 monitoring stations in YRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by GB. The warmer the color tone on the whole rows, the more important the corresponding feature.

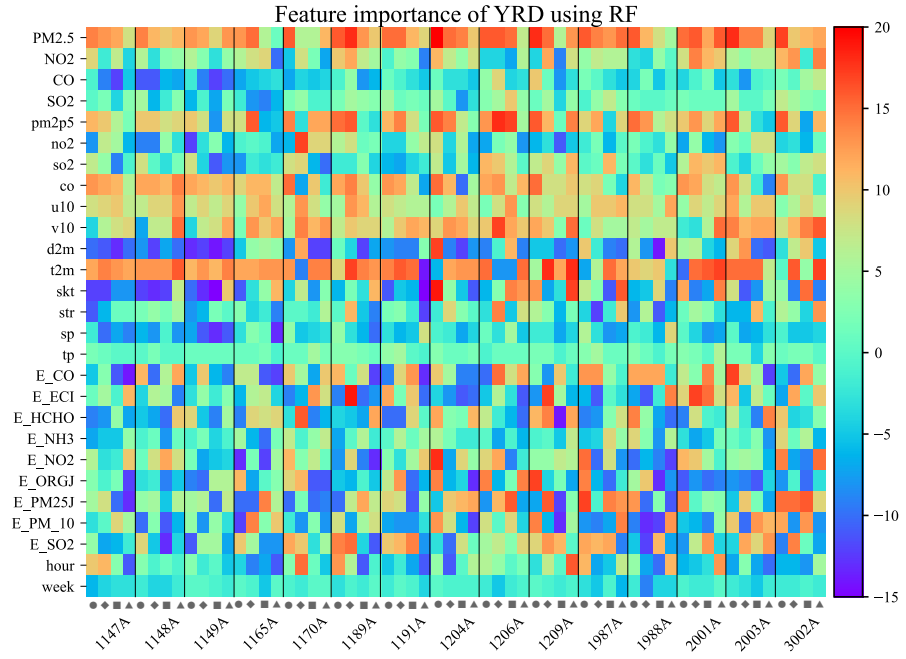

Figure S9. Heatmap of all empirical features with random 15 monitoring stations in YRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by RF. The warmer the color tone on the whole rows, the more important the corresponding feature.

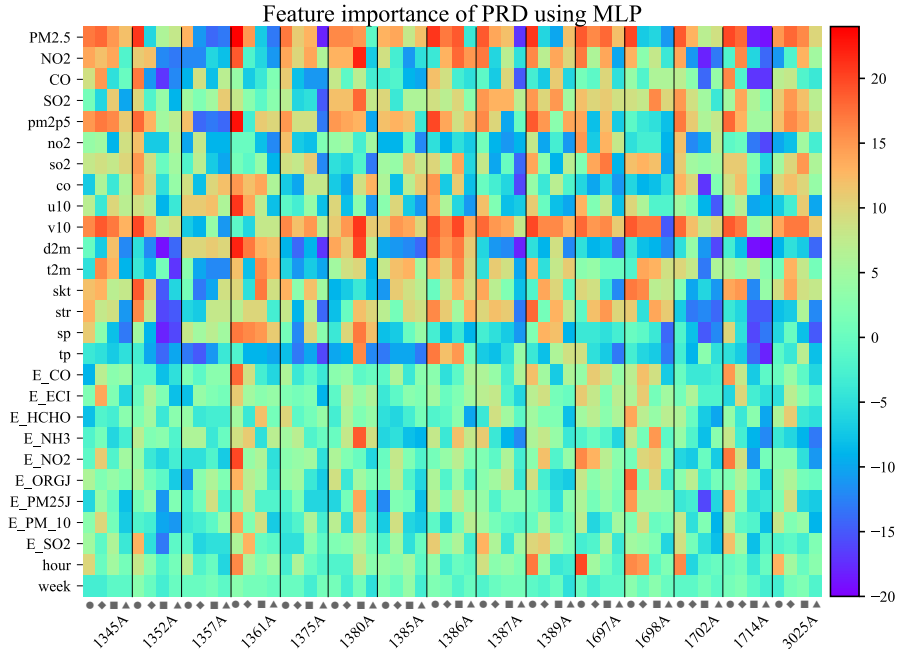

Figure S10. Heatmap of all empirical features with random 15 monitoring stations in PRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by MLP. The warmer the color tone on the whole rows, the more important the corresponding feature.

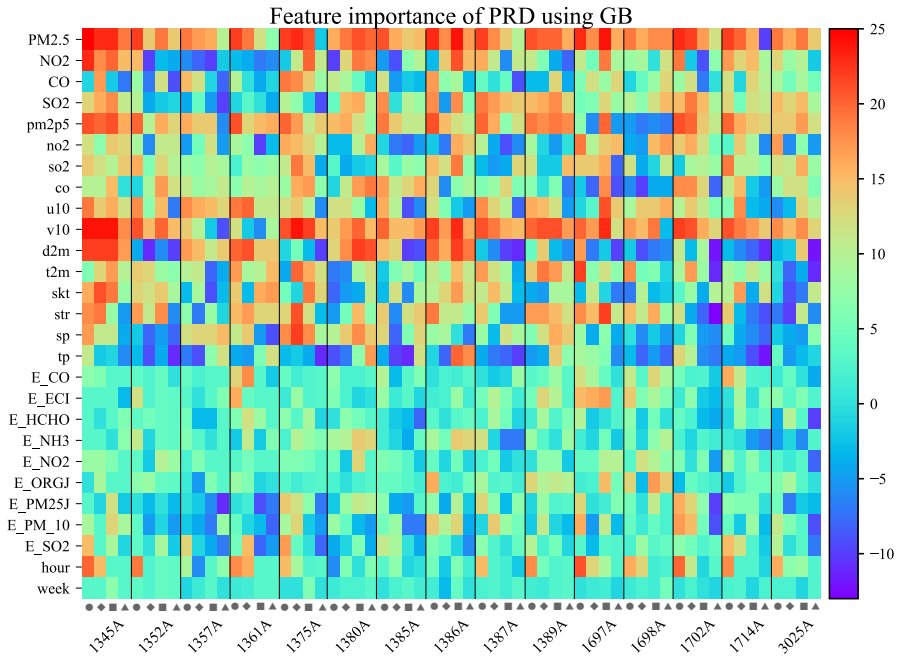

Figure S11. Heatmap of all empirical features with random 15 monitoring stations in PRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by GB. The warmer the color tone on the whole rows, the more important the corresponding feature.

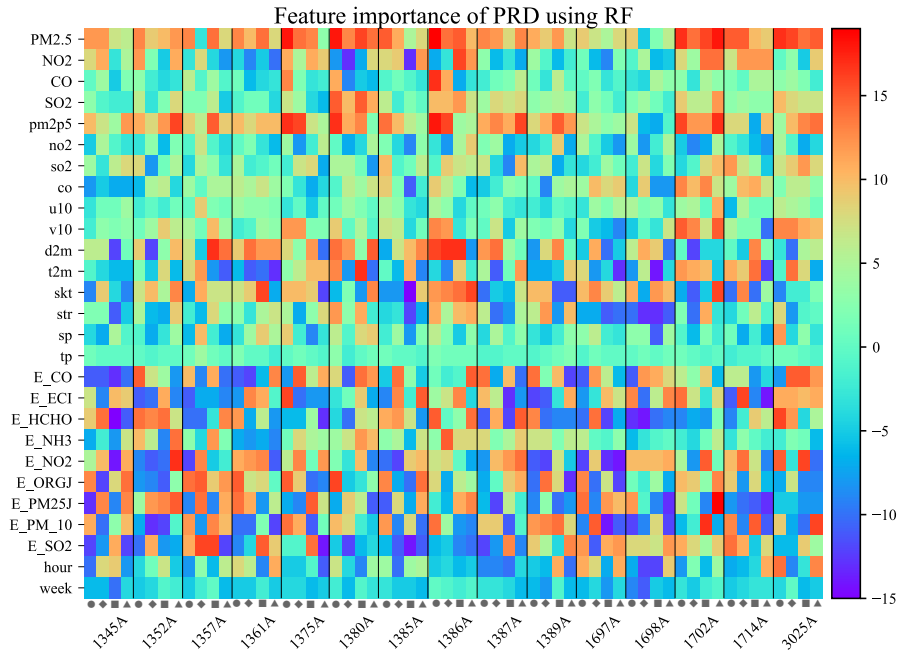

Figure S12. Heatmap of all empirical features with random 15 monitoring stations in PRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by RF. The warmer the color tone on the whole rows, the more important the corresponding feature.

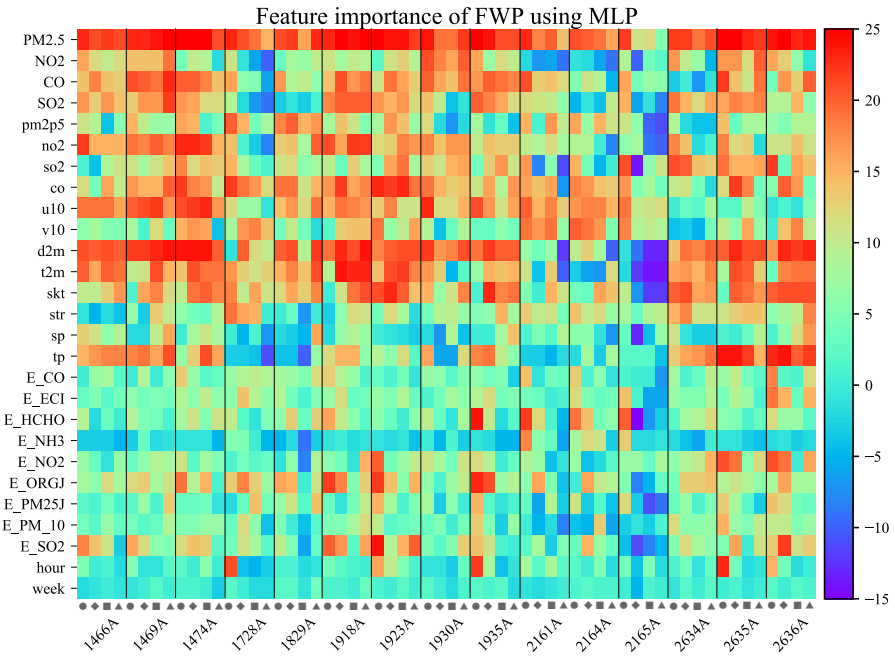

Figure S13. Heatmap of all empirical features with random 15 monitoring stations in FWP and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by MLP. The warmer the color tone on the whole rows, the more important the corresponding feature.

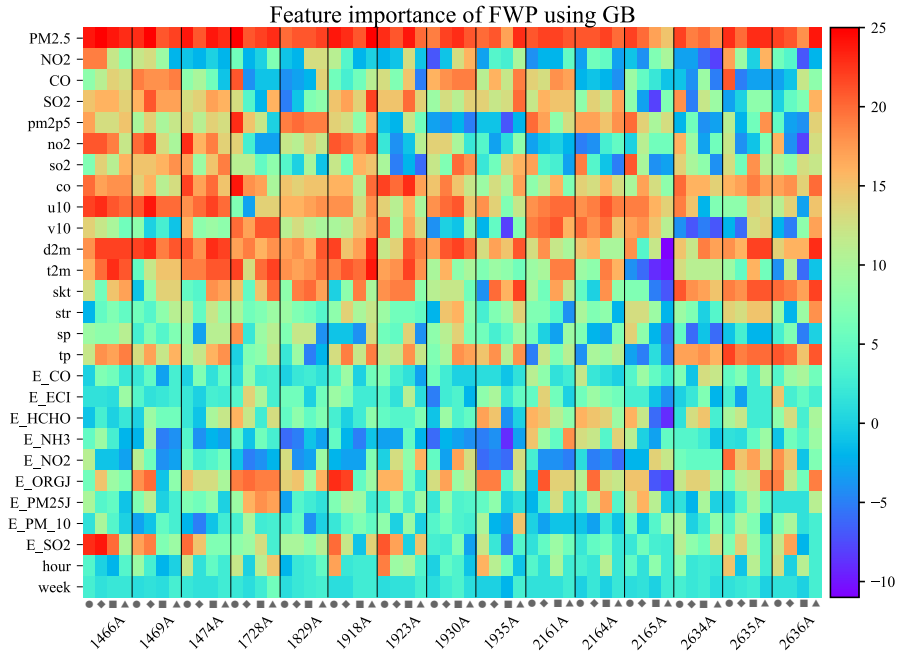

Figure S14. Heatmap of all empirical features with random 15 monitoring stations in PRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by GB. The warmer the color tone on the whole rows, the more important the corresponding feature.

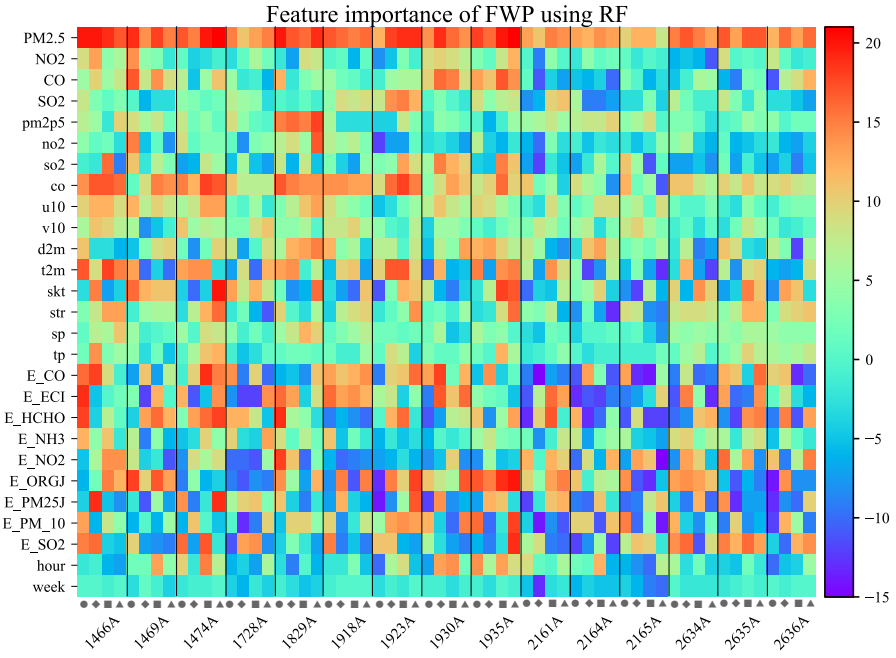

Figure S15. Heatmap of all empirical features with random 15 monitoring stations in PRD and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by RF. The warmer the color tone on the whole rows, the more important the corresponding feature.

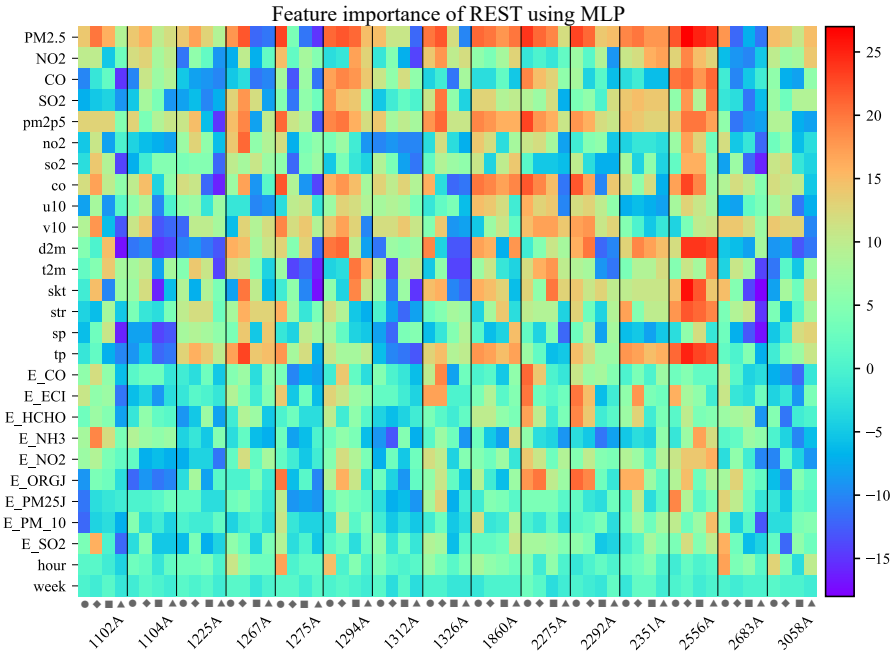

Figure S16. Heatmap of all empirical features with random 15 monitoring stations in the rest area of China and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by MLP. The warmer the color tone on the whole rows, the more important the corresponding feature.

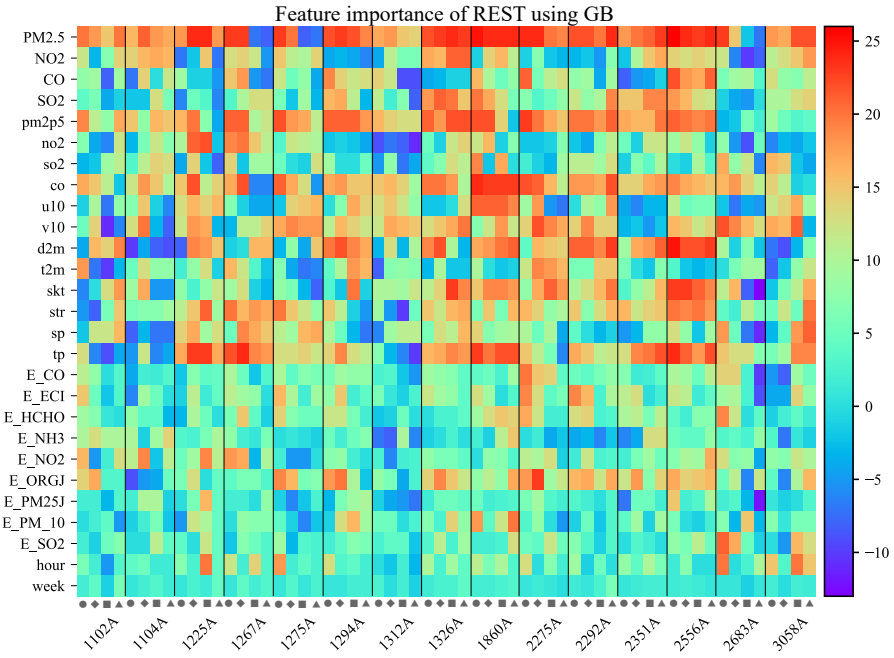

Figure S17. Heatmap of all empirical features with random 15 monitoring stations in the rest area of China and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by GB. The warmer the color tone on the whole rows, the more important the corresponding feature.

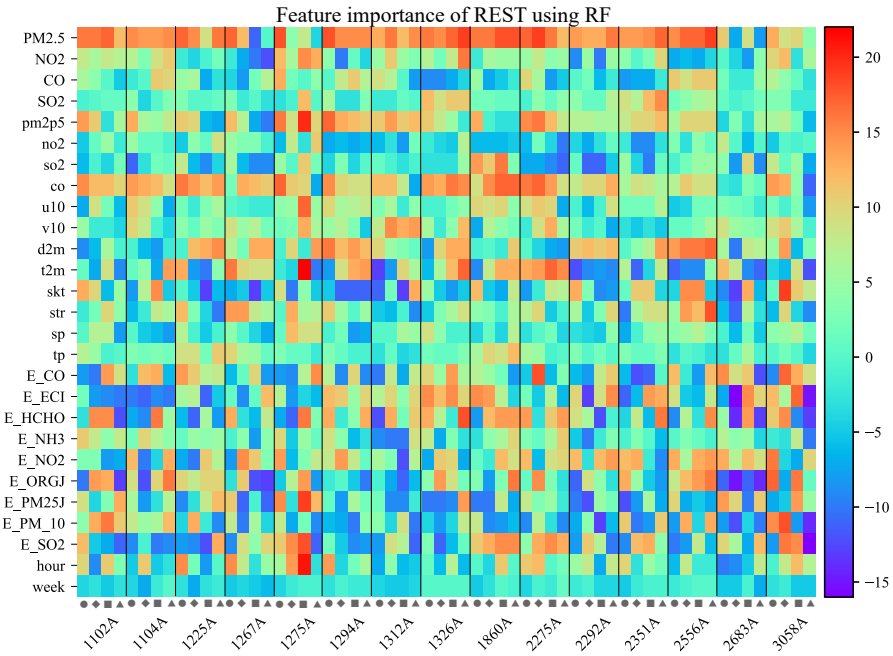

Figure S18. Heatmap of all empirical features with random 15 monitoring stations in the rest area of China and four predicting horizons. The circle, diamond, square and triangle represent four predicting horizons 6, 12, 18 and 24 h respectively. The heatmap is based on ranking the SAGE analysis of features training by RF. The warmer the color tone on the whole rows, the more important the corresponding feature.

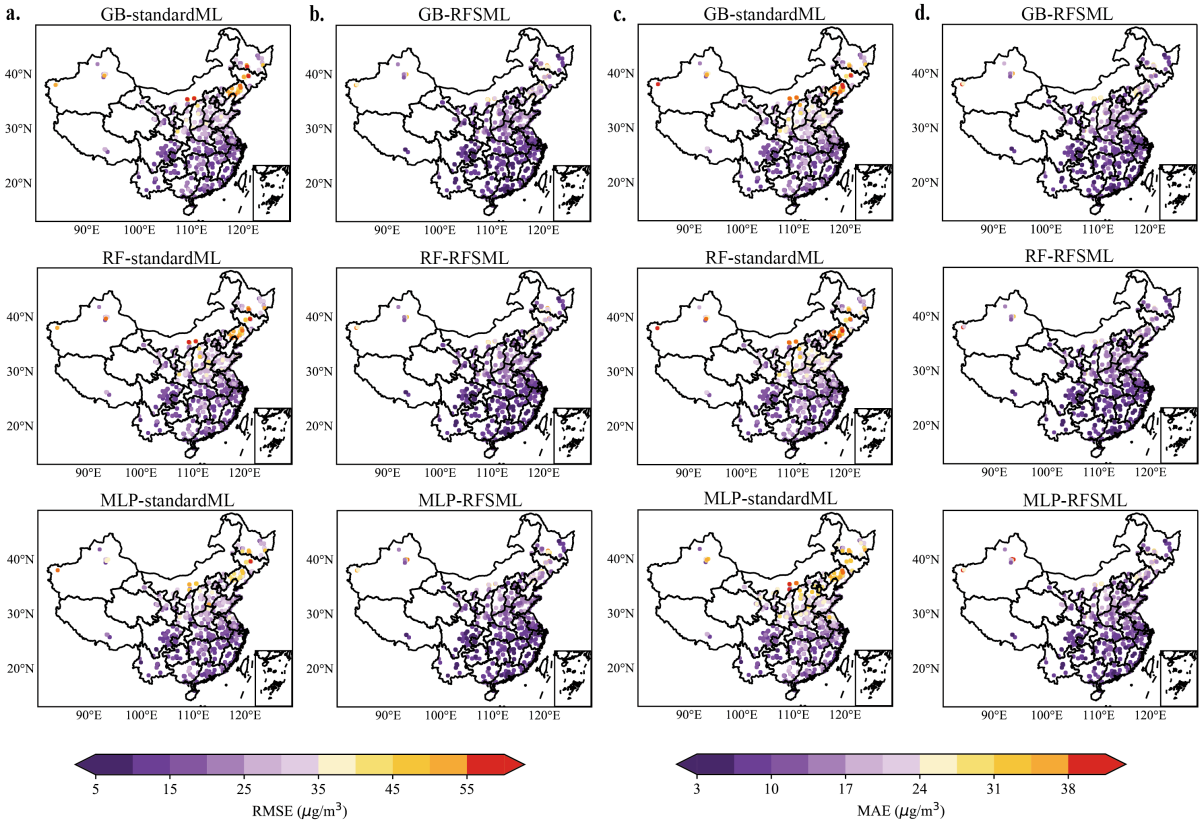

Figure S19. Spatial distribution of RMSE and MAE in a predicting horizon of 6 hours. Panel a and c are results of standard machine learning system while panel b and d are results of RFSML. The cooler the color tone, the lower the RMSE and MAE, thus the better predicting performance.

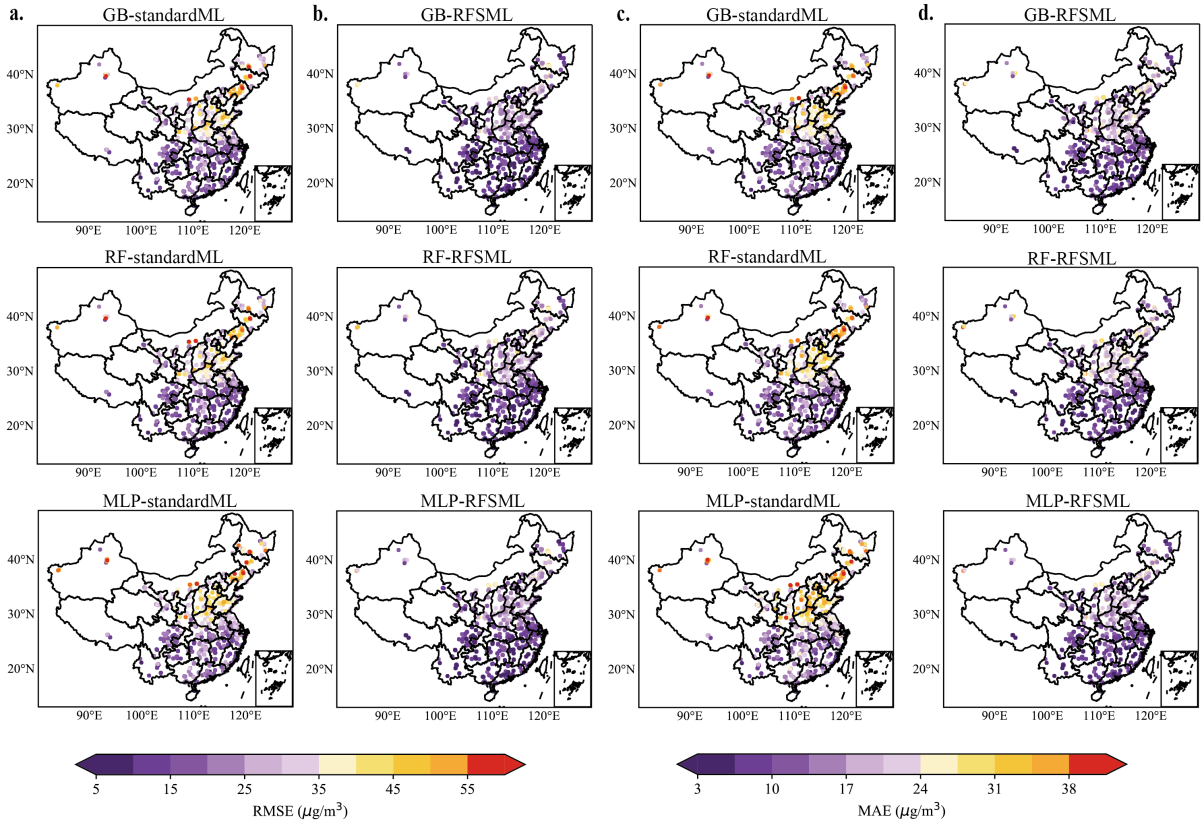

Figure S20. Spatial distribution of RMSE and MAE in a predicting horizon of 18 hours. Panel a and c are results of standard machine learning system while panel b and d are results of RFSML. The cooler the color tone, the lower the RMSE and MAE, thus the better predicting performance.

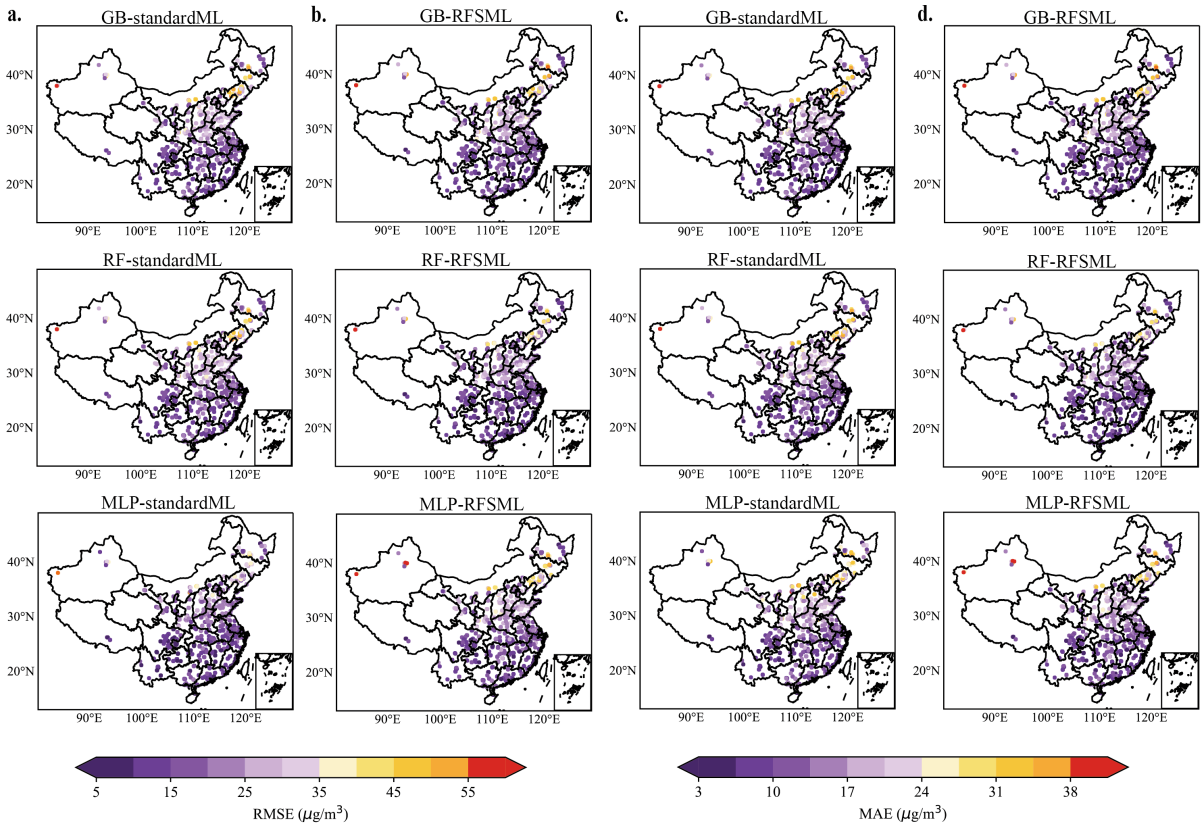

Figure S21. Spatial distribution of RMSE and MAE in a predicting horizon of 24 hours. Panel a and c are results of standard machine learning system while panel b and d are results of RFSML. The cooler the color tone, the lower the RMSE and MAE, thus the better predicting performance.