Supplementary Information Appendix

Contents

Text S1 Generalized Additive model (GAM) for the effects of meteorological factors on COVID-19

We aimed to describe the correlation between the intensity of COVID-19 transmission and meteorological factors in order to identify good predictors of the areas of high risk.

S1.1 Data and data exploration

The intensity of local transmission of COVID-19 was estimated by calculating the attack rate (AR_c) of affected county *c*. Log-transformed (log10)AR_c as of Mar 21 were selected to be dependent variable in GAM.

Meteorological variables included: monthly average temperature (AT, °C), temperature difference (TD, °C), relative humidity (RH, %), sunshine duration (SD, h), wind speed (WS, 0.1m/s) and cumulative precipitation (CP, mm).

Besides, distance to Wuhan (DW, 100km) and population density (PD, 1000 per km^2) of each county were also included from the start of the development of GAM to adjust for potential confounds.

Collinearity between factors to be included into the GAM, especially meteorological factors, was expected. Spearman correlation coefficient was calculated between pairwise factors, and the correlation matrix was visualized (Fig. S2).

S1.2 Generalized Additive model (GAM)

S1.2.1 Model description

We modelled the log(AR) using a GAM that included meteorological variables with adjustment for DW and PD. The model were fitted to data using the R package "mgcv".

The models used were described by the following equation:

$$\text{Log}(AR_c)=\beta_0+s(X_{1,c})+s(X_{2,c})+\ldots+ s(X_{n,c})+s(DW)+s(PD)$$

where $X_1,\ldots, X_n$ are continuous meteorological variables, *s*() are penalized spline function. *s*(DW) and *s*(PD) indicate the spline functions of DW and PD to adjust the confounders.

S1.2.2 Model selection: bottom-up strategy

Because some pairs of meteorological variables were highly correlated, and including correlated factors in the model may result in misleading, we used a bottom-up strategy using Akaike information criterion (AIC) as selection criteria. We started with all possible models with one meteorological variable with adjustment for DW and PD, and chose the one with the smallest AIC. At each step, we refitted all the possible models generated from adding one meteorological covariable, and chose the one that made the largest decreasing of AIC. We repeated the model selecting process until AIC was no longer decreased, and all the variables were significant ($p<0.05$).

S1.2.3 Results of model selection

In the basic GAM adjusting for DW and PD, WS was selected as the optimal meteorological factor due to the minimum AIC (Table S1). Spearman correlation analysis showed high correlations ($R_{Spearman}>0.6$) of various meteorological factors (Fig. S2). So we did not include any meteorological factor having high correlation with the factors already selected into the model in subsequent GAM selection process. The AIC of every GAM in the process of selection was listed (Table S1). Final model included AT, CP and WS, with adjustment for DW and PD (Table S2).

S1.3 The interactions of meteorological factors on log(AR)

The interactions were explored using GAM, and three-dimensional diagrams were made to visualized the pairwise interactions.

The model for interaction between AT and CP is as follow:

$Log(AR_c)=\beta_0+s(AT_{c,} CP_c)$

The model for interaction between AT and WS is as follow:

$Log(AR_c)=\beta_0+s(AT_{c,} WS_c)$

The model for interaction between WS and CP is as follow:

$Log(AR_c)=\beta_0+s(WS_{c,} CP_c)$

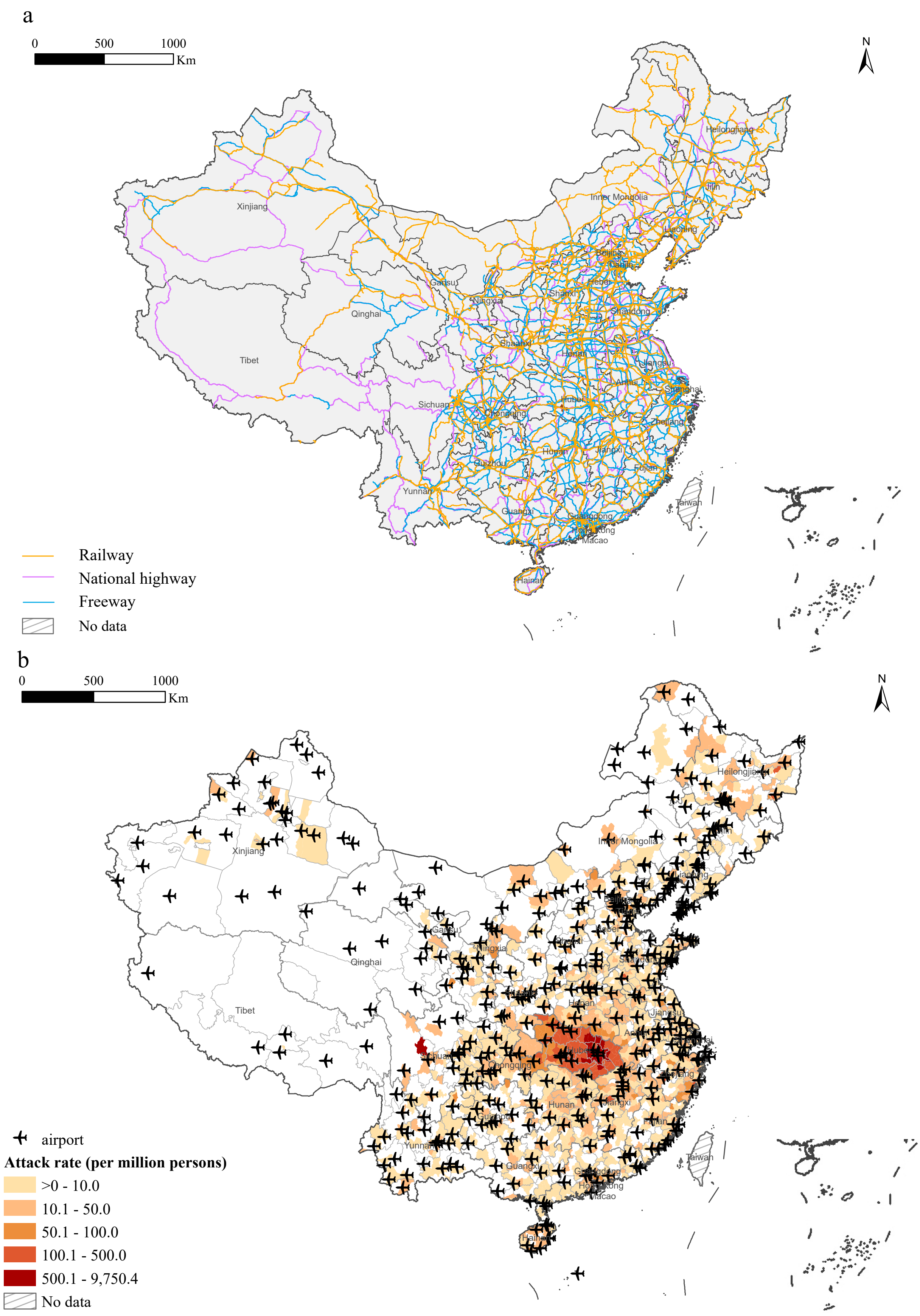

Fig. S1. Spatial overview of transportation routes (a) and locations of airports (b) in mainland China.

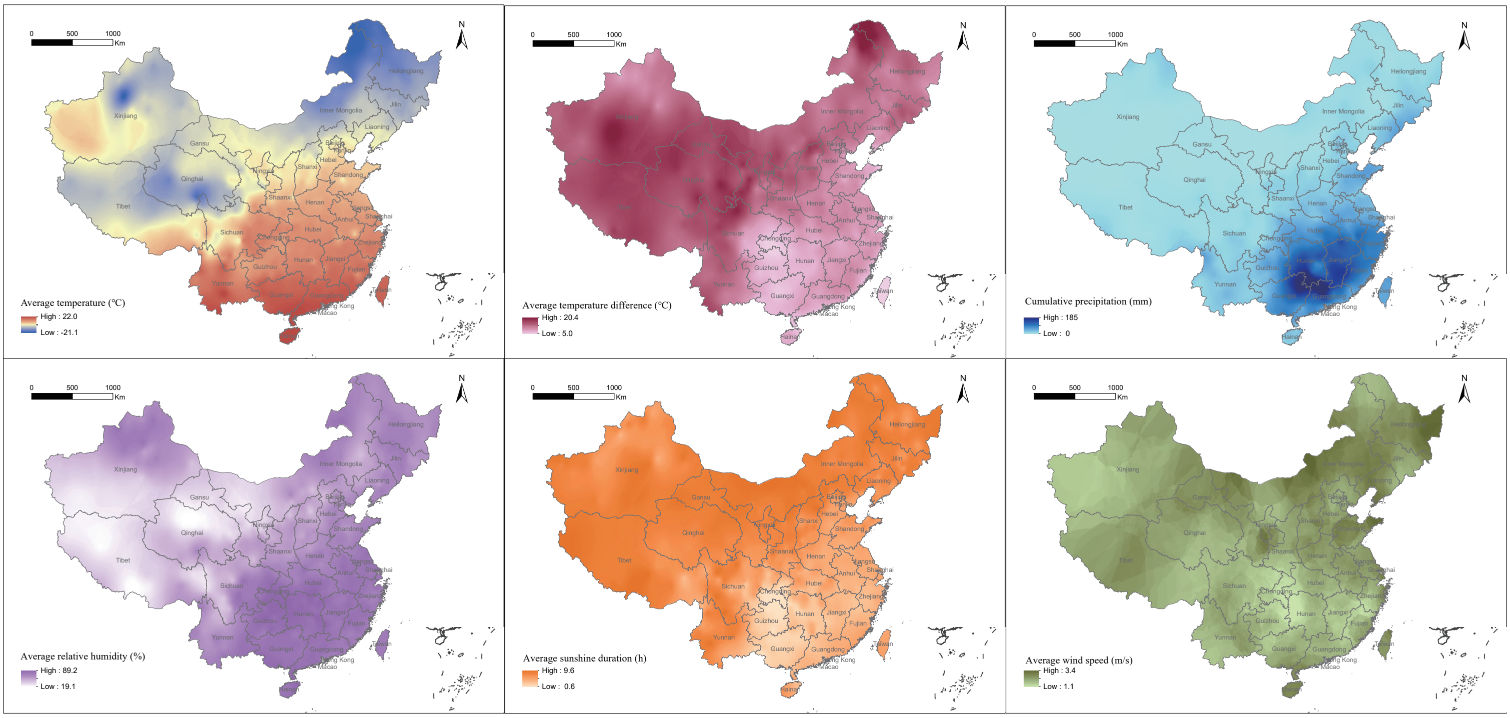

Fig. S2. Meteorological factors in February 2020 in China.
Meteorological factors including monthly average temperature, average temperature difference, cumulative precipitation, average relative humidity, average sunshine duration, and average wind speed are presented.

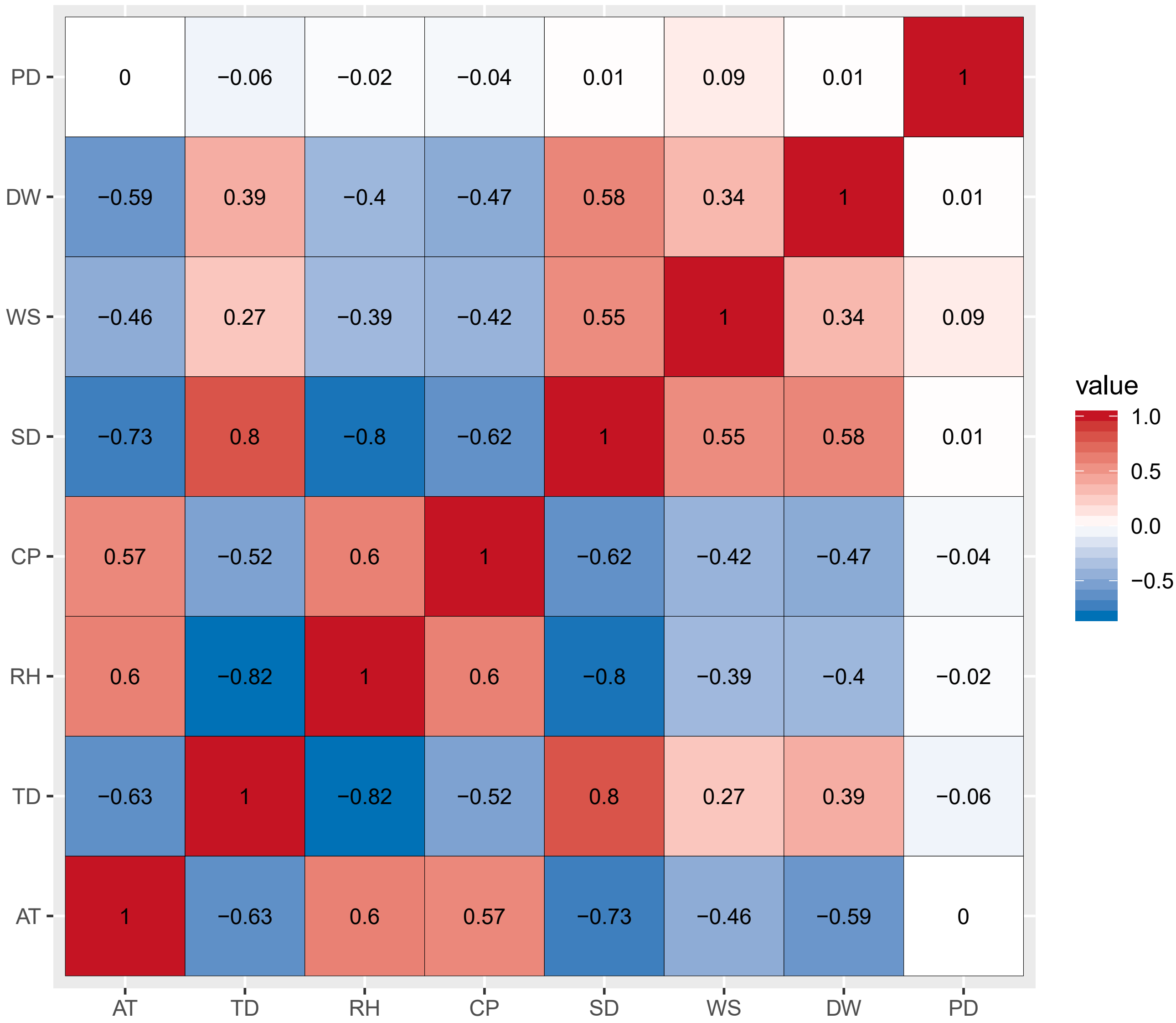

Fig. S3. Spearman correlation matrix of multiple factors.
Pairwise Spearman correlation efficient of monthly average temperature (AT, °C), average temperature difference (TD,°C), average relative humidity (RH, %), cumulative precipitation (CP, mm), average sunshine duration (SD, h), average wind speed (WS, 0.1m/s), distance to Wuhan (DW, 100 km) and population density (PD, 1000 per km^2).

Table S1. Generalized additive model selection using a bottom-up strategy.

Model	**Degree of freedom**	**AIC**
Univariate meteorological factor		
Average temperature (AT, ℃)	27	2020
Temperature difference (TD, ℃)	27	2029
Relative humidity (RH, %)	23	2044
Cumulative precipitation (CP, mm)	23	2042
Sunshine duration (SD, h)	21	2044
Wind speed (WS, 0.1m/s)	26	**2012**
Two meteorological factors		
WS+AT	33	1994
WS+CP	31	**1990**
WS+TD	28	2008
WS+RH	30	2009
WS+SD	30	2011
Three meteorological factors		
WS+CP+AT	37	**1970**
WS+CP+TD	32	1987
WS+CP+RH	34	1986

Table S2. Estimated parameters of the best model of generalized additive models.

	Estimated degree of freedom	**Reference degrees of freedom**	**F value**	**p value**
s(AT)	2	2	15.7	<0.001*
s(CP)	3	3	12.6	<0.001*
s(WS)	3	3	17.0	<0.001*
s(DW)	4	4	133.2	<0.001*
s(PD)	4	4	19.1	<0.001*

* Statistically significant.