Supplementary Materials and Methods

<u>Epidemiological data</u>

We scrutinized WHO's situation reports to rule out these countries with only imported cases, and only collected the confirmed cases with possible or confirmed local transmission (i.e., without recent travel history to China).

For Wuhan city, there was a shortage of test kits at the beginning of the pandemic, which would make confirmed case counts much lower than the actual data, thus, we discarded epidemic data before January 28th, the day when domestic test kits have been approved, produced in large quantities, and were available for Wuhan hospitals. As there was a cut down problem for the existing confirmed case count on February 20th for Wuhan, when modeling with the existing confirmed case count, only data before February 20th were used.

<u>Weather data</u>

Temperature and dew point displayed in Fahrenheit were transformed into Celsius forms, and relative humidity was calculated from temperature and dew point using the following formula for each time point:

$$\mathrm{RH} = \begin{cases} e^{\frac{7.5\mathrm{D}}{237.3+\mathrm{D}}-\frac{7.5\mathrm{T}}{237.3+\mathrm{T}}} \times 100\%, & T < 0 \\ 10^{\frac{7.5\mathrm{D}}{237.3+\mathrm{D}}-\frac{7.5\mathrm{T}}{237.3+\mathrm{T}}} \times 100\%, & T \geq 0 \end{cases}$$

where RH is the relative humidity, D is the dew point in degrees Celsius, T is the temperature in degrees Celsius, and *e* is the base of the natural log.

For each city with epidemiological data, the meteorological station in that city or that was closest to the latitude and longitude coordinates of the city center was chosen.

For a city with more than one meteorological stations, the one nearest to the city center was chosen. For a province with epidemiological data, the meteorological station in the capital city of that province was chosen. For a country with only national wide epidemiological data, weather data were averaged across all the meteorological observatories in the cities where outbreak was officially reported. Latitude and elevation for the meteorological observatories were also collected.

Statistical modeling

Only one city Wuhan was chosen for illustrating the time delay effect because it is the first city to have an outbreak of COVID-19, there was none reported imported cases for Wuhan, which might obscure the correlation between weather and virus transmission.

Supplementary Results

Datasets description

Only Chinese cities with monthly confirmed cases over 50 were included in the discovery dataset, which was 60 cities including Wuhan. The confirmed new cases in Wuhan on February 13, 2020, reached 13,436, which was oddly high as the daily confirmed new cases were no larger than 3,000 on all the other dates in Wuhan or in all the other Chinese cities. We suppose that it might be due to abrupt large supplement of virus test kits or data correction on that day. In order to reduce the potential contamination of modeling by this outlier, data on that day were discarded from the subsequent analysis. There were also two oddly large new confirmed case counts for Lombardy, which were discarded from the subsequent analysis. Except the outliers, the daily confirmed new cases in the discovery dataset ranged from 1 to 2,997, the average temperature ranged -22.54℃ ~ 22.16℃, the wind speed ranged 0.56 ~ 9.29 meter per second, visibility ranged 1.3 ~ 18.8 statute miles, and relative humidity ranged 30.84% ~ 98.52%.

Model selection

With the increase of relative humidity, the amount of droplets in the air increases, leading to more virus load. However, as the air gets humid, human's respiratory tract could better defend virus infection. Thus, the relationship of relative humidity could be complex, not pure linear. Giving comprehensive consideration, we defined the effect of relative humidity to be quadric. As for visibility, it only affects the amount of particles in the air, which is positively correlated with virus load. Thus, it is most probably to exert its effect linearly.

Although relative humidity and visibility 7 days ago correlated with the confirmed new case counts best, there was not great loss of model fitting statistics for relative humidity and visibility 3~7 days ago, as compared to the loss between 7 days time delay and 3~7 days time delay for temperature.

Fitted models

The fitted single-factor models were as follows:

$$\text{New Case Count} = -0.11305 \times \mathrm{T}^2 + 1.39819 \times \mathrm{T} + 45.11405$$

where T is temperature in ℃.

The estimate p-value for constant was < 0.001. The extremum was $-1.39819/(2 \times (-0.11305)) = 6.183945$ ℃.

$$\text{New Case Count} = -0.05759 \times \text{RH}^2 + 9.038 \times \text{RH} - 303.0$$

where RH is relative humidity in percentage.

The extremum was $-9.038/(2 \times (-0.05759)) = 78.46848$ %.

$$\text{New Case Count} = -1.360056 \times \text{SPD}^2 + 5.120123 \times \text{SPD} + 42.1855$$

where SPD is wind speed in meter per second (m/s).

The extremum was $-5.120123/(2 \times (-1.360056)) = 1.882321$ m/s.

$$\text{New Case Count} = -7.021 \times \text{VSB} + 89.041$$

where VSB is visibility in statute miles.

The estimate p-value for VSB was < 0.01, constant was < 0.001.

Thus, the complex short-term model to be regressed was

$$\begin{aligned}\text{New Case Count} \\ &= (-0.11 \times \text{T}^2 + 1.40 \times \text{T} - 0.058 \times \text{RH}^2 + 9.04 \times \text{RH} - 1.36 \\ &\times \text{SPD}^2 + 5.12 \times \text{SPD} - 7.02 \times \text{VSB} - 126.66) \times \text{a} \\ &\times \text{Existing Confirmed Case Count}\end{aligned}$$

where a is a constant to be fitted. All parameters take values 3~7 days before the day new case count is confirmed.

Through fitting this full model with the discovery data, a was estimated to be 0.0004786 (standard error 0.0000128, *p*-values < 2e-16).

For long-term model, the fitted model with temperature 14 days ago was as follows:

$$\text{New Case Count} = -0.10062 \times \mathrm{T}^2 + 1.11189 \times \mathrm{T} + 46.41792$$

The estimate p-value for constant was < 0.001. The extremum was $-1.11189/(2 \times (-0.10062)) = 5.525194$.

Thus, the simplified long-term model to be regressed was:

$$\text{New Case Count} = (-0.10 \times \mathrm{T}^2 + 1.11 \times \mathrm{T} + 46.42) \times \mathrm{b} \times \text{Existing Confirmed Case Count}$$

where b is a constant to be fitted. All parameters take values 14 days before the day new case count is confirmed.

Through fitting this simplified model with the discovery data, b was estimated to be 0.0061382 (standard error 0.0002666, *p*-values < 2e-16).

Table S1. Model fitness statistics for comparing and selecting proper fitting relationship

	sigma	finTol	logLik	AIC	BIC	deviance	Corr
Temperature							
Linear	493	4.5×10^{-8}	-167	339	342	4860391	0.757
Quadric	421	1.3×10^{-7}	-163	333	337	3370230	0.812
Relative humidity							
Linear	627	9.8×10^{-8}	-172	350	353	7855418	0.401
Quadric	626	8.4×10^{-6}	-171	351	355	7442367	0.358
Wind speed							
Linear	585	3.1×10^{-8}	-170	347	350	6840545	0.380
Quadric	546	2.4×10^{-7}	-168	344	349	5654728	0.423
Visibility							
Linear	594	3.3×10^{-8}	-171	347	351	7059799	0.354
Quadric	598	7.9×10^{-7}	-170	349	353	6799355	0.358

Note: sigma, estimated standard error of the residuals; finTol, the achieved convergence tolerance; logLik, the log-likelihood of the model; AIC, Akaike's Information Criterion for the model; BIC, Bayesian Information Criterion for the model; deviance, deviance of the model; Corr, Spearman's correlation coefficient between the real values and the predicted values by the predisposed model.

Table S2. Model fitness statistics for comparing and selecting proper time delay of virus exposure

	sigma	finTol	logLik	AIC	BIC	deviance	Corr
Temperature							
Day 0	626	2.6×10^{-8}	-171	351	355	7441513	0.330
Day -3	605	1.3×10^{-8}	-171	349	353	6953553	0.479
Day -7	664	5.4×10^{-8}	-173	353	358	8386957	0.262
Day -14	528	1.1×10^{-7}	-168	343	347	5297229	0.534
Day -3 ~ -7	421	1.3×10^{-7}	-163	333	337	3370230	0.812
Relative humidity							
Day 0	605	5.9×10^{-6}	-171	349	353	6953396	0.389
Day -3	679	4.3×10^{-6}	-173	354	359	8768069	0.065
Day -7	560	5.0×10^{-8}	-169	346	350	5962416	0.524
Day -14	605	9.1×10^{-6}	-171	349	353	6962609	0.326
Day -3 ~ -7	626	8.4×10^{-6}	-171	351	355	7442367	0.358
Wind speed							
Day 0	526	7.4×10^{-8}	-167	343	347	5251026	0.500
Day -3	663	1.4×10^{-8}	-173	353	357	8343427	0.268
Day -7	559	1.1×10^{-8}	-169	346	350	5926891	0.516
Day -14	674	5.2×10^{-8}	-173	354	358.	8643076	0.014
Day -3 ~ -7	546	2.4×10^{-7}	-168	344	349	5654728	0.423
Visibility							

Day 0	646	4.2×10^{-9}	-173	351	354	8343221	0.286
Day -3	663	5.1×10^{-8}	-173	352	355	8804055	0.016
Day -7	514	3.9×10^{-8}	-168	341	344	5290247	0.502
Day -14	635	1.1×10^{-8}	-172	350	354	8052388	0.272
Day -3 ~ -7	594	3.3×10^{-8}	-171	347	351	7059799	0.354

Note: sigma, estimated standard error of the residuals; finTol, the achieved convergence tolerance; logLik, the log-likelihood of the model; AIC, Akaike's Information Criterion for the model; BIC, Bayesian Information Criterion for the model; deviance, deviance of the model; Corr, Spearman's correlation coefficient between the real values and the predicted values by the predisposed model.

Table S3. Model fitness statistics for weather-combined model and epidemic only model

Model	sigma	finTol	logLik	AIC	BIC	deviance	Corr
Weather-combined	147	1.8×10^{-9}	-6239	12481	12491	21128810	0.171
Epidemic-only	149	2.1×10^{-8}	-6251	12507	12517	21689551	0.152

Note: The weather-combined model is the short-term model with multiplicative constant to be fitted. The epidemic-only model is the model only with existing confirmed case count as an independent variable, assuming a linear function.

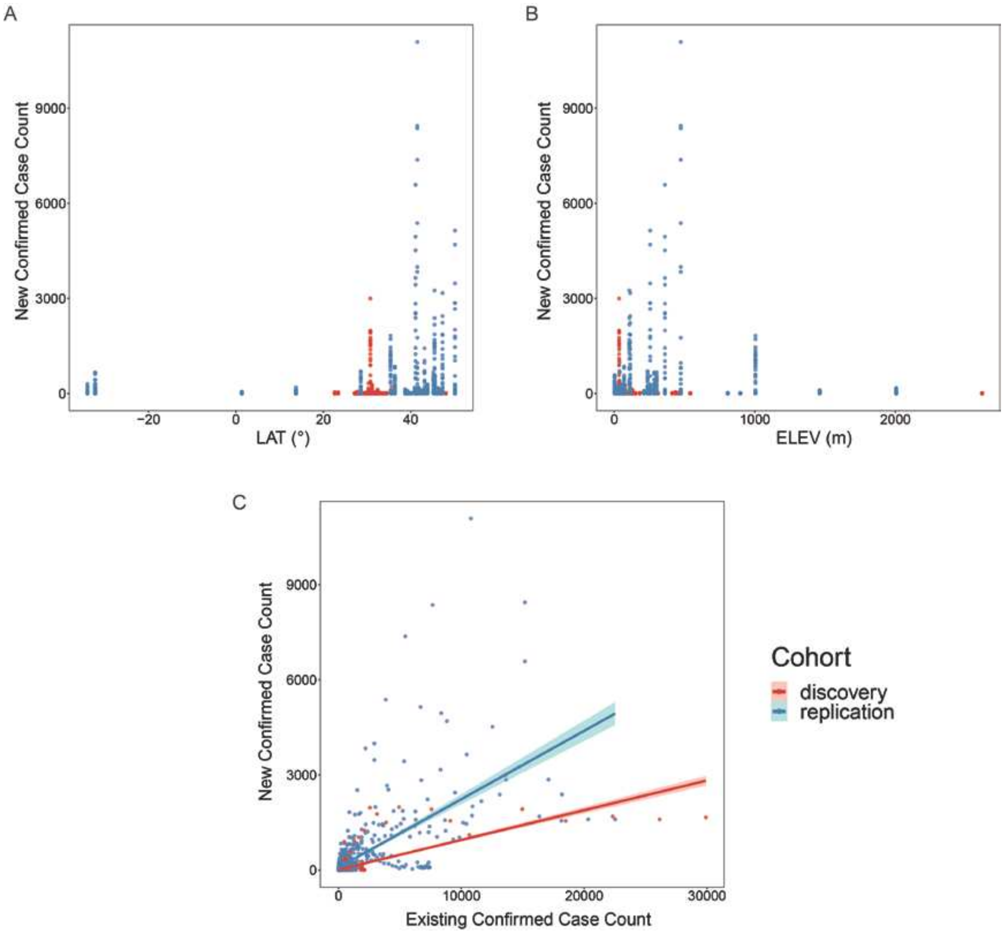

Fig. S1. Scatterplots of new confirmed case count to (A) latitude, (B) elevation, and (C) the existing confirmed case count, for all the studied sites. Linear regression (C) interpolation curves are illustrated for each dataset, with 95% confidence intervals showing in shadow.