Meteorological Applications

RESEARCH ARTICLE OPEN ACCESS

Meteorological Factors and the Spread of COVID-19: A Territorial Analysis in Italy

Vito Telesca[1] | Gianfranco Castronuovo[1] | Gianfranco Favia[2] | Mariarosaria Marra[1] | Marica Rondinone[1] | Alessandro Ceppi[3]

[1]School of Engineering, Università Della Basilicata, Potenza, Italy | [2]Department of Interdisciplinary Medicine, Università di Bari, Bari, Italy | [3]Department of Civil and Environmental Engineering (D.I.C.A.), Politecnico di Milano, Milano, Italy

Correspondence: Marica Rondinone (marica.rondinone@unibas.it)

Received: 1 August 2024 | **Revised:** 17 February 2025 | **Accepted:** 29 March 2025

Funding: This work was supported by the School of Engineering (University of Basilicata).

Keywords: artificial intelligence | atmospheric moisture | COVID-19 | dew point temperature | machine learning models | random Forest | Shapley additive explanation | STL decomposition

ABSTRACT

The COVID-19 pandemic has generated significant global impacts on health and society, imposing a comprehensive analysis of its influencing factors, including weather variables. This study investigates the interaction between meteorological conditions and the spread of COVID-19 in three Italian regions: Lombardia, Emilia-Romagna, and Puglia. Effects of weather variables, such as air temperature, relative humidity, dew point, solar radiation, wind speed, and barometric pressure, are explored in the incidence of disease. Observed meteorological and health data are taken from various sources, such as the citizen-science Meteonetwork Association and the National Department of Civil Protection, respectively, and they are analyzed with statistical methods and machine learning algorithms. The study emphasizes the necessity of carefully considering key meteorological quantities as primary drivers in illness diffusion and prevention strategies, offering valuable insights to address challenges to the pandemic and ensure the safety of global communities. The results reveal a significant correlation between specific atmospheric variables and the spread of COVID-19, with dew point temperature as the most influential parameter at low air temperature values.

1 | Introduction

The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has been one of the most serious global health crises in the last decade. Identified for the first time in Wuhan (China) at the end of 2019, the infection quickly spread worldwide, leading the World Health Organization (WHO) to declare a global pandemic on March 11, 2020. COVID-19 has affected millions of people, causing significant impacts on health, economy, and society. The response has involved rigorous containment measures, mass vaccination campaigns, and ongoing efforts to develop effective treatments. Considering the global impact of the infection, questions have arisen about which factors contributed to the disease's diffusion, with particular attention focused on climatic conditions.

COVID-19 belongs to the family of viruses that provoke severe acute respiratory syndrome (SARS): its risk lies in the common ability to be transmitted from person to person through the air, and it presents symptoms similar to influenza, while in severe cases, respiratory difficulties, pneumonia, multi-organ failure, and death may even occur. Like SARS, COVID-19 can be strongly correlated with meteorological factors such as temperature, humidity, solar radiation, and wind speed: these aspects influence transmission, environmental stability, and virus vitality. Meteorological variables have, in fact, a dual influence on pathogens: on one hand, they directly affect their survival, reproduction, and life cycle, while, on the other hand, by affecting the environment, they indirectly condition their habitat.

Literature studies have already shown that the onset, development, survival, and spread of diseases, particularly infectious ones, like 229E (alpha coronavirus), SARS-CoV (Severe Acute Respiratory Syndrome Coronavirus), and MERS-CoV (the Middle East Respiratory Syndrome Coronavirus), are closely correlated with temperature and relative humidity (Chan et al. 2011; Geller et al. 2012; van Doremalen et al. 2013; Pirouz et al. 2020a; Pirouz et al. 2020b). For instance, Qi et al. (2020) examined their validity on the COVID-19 cases in 30 Chinese provinces, demonstrating that temperature and moisture have a significant impact on the COVID-19 cases daily registered. His research revealed that the combination of increasing daily temperatures by 1°C, when relative humidity is between 67% and 86%, reduces the daily rate of COVID-19 cases by 36%–57%, while a growth of +1% in relative humidity, when daily average temperatures are between 5°C and 8°C, diminishes the daily COVID-19 cases by 11%–22%.

Furthermore, among the scientific community, there has been much discussion about the effect of temperature on COVID-19 transmissions in some countries with warm climates during the epidemic year 2020 (Demongeot et al. 2020; Iqbal et al. 2020; Mecenas et al. 2020). Despite initial speculation that warm and humid climates could slow the virus diffusion, facts have later shown that COVID-19 can quickly and widely spread in any type of climate, including hot and humid ones like in India (Vinoj et al. 2020). Additionally, Gupta et al. (2020) confirmed that temperature and humidity are suitable quantities in predicting COVID-19 transmission in the United States as well; Wu et al. (2020) tested the influence of meteorological factors on COVID-19 deaths, revealing a significant positive correlation with daytime temperature and a negative correlation with humidity, while Sahin (2020) highlighted a positive correlation between COVID-19 transmission rates and meteorological variables such as daily temperature and wind speed, emphasizing how this latter factor affects virus transmission with a two-week lag time. Pathogens can, in fact, spread from endemic regions to other areas through windstorms and dust, as in the case of avian flu outbreaks developed in some downwind regions, especially during dust storm seasons, or through the transport of influenza viruses during winter months from Asia to the Americas.

Nevertheless, no general conclusions have been drawn, and different findings were obtained around the world. Chin et al. (2020) stated that the virus is highly stable at 4°C, but more sensitive to hot temperatures. Conversely, Zhu et al. (2020) found that temperature has a positive correlation with COVID-19 cases when it is below 3°C, while Shi et al. (2020) showed the opposite effects in China, stating that temperatures between 8°C and 10°C reduced the odds of daily increases in the COVID-19 cases. In Brazil, Prata et al. (2020) demonstrated that a temperature rise of 1°C is associated with a 4.9% decrease in the daily COVID-19 transmissions when values are below 25.8°C. Lastly, low humidity rates and cold temperatures seem to significantly affect the survival and spread of various types of respiratory viruses, with influenza being a typical example as its spread has significantly grown.

In this context, our study proposes an in-depth investigation to identify whether and which weather factors among barometric pressure, air temperature, relative humidity, dew point temperature (hereafter dew point), incoming solar radiation, and wind speed have mostly influenced the COVID-19 transmission in three Italian regions: Lombardia and Emilia-Romagna in the north, and Puglia in the south, to ascertain whether influential meteorological variables were consistent or varied based on the weather conditions of each area. Based on the assumptions learned from the scientific literature, it is expected that some connections between COVID-19 and meteorological features during the period of the pandemic were observed in Italy as well. These relationships are here examined using artificial intelligence algorithms supported by the application of a data decomposition model; this latter allows for obtaining the trend component for the examined data, thereby eliminating the noise and seasonality.

2 | Area of Study

The focus of this research is on three Italian regions: Lombardia, Emilia-Romagna, and Puglia. Figure 1 shows the area of study with the location of the selected weather stations by Meteonetwork (MNW).

The first case of COVID-19 was recorded in February 2020 in Lombardia, which is the Italian region with the largest population (9,981,554) according to the ISTAT (2024) (Italian National Institute of Statistics: https://www.istat.it/ accessed on July 1st, 2024). Located in the northwest of the Italian peninsula, Lombardia mainly consists of flat terrain (47%) along with mountainous areas (41%). Its climate is complex due to the diverse natural features, such as the presence of rivers, lakes, and mountains. Like in the entire northern Italy, Lombardia's climate is strongly influenced by the Alps to the north and the Mediterranean winds from the south. Although belonging to a temperate climatic zone, the Po Valley is a transitional area between the Mediterranean Sea, and the continental and oceanic climate of central and western Europe, respectively. However, the complex Alpine orography acts as a barrier to cold and moist winds from the north and the Atlantic Ocean, resulting in a pronounced continental component; consequently, significant annual temperature variations occur, with cold winters and very hot summers. Lastly, as one ascends in altitude, the climate becomes predominantly mountainous, although with different characteristics between the Alps and the Apennines, especially concerning precipitation regimes (ARPA 2024, Regional Environmental Protection Agency: https://www.arpalombardia.it/temi-ambientali/meteo-e-clima/clima/il-clima-in-lombardia/ accessed on July 1st, 2024).

Belonging to the northern regions of Italy, the Emilia-Romagna (with 4,458,006 inhabitants) is mainly characterized by flat terrain (48%), with hills (27%) and mountains (25%). Its climate is predominantly subcontinental, with cold winters and hot summers. However, there are significant climatic variations between different places in the region. In the mountainous areas of the Apennines, the climate is colder, and precipitation can be abundant, especially during winter. In the plains and coastal areas, winters can be milder, and summers hotter, with

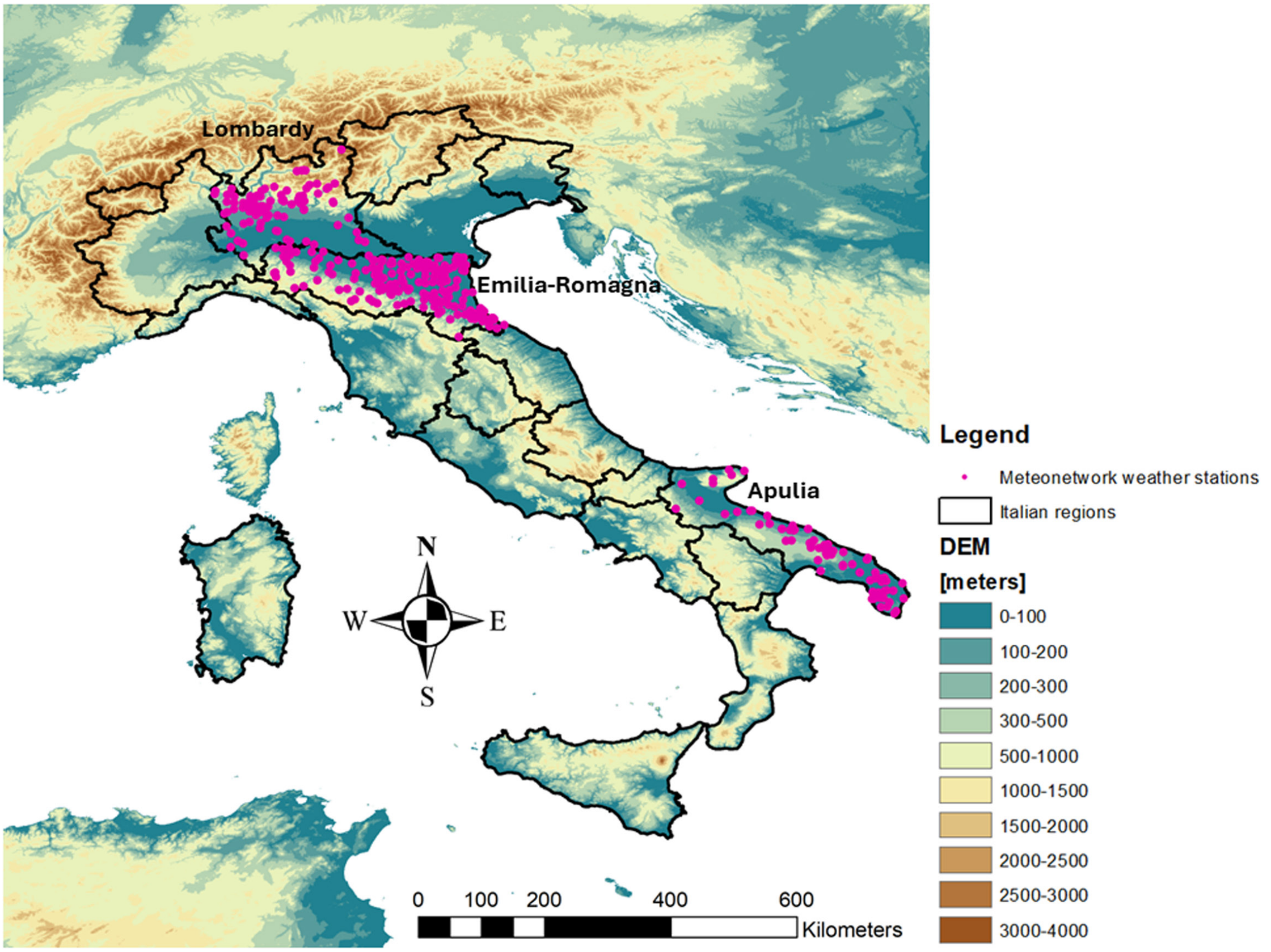


FIGURE 1 | The three investigated Italian regions: Lombardia, Emilia-Romagna, and Puglia. Magenta dots show the selected weather stations from the MNW meteorological network.

a greater influence of the Adriatic Sea. Precipitation is evenly distributed throughout the year, although there can be dry periods during summer (ARPAE 2024, Regional Environmental Protection Agency: https://www.arpae.it/it/temi-ambientali/clima/rapporti-e-documenti/atlante-climatico/ accessed on July 1st, 2024; Antolini et al. 2017).

The last examined region is Puglia (with 3,933,777 inhabitants), consisting of 53% plains, 45.5% hills, and only 1.5% mountains. Located in the southeast of the Italian peninsula, it borders the Adriatic Sea to the northeast, and the Ionian Sea to the south. The climate is typically Mediterranean, characterized by fairly warm and dry summers, mild and moderately rainy winters, and abundant precipitation during the autumn season (ISPRA, Italian National Institute for Environmental Protection and Research: https://www.isprambiente.gov.it/it accessed on July 1st, 2024).

3 | Materials and Methods

This section depicts the meteorological and epidemiological data used for the current analysis, with a detailed description of the applied methodology.

3.1 | Meteorological Data

Meteorological observations used in this study are taken from a citizen science open crowdsourced weather data network managed by Meteonetwork, an Italian meteorological association that shares more than 6500 weather station data over the European continent, with a higher density over Italy. In particular, more than 4700 sites daily update their databases, and 3400 are accessible in real-time at this website: https://www.meteonetwork.it/rete/livemap/ (accessed on July 1st, 2024). For further information about the MNW network, its weather data system, quality control procedures, metadata, etc., the reader can refer to Giazzi et al. (2022).

In this paper, the analyzed meteorological data span from February 24th, 2020 (immediately after the COVID-19 outbreak in Italy) till August 31st, 2023: thus, more than 3 years are available to compute main statistics and bioclimatological investigations. Main meteorological variables are collected at an hourly time step, in particular, 2-m air temperature (T) [°C], 2-m relative humidity (RH) [%], 2-m dew point (DEW) [°C], incoming shortwave solar radiation (RAD) [W/m2], wind speed (W) [m/s], and barometric pressure reduced at sea level (BARO) [hPa], and then averaged out at a daily scale. A total number

TABLE 1 | Distribution of the MNW weather stations for each Italian province.

Lombardia (98)	Emilia-Romagna (188)	Puglia (77)
Bergamo (18), Brescia (8), Como (9), Cremona (2), Lecco (4), Lodi (4), Mantova (4), Milano (13), Monza (17), Pavia (5), Sondrio (5), Varese (9).	Bologna (28), Ferrara (45), Forlì-Cesena (21), Modena (30), Parma (12), Piacenza (15), Ravenna (15), Reggio Emilia (7), Rimini (15).	Bari (19), Barletta-Andria-Trani (4), Brindisi (14), Foggia (9), Lecce (27), Taranto (4).

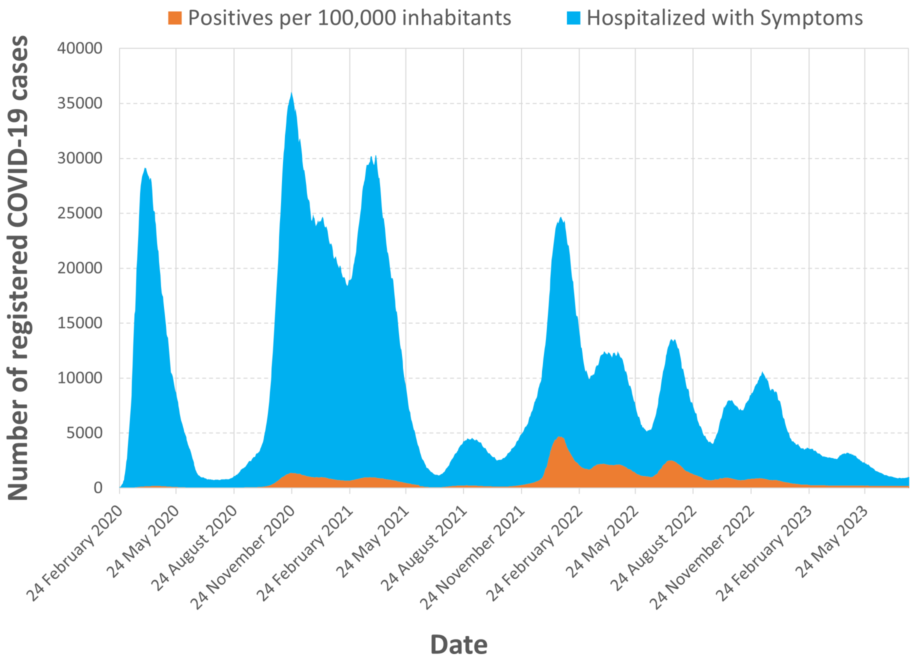


FIGURE 2 | Number of registered COVID-19 cases in Italy during the period 2020–2023. Data source: https://lab24.ilsolE24ore.com/coronavirus/?refresh_ce; accessed on July 1st, 2024.

of 363 meteorological sites is chosen from three main Italian regions (Table 1): Lombardia (98), Emilia-Romagna (188), and Puglia (77).

3.2 | Epidemiological Data

In addition to meteorological data, two key variables (hereafter, named target variables) related to infections are included and derived from the database of the National Department of Civil Protection: https://github.com/pcm-dpc/COVID-19 (accessed on July 1st, 2024): in detail, (i) the number of positive cases per 100,000 inhabitants, which is obtained by dividing the daily recorded positive cases (regardless of hospitalization status) by the population data from ISTAT of the respective regions and multiplying the result by 100,000; (ii) the number of hospitalized individuals with symptoms. Information regarding contagions is accessible through an interactive geographic dashboard where values concerning the trend of infections at the national, regional, and district scales are displayed. To provide a clear overview during the pandemic period in the Italian country, a graph illustrating the trend of positive cases per 100,000 inhabitants and hospitalized individuals with symptoms from 2020 to 2023 is shown in Figure 2.

These records of meteorological and epidemiological data allow for the creation of a dense database to analyze the spread of COVID-19, counting the number of positive cases per 100,000 inhabitants and the hospitalized individuals with symptoms over the three investigated regions.

3.3 | The Adopted Methodology

Since this paper aims to identify which meteorological factors potentially influence the virus spread, our study investigates the interaction between weather variables and the incidence of COVID-19 in regions characterized by different demographic features and climate conditions using an artificial intelligence (AI) model. The workflow of the applied methodology is illustrated in Figure 3 and summarized in the following four steps.

i. In the initial phase, a correlation analysis (Gogtay and Thatte 2017) is performed between meteorological variables and the ones of interest, i.e., the hospitalizations with symptoms and the positivity rates per 100,000 inhabitants. They are evaluated through the Pearson coefficient, which measures how two sets of data are related to each other: if the value is close to 1 (−1), it means there

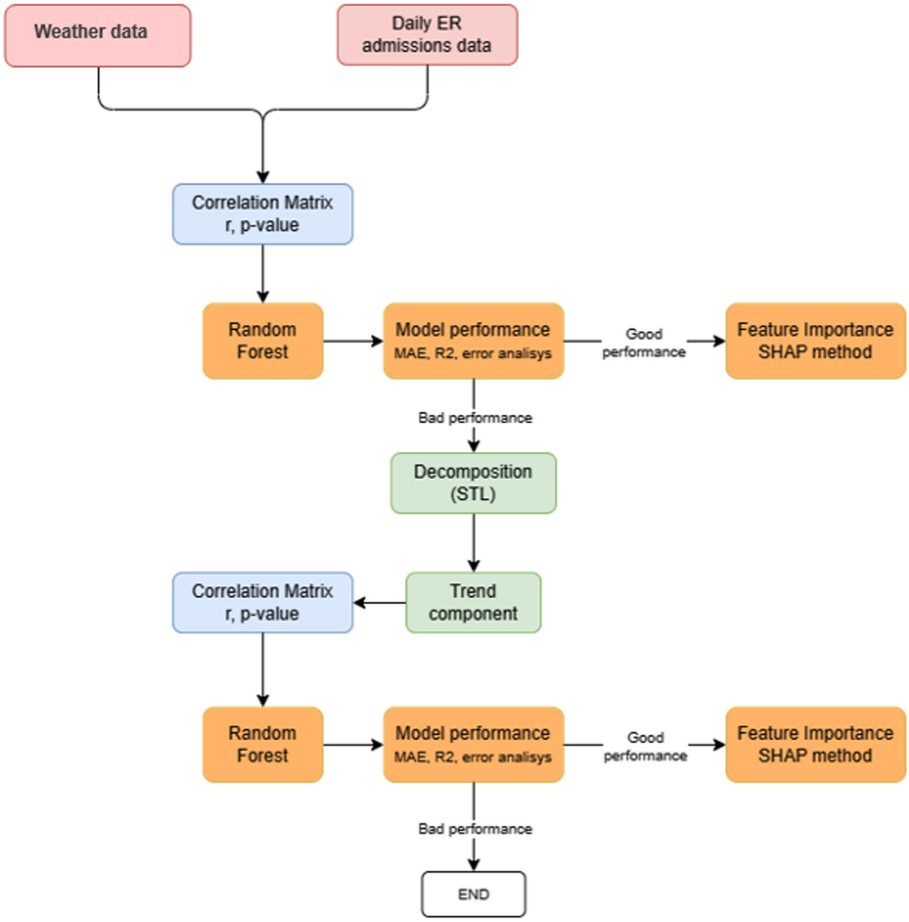


FIGURE 3 | Design of the tree diagram for the adopted methodology.

is a strong positive (negative) correlation, and both datasets increase (decrease) together; if it is around 0, no correlation exists.

ii. Afterward, the training and test phase with a simulation model, based on a supervised machine learning (ML) method (Shetty et al. 2022), is implemented. In this scenario, the COVID-19 dataset (i.e., the hospitalizations with symptoms and positives per 100,000 inhabitants) is the two target variables, while the predictors are represented by meteorological data. The algorithm, here used for constructing the regression model, is the Random Forest (RF) (Rodriguez-Galiano et al. 2015), which is generally effective for modeling nonlinear phenomena, avoiding the overfitting of data. This process is further evaluated with a k-fold cross-validation procedure that verifies whether the initial sample is subject to overfitting.

iii. If the results from the Pearson coefficient and RF simulation do not meet expectations (as it happened), the outlined methodology involves applying a data filter using a statistical technique called Seasonal and Trend decomposition with Loess (STL) (Cleveland et al. 1990). This procedure Equation (1) separates data into their fundamental components (trend, seasonality, and residuals) to remove intrinsic noise:

$$Y_t = T_t + S_t + R_t \tag{1}$$

where:

Y_t is the observed value of the time series at time t;

T_t is the trend component at time t;

S_t is the seasonal component at time t;

R_t is the residual (or noise) component at time.

The STL model utilizes a robust weighted local regression for the time series decomposition. During the estimation of a target variable, a subset of data points near the predicted variable is selected, and linear or quadratic regression is performed on this subset using the weighted least squares. The use of STL is applied to obtain the trend component both for the meteorological factors and the two target variables. Subsequently, a new correlation analysis is again conducted to find out the desired Pearson coefficient.

iv. Lastly, the weight of each variable in defining the model is assessed, i.e., how much each independent variable will influence the target one. This is accomplished through the application of the SHAP method (SHapley Additive exPlanation) (Mangalathu et al. 2020; Nohara et al. 2022), a popular Explainable Artificial Intelligence (XAI) approach (Barredo et al. 2020), generally used in game theory. The SHAP method calculates the Shapley value for each feature i in a model, according to Equation (2):

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \bullet [v(S \cup \{i\}) - v(S)] \quad (2)$$

where:

ϕ_i is the Shapley value for the feature i;

N is the set of all features;

S is a subset of features not including i;

$v(S)$ is the value function (model prediction) for the subset S.

The primary purpose of XAI methods is to make the functioning of artificial intelligence models transparent, avoiding the so-called black-box processes. Using XAI methods, it is possible to verify the importance of each observed variable and deduce whether, according to the model, it can influence the predicted values; upon this assumption, the model's reliability can be estimated.

4 | Results

The results of the four steps highlighted in the methodology are presented in this section. The first part of the analysis is conducted without applying the STL model, generally obtaining poor scores for the RF model simulation and with the k-fold cross-validation. Once the decomposition function is applied and better outcomes are achieved, the investigations continue by introducing the SHAP method to determine the Feature Importance of meteorological factors on the two target variables during the period from February 24, 2020, to August 31, 2023.

4.1 | Preliminary Analysis

The initial phase of this study evaluates the relationship between the selected meteorological and the epidemiological variables in the investigated regions of Lombardia, Emilia-Romagna, and Puglia. By calculating the Pearson correlation coefficient and the corresponding correlation matrix with a p-value set to 0.01, the statistical significance of correlations is determined.

Subsequently, a machine learning model (the Random Forest regressor) is applied to simulate the evolution of the epidemiological data over the entire period; this involves dividing the dataset into two parts: one for training the model (80%), and the other for testing (20%).

In addition, the goodness of the model is assessed using the cross-validation method with the k-fold technique with k equal to 10 over the entire dataset. Cross-validation is a statistical method used to evaluate the generalization ability of a machine learning model, i.e., its skill to predict new data not seen during the training phase. The model is thus trained k times, each time using k-1 folds for training and the remaining fold for testing. This process helps to reduce errors due to variability in training/test data, avoiding overfitting, providing a more reliable estimate of the model performance for new values, and ensuring that the model performance does not depend on the split interval between training and test data.

The performance of the training/test phases of the predictive RF model and the k-fold cross-validation are assessed through some statistical indexes (Chicco et al. 2021), commonly known in the literature: the determination coefficient (R^2) and the Mean Absolute Error (MAE). For the k-fold cross-validation, the computed MAE considers the average accuracy of the model across the distinct parts of the training and test subset of data. If the cross-validation values are in the same order of magnitude as those initially obtained in the test phase by the predictive RF model, the model performance can be considered acceptable.

In this preliminary examination, neither the initial correlation analysis nor the application of the RF model yielded the desired results: in fact, the correlations obtained are generally weak, with absolute values of the Pearson coefficient well below 0.5 (magenta values in Table 2), suggesting there is no strong correlation between the selected meteorological variables and the two epidemiological ones. Furthermore, unsatisfactory values of R^2 and MAE, not reported here for brevity,

TABLE 2 | Results of the correlation matrix for Lombardia, Emilia-Romagna, and Puglia before (in magenta) and after (in green) introducing the STL method for the number of hospitalized with symptoms and positives per 100,000 inhabitants.

	Lombardia		Emilia-Romagna		Puglia	
Weather variables	**Hospitalized with symptoms**	**Positives per 100,000 inhabitants**	**Hospitalized with symptoms**	**Positives per 100,000 inhabitants**	**Hospitalized with symptoms**	**Positives per 100,000 inhabitants**
BARO	0.230 (0.150)	0.390 (0.200)	0.0001 (0.000)	0.340 (0.190)	0.190 (0.100)	0.280 (0.140)
RH	−0.078 (−0.056)	0.011 (−0.038)	0.1900 (0.120)	0.300 (0.190)	0.260 (0.200)	0.044 (0.012)
RAD	−0.080 (−0.080)	−0.280 (−0.230)	−0.350 (−0.320)	−0.270 (−0.240)	−0.340 (−0.310)	−0.180 (−0.160)
DEW	−0.450 (−0.410)	−0.380 (−0.350)	−0.610 (−0.580)	−0.380 (−0.350)	−0.490 (−0.430)	−0.460 (−0.410)
T	−0.360 (−0.350)	−0.330 (−0.310)	−0.560 (−0.540)	−0.390 (−0.370)	−0.460 (−0.450)	−0.350 (−0.340)
W	0.0087 (0.056)	−0.230 (−0.099)	−0.190 (−0.002)	−0.230 (−0.060)	−0.0097 (0.025)	0.110 (0.093)

are even obtained for the RF model simulation and with the k-fold cross-validation.

4.2 | The STL Method

Following the methodological framework, the STL method was applied. This process filters the predictive data through a decomposition function, allowing data to be divided into three components: seasonal, trend, and residuals. Using the trend component, which better represents the data progression over time, a new sub-dataset is created with both meteorological and epidemiological values; hence, a new analysis is carried out. The Pearson correlation coefficient is once again calculated, but this time better (absolute) values, compared to those previously obtained, are achieved (green values in Table 2).

Subsequently, we apply the supervised machine learning model using the RF regression algorithm in its training and test phase, but it is now executed using this daily dataset, created with the trend variables obtained from the STL decomposition model. Its performance is again evaluated using the R^2 and MAE metrics for the number of hospitalized with symptoms and the positives per 100,000 inhabitants (first four columns of Tables 3 and 4, respectively). Very high scores prompt us to use the regressor model over the entire data series simulation, thereby evaluating its overall performance (Figures 4–6). Small differences are

TABLE 3 | Results of the training/test phases of the model performance and the k-fold cross-validation over the three Italian regions regarding the hospitalized with symptoms.

Region	R^2 train	R^2 test	MAE train	MAE test	R^2 mean train	R^2 mean test	MAE mean train	MAE mean test
Lombardia	0.991	0.918	0.008	0.022	0.993	0.946	−0.007	−0.017
Emilia-Romagna	0.994	0.958	0.009	0.024	0.994	0.966	−0.008	−0.022
Puglia	0.985	0.926	0.014	0.035	0.989	0.920	−0.012	−0.033

TABLE 4 | Results of the training/test phases of the model performance and the k-fold cross-validation over the three Italian regions: the positives per 100,000 inhabitants.

Region	R^2 train	R^2 test	MAE train	MAE test	R^2 mean train	R^2 mean test	MAE mean train	MAE mean test
Lombardia	0.989	0.884	0.006	0.017	0.993	0.949	−0.005	−0.013
Emilia-Romagna	0.992	0.964	0.005	0.012	0.994	0.963	−0.004	−0.010
Puglia	0.982	0.880	0.016	0.044	0.984	0.898	−0.014	−0.037

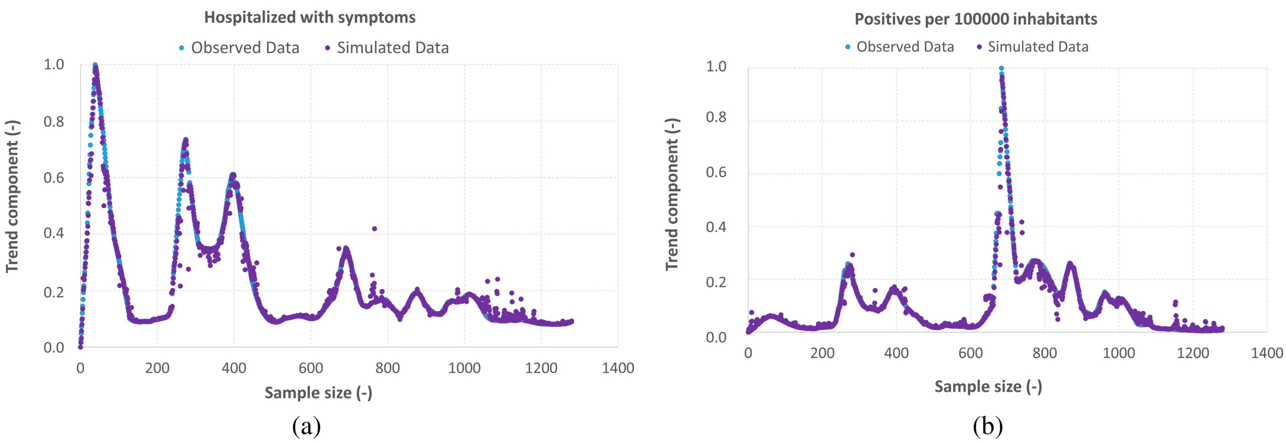
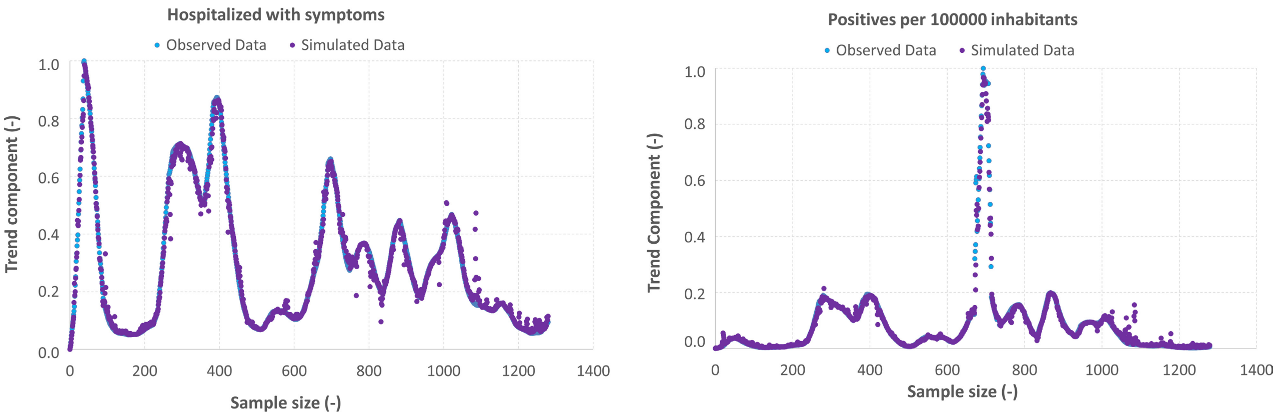


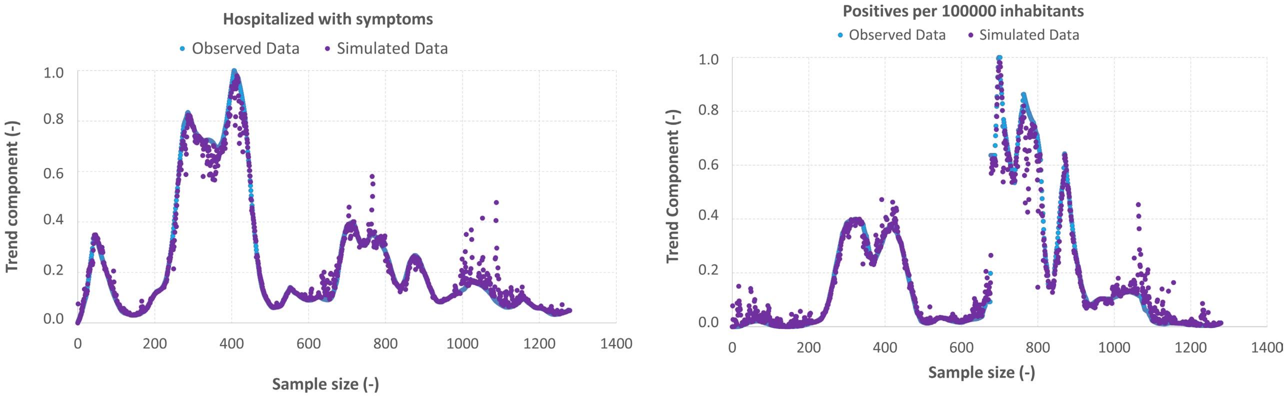
FIGURE 4 | Trend component analysis between observed (in blue) and simulated (in magenta) data for the number of hospitalized with symptoms (a) and the positives per 100,000 inhabitants (b) in the Lombardia region.

seen between simulated and measured values for the two target variables, indicating that the model well-simulates the observed data trend. The whole simulation is then assessed with the k-fold cross-validation (second four columns in Tables 3 and 4), where the skill scores achieved are consistent with those initially obtained in the test phase by the predictive RF model, highlighting the robustness of the model, and pointing out the absence of overfitting phenomena.



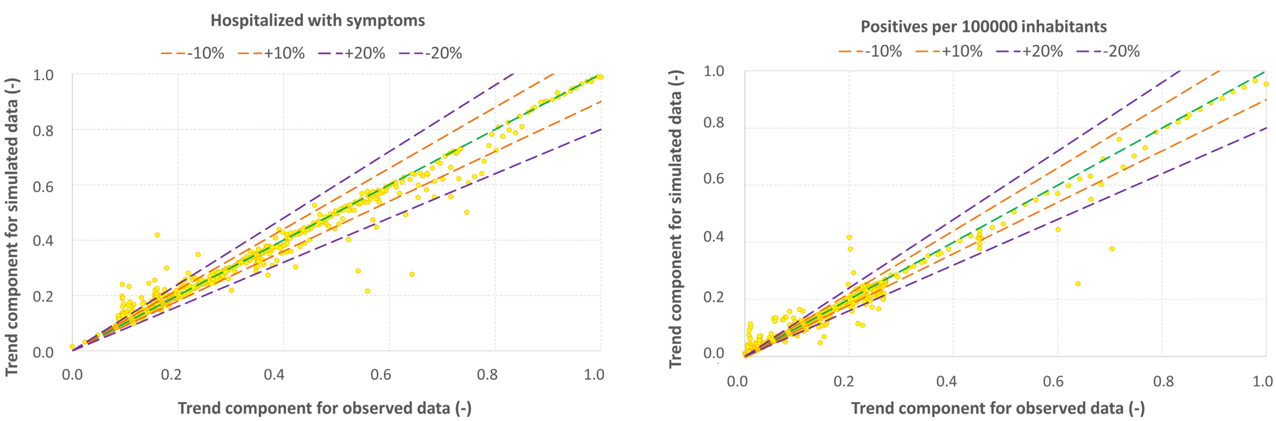
(a) (b)

FIGURE 5 | Trend component analysis between observed (in blue) and simulated (in magenta) for the number of hospitalized with symptoms (a) and the positives per 100,000 inhabitants (b) in the Emilia-Romagna region.



(a) (b)

FIGURE 6 | Trend component analysis between observed (in blue) and simulated (in magenta) for the number of hospitalized with symptoms (a) and the positives per 100,000 inhabitants (b) in the Puglia region.



(a) (b)

FIGURE 7 | Scatterplot of the trend component between simulated vs. observed data for the number of hospitalized patients with symptoms (a) and positives per 100,000 inhabitants (b) in the Lombardia region.

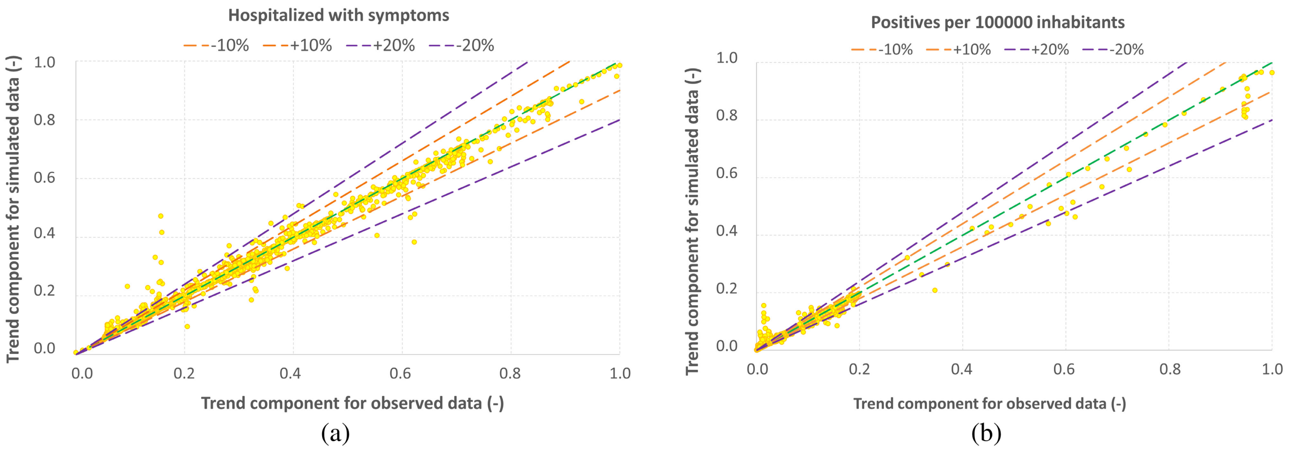


FIGURE 8 | Scatterplot of the trend component between simulated vs. observed data for the number of hospitalized patients with symptoms (a) and positives per 100,000 inhabitants (b) in the Emilia-Romagna region.

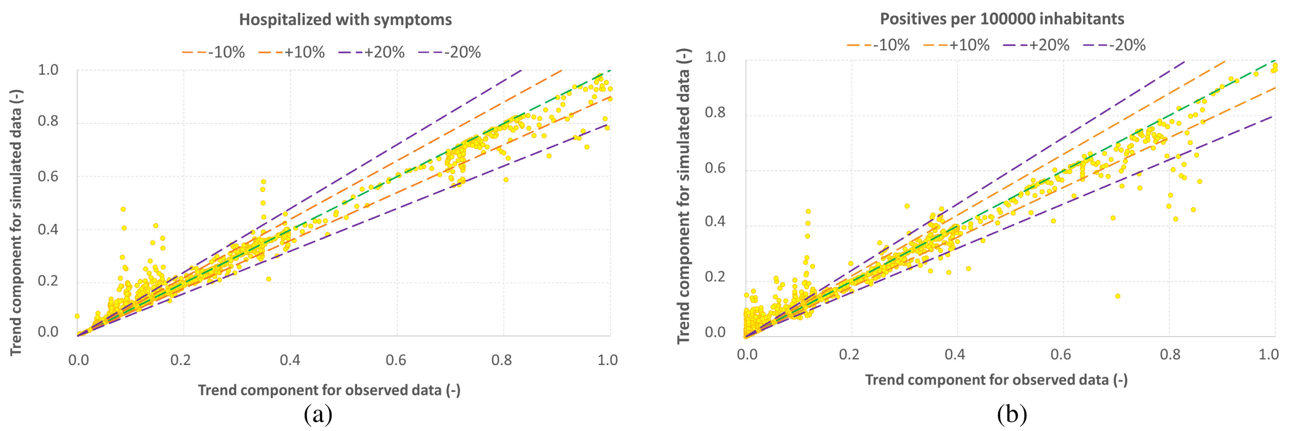


FIGURE 9 | Scatterplot of the trend component between simulated vs. observed data for the number of hospitalized patients with symptoms (a) and positives per 100,000 inhabitants (b) in the Puglia region.

TABLE 5 | Percentage of data points falling within the ±5% error range for the number of hospitalized patients with symptoms and positives per 100,000 inhabitants for Lombardia, Emilia-Romagna, and Puglia regions.

Region	Hospitalized patients with symptoms	Positives per 100,000 inhabitants
Lombardia	78%	58%
Emilia-Romagna	71%	52%
Puglia	63%	42%

Lastly, the scatterplots in Figures 7–9 summarize the agreement between observed and simulated data for the two target variables under investigation; the error bands shown as confidence levels denote the RF model performance.

Over the three regions, most of the points fall within the error bands of ±20%, with many even in the ±10% range for both the target variables. Over the Lombardia and Emilia-Romagna regions, it is clear how the points cluster around the bisector, and only a few of them are outside the confidence levels, while in the Puglia region, a greater number of dots fall outside the ±20% bands.

In addition, Table 5 shows an in-depth analysis considering an error range of ±5% over the three regions for the number of hospitalized patients with symptoms and positives per 100,000 inhabitants, with a percentage of points in Lombardia equal to 78% and 58%, respectively, while in Emilia-Romagna, it is 71% and 52%; only in Puglia, the percentage drops to 63% and 42%. Nevertheless, the majority of points are generally distributed within the confidence levels across the three regions, demonstrating the robustness of the implemented Random Forest model.

4.3 | The SHAP Method

The satisfactory results prompt the analysis towards the implementation of the Feature Importance (FI) method (Shaikhina et al. 2021; Kumar et al. 2020). Through this calculation procedure, a weight (or importance) is assigned to each meteorological variable in the entire machine learning process over the three regions (Tables 5–7). This ultimate step involves the obtained predictors as a weight of each variable in defining the

TABLE 6 | SHAP Feature Importance for the Lombardia region; the number of hospitalized patients with symptoms (a) and positives per 100,000 inhabitants (b).

Relative importance (a)		Relative importance (b)	
DEW	0.426	DEW	0.336
RAD	0.229	T	0.180
W	0.161	W	0.176
BARO	0.074	BARO	0.146
RH	0.061	RH	0.116
T	0.047	RAD	0.043

TABLE 7 | SHAP Feature Importance for the Emilia Romagna region; the number of hospitalized patients with symptoms (a), positives per 100,000 inhabitants (b).

Relative importance (a)		Relative importance (b)	
DEW	0.471	DEW	0.414
RAD	0.213	T	0.168
W	0.095	W	0.145
BARO	0.090	BARO	0.134
RH	0.078	RH	0.087
T	0.053	RAD	0.052

predictive model to be evaluated. In particular, it was examined how much each independent (meteorological) variable affects the dependent (target) variable. This analysis is carried out using the SHAP (SHapley Additive exPlanations) method (Fryer et al. 2021; Roder et al., 2021), which is a consolidated approach for explaining machine learning models.

The results of the SHAP Feature Importance for hospitalized patients with symptoms (Table 6a) in Lombardia show that the most affecting weather variables in the transmission of COVID-19 are the DEW, RAD, and W, since the normalized values of these six variables sum up to 70% of influence, while the meteorological factors that most influence the transmission of COVID-19 for positives per 100,000 inhabitants (Table 6b) are DEW, T, and W. The same outcomes are found for the Emilia-Romagna and Puglia regions, as shown in Tables 7 and 8; in all cases, the dew point turns out to have the main importance.

In addition, to evaluate the importance of lag time on the impact of dew point with the spread of COVID-19 and its associated Relative Risk (RR), a DLNM (Distributed Lag Non-Linear Model) is implemented following the methodology proposed by Gasparrini (2011), Gasparrini (2021) and Gasparrini 2021, with a maximum lag here set at 10 days. The RR is an index that quantifies the probability of an occurring event in the presence of a specific exposure compared to its absence (McNutt et al. 2003). It is calculated by comparing the event probability (e.g., an increase in COVID-19 cases) associated with a specific level of exposure (in our case a given dew point value) to the probability of the event under reference conditions (e.g., an average value of

TABLE 8 | SHAP Feature Importance for the Puglia region; the number of hospitalized patients with symptoms (a), positives per 100,000 inhabitants (b).

Relative importance (a)		Relative importance (b)	
DEW	0.501	DEW	0.451
RAD	0.155	T	0.158
W	0.141	W	0.127
BARO	0.073	BARO	0.100
RH	0.069	RH	0.094
T	0.061	RAD	0.071

the meteorological variable). In building the model, penalized splines (Economou et al. 2024) are used to optimize the degrees of freedom for both meteorological variables and lag times. These optimized parameters are then applied to the modeling process, generating graphs that describe the RR as a function of lag time variations.

By the graphs shown in Figure 10 with a particular reference to dew point values in the Lombardia region, the lag times associated with the highest RR differ between the two cases. Regarding hospitalized patients with symptoms (Figure 10a), the highest RR is observed for lags between 0 and 1 day, while for positives per 100,000 inhabitants (Figure 10b), significant lags are found between 0–1, 4–6, and around 10 days; it is relevant to note the highest RR scores are associated when low dew point temperatures are observed.

5 | Discussion

The COVID-19 infection, caused by the novel coronavirus SARS-CoV-2, was declared a pandemic by the WHO in March 2020 after the virus was identified in Wuhan (China) at the end of 2019. Since it has been one of the most serious global health crises in recent decades, assessing the impact of meteorological factors on the spread of COVID-19 led the scientific community to investigate several methods and analyses in recent years (Yuan et al. 2021; Balboni et al. 2023; Tan and Schultz 2023). As mentioned in the introduction, different achievements were found, but no general conclusions were obtained around the world.

In this paper, we analyzed common meteorological variables over three regions of Italy: Lombardia, Emilia-Romagna, and Puglia, using citizen science meteorological values by Meteonetwork and epidemiological data by the national civil protection database.

Initially, a preliminary correlation analysis between meteorological data and the COVID-19 values was implemented, but a weak outcome between the two types of analyzed data was obtained. Thus, the STL decomposition model was introduced to decompose both meteorological data and COVID-19 infection-related values into their three components: seasonality, trend, and residuals.

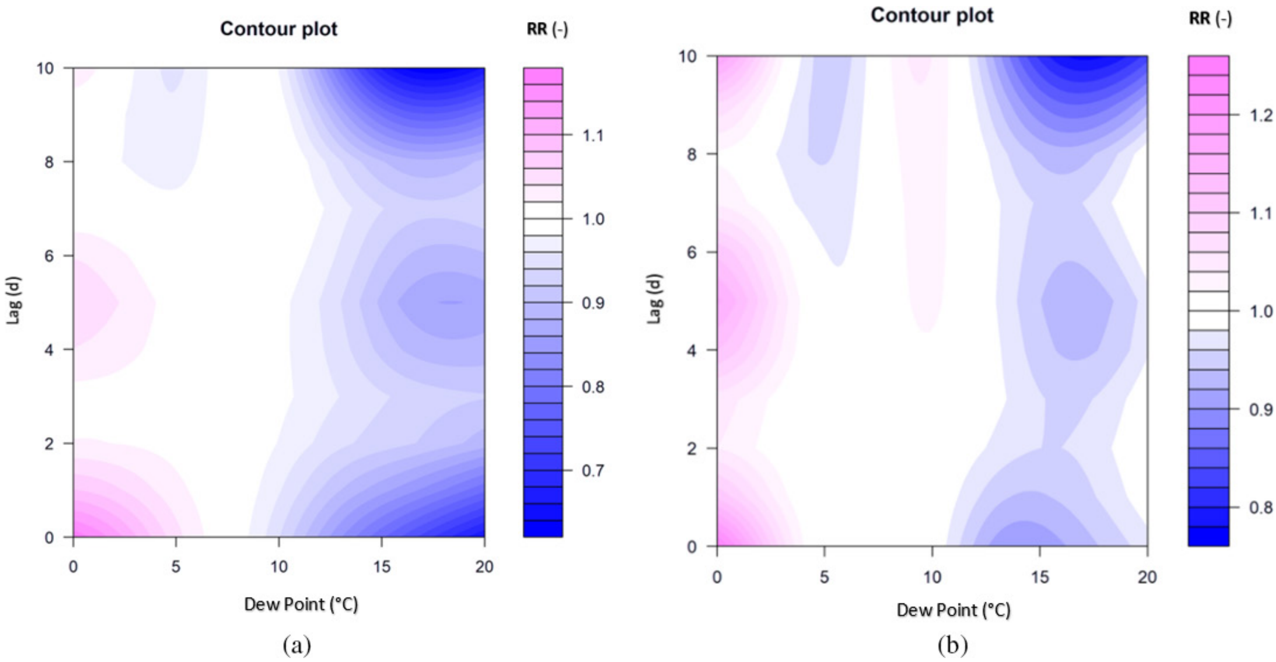


FIGURE 10 | Contour plot of the relationship between dew point, lag time, and relative risk (RR) for the number of hospitalized patients with symptoms (a) and positives per 100,000 inhabitants (b) in the Lombardia region.

Removing noise from a time series is crucial to highlight the trend, improving long-term data analysis. Without noise, it becomes easier to identify regular patterns and enhance the prediction accuracy. This process not only facilitates the understanding of relationships between variables and system dynamics, but also reduces the risk of overfitting in machine learning models, making them more robust and suitable for generalizing new data. Hence, the machine learning model (RF) was again run to simulate the trend of observed values compared to a randomly selected sample of data. After introducing the STL, it was observed that: (i) the predicted data well simulate the trend of the observed data; (ii) analyzing the results of the k-fold cross-validation technique, the R^2 and MAE were, respectively, close to 1 and 0, indicating the applied model is robust; (iii) the problem of overfitting was not found, since after the model test phase, it was observed that the average metrics of the k-fold cross-validation are in line with, and even better than those initially obtained by the predictive RF model; (iv) analyzing the scatterplots in Figures 7–9, most of the points representing the observed vs. simulated data are within ±10% of the error bands, with more than 75% in the ±5% confidence levels, and very few out of the ±20% range, mainly in Lombardia and Emilia-Romagna regions, while a decrease is observed in Puglia for data in the error range of ±5%, where the percentage drops to 63% for hospitalized patients with symptoms and to 42% for positives per 100,000 inhabitants, respectively.

Nevertheless, the main finding of the study emerges from the Feature Importance (FI) analysis. In fact, it is significant to observe that regarding both the hospitalized patients with symptoms and positives per 100,000 inhabitants in the three examined regions, the variable that consistently shows the greatest impact on the spread of COVID-19 is the dew point.

The dew point is defined as the temperature to which the air must be cooled to reach saturation (at constant pressure), that is, to achieve 100% relative humidity, causing the contained water vapor to condense into dew or fog. When this phenomenon occurs, the dew point is equal to the air temperature, indeed bringing humidity to saturation. This achievement reveals the fundamental role of atmospheric moisture in the spread of COVID-19: if the water vapor is closer to saturation, the virus would have a greater persistence in the air. The trapping of droplets by humidity creates, in fact, ideal conditions for virus transmission among people.

Hence, high values of relative humidity could theoretically play an important role in the spread of COVID-19. In reality, factors influencing the virus spread are numerous and interconnected, making it difficult to disentangle them as predominant, especially considering that relative humidity does not show high feature importance in the analyzed cases. On the other hand, contagion strongly depends on the sharing of indoor environments by individuals as well, which is frequent in winter time, and significantly reduces during the summer season, characterized by higher temperature and radiation. As an example for hospitalized patients with symptoms in the Lombardia region, Figure 11 depicts the SHAP dependence plot of the dew point in relation to the air temperature values (both expressed in dimensionless form) and reveals that the highest SHAP index values (indicative of the model's predictive performance) occur at low temperatures (shown in blue points) with low dew point values, generally, in the colder seasons.

This finding is not so clear-cut in the same graph of the dew point in relation to relative humidity values, still in dimensionless form (Figure 12). Nevertheless, SHAP values are prevalently greater when higher levels (magenta colors) of relative humidity occur at low dew points (in the colder months). This score

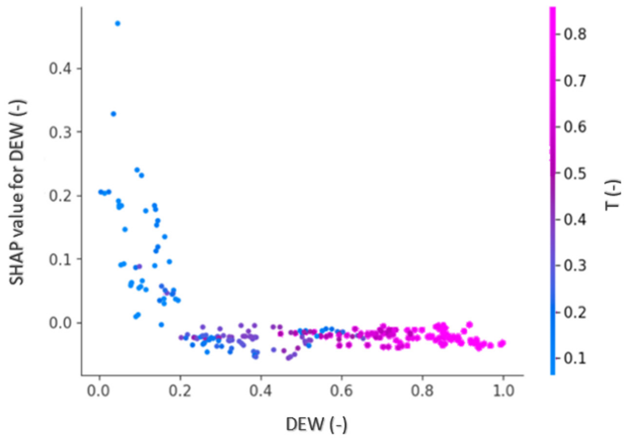


FIGURE 11 | SHAP dependence plot: Relationship between the dew point and temperature values. Blue points indicate low-temperature ranges, while magenta points indicate high-temperature intervals (both in dimensionless form).

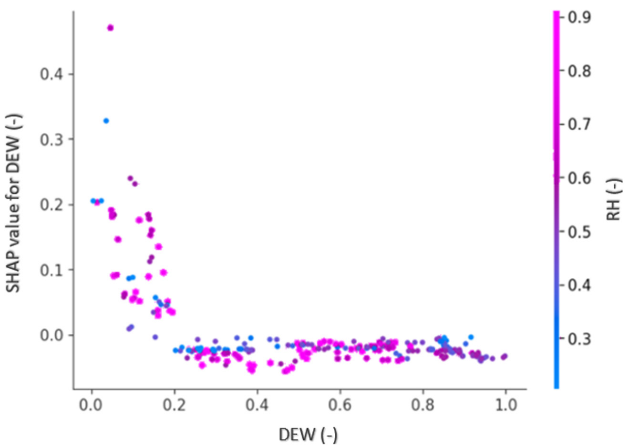


FIGURE 12 | SHAP Dependence Plot: Relationship between dew point and relative humidity. Blue points indicate low relative humidity ranges, while magenta points indicate high intervals of relative humidity (in dimensionless form).

underlines an original and significant result, which is also useful in guiding future research.

5.1 | Limitations of the Study

The proposed study presents some methodological limitations which deserve attention. In particular, no temporal distinction has been made between the pre- and post-vaccine periods or pre- and post-curfew periods. This choice is motivated by the need to retain a sufficient quantity of data for applying machine learning techniques. Dividing the analysis period into too many small segments would have shortened the availability of data for each interval, compromising the reliability and robustness of the results. Similarly, the vaccination rate has not been explicitly considered. Including this variable would have required a more detailed fragmentation of data and a temporal division of the analyzed period, further reducing the amount of data accessible for robust training of models. A similar issue arises with age, particularly in the presence of multiple comorbidities. Although it is a relevant variable, it has not been explicitly included in the model. Therefore, the decision to not explicitly consider age and comorbidities fits within a methodological approach aimed at maintaining a large dataset ensuring robust and predictive models. However, a future hybrid statistical-probabilistic model combined with machine learning techniques could be implemented to overcome this limitation related to big data, allowing for a more nuanced analysis while maintaining the robustness of the predictive models.

An additional limitation concerns the absence of a specific investigation of air pollutants. Sensors of monitoring stations used in this study are exclusively meteorological; consequently, the analysis focuses on these aspects only. However, in the future, to provide a more comprehensive assessment, air pollutant data can be integrated. For instance, a further aspect that could be considered for future developments is the inclusion of the atmospheric boundary layer height (BLH) as an additional meteorological factor potentially influencing the spread of COVID-19. This parameter has been analyzed in some previous studies that compared meteorological conditions during and before the pandemic in order to assess the impact of meteorological variability during those periods (Petetin et al. 2020; García-Dalmau et al. 2022). BLH is, in fact, a classical variable considered in the atmospheric dispersion of air pollutants; thus, it could also play a role in the spread of COVID-19. In the same way, a limit regarding possible biological justifications in the relationship between meteorological factors and the spread of COVID-19 has not been explored, as biological factors fall outside the aim of our study.

Most of the methodological choices are here made to maximize the robustness and reliability of predictive models. However, the possibility of expanding and further exploring these aspects in upcoming research is surely a future development to investigate.

6 | Conclusion

The current study analyzed the influence of meteorological factors on the COVID-19 contagion over three Italian regions with different climatic conditions (Lombardia, Emilia-Romagna, and Puglia).

First, we conducted a preliminary correlation analysis between the daily average weather data and health target variables (number of individuals hospitalized with symptoms and number of positive cases per 100,000 inhabitants), followed by a data filtering analysis using the STL decomposition. The trend values determined through the STL function were subject to a second investigation, revealing a good correlation between the meteorological and epidemiological data.

To carry on the simulations, the Random Forest algorithm, which is a regression model based on machine learning algorithms, was implemented and evaluated through the calculation of skill scores, even with a k-fold cross-validation procedure. Values of R^2 and MAE, close to 1 and 0, respectively, highlighted the accuracy and robustness of the model in simulating the measured data in the entire observation period between February 24, 2020, and August 31, 2023.

Finally, the importance of each weather variable was assessed using the SHAP methodology, finding the dew point as the most significant meteorological parameter for promoting virus survival (and spread) in the environment.

The conducted analyses highlight that the highest SHAP index values related to dew point occur at low temperatures and high relative humidity levels, particularly when the dew point is low (typically in the colder months). This conclusion is further emphasized by the evaluation of the relative risk associated with lag time: the application of the DLNM model, in fact, reveals that, both for hospitalized patients with symptoms and for positives per 100,000 inhabitants, the greatest risk of COVID-19 spread occurs when the dew point values are low.

Although this investigation has been carried out in three regions of Italy only, and future developments with additional analyses may confirm this hypothesis, the study provides important insights into the spread of COVID-19, emphasizing the importance of considering meteorological factors in designing prevention and control strategies for the virus diffusion, even for other pathologies, particularly involving the respiratory system.

Author Contributions

Vito Telesca: methodology, investigation, conceptualization, data curation, writing – original draft, writing – review and editing, supervision. **Gianfranco Castronuovo:** methodology, software, data curation, writing – original draft. **Gianfranco Favia:** methodology, investigation, writing – original draft. **Mariarosaria Marra:** data curation, writing – original draft. **Marica Rondinone:** software, formal analysis, data curation, writing – original draft, writing – review and editing. **Alessandro Ceppi:** data curation, conceptualization, supervision, writing – original draft, writing – review and editing.

Acknowledgments

Open access publishing facilitated by Universita degli Studi della Basilicata, as part of the Wiley - CRUI-CARE agreement.

Consent

The data provided does not contain any personal information about patients.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Antolini, G., V. Marletto, V. Pavan, and R. Tomozeiu. 2017. Atlante climatico dell´Emilia-Romagna (edizione 2017). Arpae Emilia-Romagna.

Balboni, E., T. Filippini, K. J. Rothman, et al. 2023. "The Influence of Meteorological Factors on COVID-19 Spread in Italy During the First and Second Wave." *Environmental Research* 228: 115796. https://doi.org/10.1016/j.envres.2023.115796.

Barredo, A. A., N. Díaz-Rodríguez, J. Del Ser, et al. 2020. "Explainable Artificial AI." *Information Fusion* 58: 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Chan, K. H., J. S. M. Peiris, S. Y. Lam, L. L. M. Poon, K. Y. Yuen, and W. H. Seto. 2011. "The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus." *Advances in Virology* 2011: 734690. https://doi.org/10.1155/2011/734690.

Chicco, D., M. J. Warrens, and G. Jurman. 2021. "The Coefficient of Determination R-Squared is More Informative than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation." *PeerJ Computer Science* 7, no. 3: e623. https://doi.org/10.7717/peerj-cs.623.

Chin, A. W. H., J. T. S. Chu, M. R. A. Perera, et al. 2020. "Stability of SARS-CoV-2 in Different Environmental Conditions." *Lancet Microbe* 1, no. 1: e10. https://doi.org/10.1016/S2666-5247(20)30003-3.

Cleveland, R. B., W. S. Cleveland, J. E. Mc Rae, and I. Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Based on Loess." *Journal of Official Statistics* 6: 3–73.

Demongeot, J., Y. Flet-Berliac, and H. Seligmann. 2020. "Temperature Decreases Spread Parameters of the New Covid-19 Case Dynamics." *Biology* 9, no. 5: 94. https://doi.org/10.3390/biology905004.

Economou, T., D. Parliari, A. Tobias, et al. 2024. "A Unifying Modelling Approach for Hierarchical Distributed Lag Models." *arXiv*. https://doi.org/10.48550/arXiv.2407.13374.

Fryer, D., I. Strümke, and H. Nguyen. 2021. "Shapley Values for Feature Selection: The Good, the Bad, and the Axioms." *IEEE Access* 9: 144352–144360. https://doi.org/10.1109/ACCESS.2021.3119110.

García-Dalmau, M., M. Udina, J. Bech, Y. Sola, J. Montolio, and C. Jaén. 2022. "Pollutant Concentration Changes During the COVID-19 Lockdown in Barcelona and Surrounding Regions: Modification of Diurnal Cycles and Limited Role of Meteorological Conditions." *Boundary-Layer Meteorology* 183, no. 2: 273–294. https://doi.org/10.1007/s10546-021-00679-1.

Gasparrini, A. 2011. "Distributed Lag Linear and Non-Linear Models in R: The Package Dlnm." *Journal of Statistical Software* 43, no. 8: 1–20. https://doi.org/10.18637/jss.v043.i08.

Gasparrini, A. 2021. "Distributed Lag Linear and Non-Linear Models for Time Series Data." London School of Hygiene & Tropical Medicine, dlnm Version 2.4.7, 2021-10-07. https://rdrr.io/cran/dlnm/.

Geller, C., M. Varbanov, and R. E. Duval. 2012. "Human Coronaviruses: Insights Into Environmental Resistance and Its Influence on the Development of New Antiseptic Strategies." *Viruses* 4: 3044–3068. https://doi.org/10.3390/v4113044.

Giazzi, M., G. Peressutti, L. Cerri, et al. 2022. "Meteonetwork: An Open Crowdsourced Weather Data System." *Atmosphere* 13, no. 6: 928. https://doi.org/10.3390/atmos13060928.

Gogtay, N. J., and U. M. Thatte. 2017. "Principles of Correlation Analysis." *Journal of Association Physicians India* 65, no. 3: 78–81. https://www.kem.edu/wp-content/uploads/2012/06/9-Principles_of_correlation-1.pdf.

Gupta, S., S. S. Hayek, W. Wang, et al. 2020. "Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US." *JAMA Internal Medicine* 180, no. 11: 1436–1447. https://doi.org/10.1001/jamainternmed.2020.3596.

Iqbal, M. M., I. Abid, S. Hussain, N. Shahzad, M. S. Waqas, and M. J. Iqbal. 2020. "The Effects of Regional Climatic Condition on the Spread of COVID-19 at Global Scale." *Science of the Total Environment* 739: 140101. https://doi.org/10.1016/j.scitotenv.2020.140101.

ISTAT. 2024. "Italian National Institute of Statistics." (in Italian, https://www.istat.it/).

Kumar, I. E., S. Venkatasubramanian, C. Scheidegger, and S. Friedler. 2020. "Problems With Shapley-Value-Based Explanations as Feature

Importance Measures." In *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, 5491–5500. PMLR. https://proceedings.mlr.press/v119/kumar20e.html.

Mangalathu, S., S. H. Hwang, and J.-S. Jeon. 2020. "Failure Mode and Effects Analysis of RC Members Based on Machine-Learning-Based SHapley Additive exPlanations (SHAP) Approach." *Engineering Structures* 219: 110927. https://doi.org/10.1016/j.engstruct.2020.110927.

McNutt, L.-A., C. Wu, X. Xue, and J. P. Hafner. 2003. "Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes." *American Journal of Epidemiology* 157, no. 10: 940–943. https://doi.org/10.1093/aje/kwg074.

Mecenas, P., R. T. R. M. Bastos, A. C. R. Vallinoto, and D. Normando. 2020. "Effects of Temperature and Humidity on the Spread of COVID-19: A Systematic Review." *PLoS One* 15, no. 9: e0238339. https://doi.org/10.1371/journal.pone.0238339.

Nohara, Y., K. Matsumoto, H. Soejima, and N. Nakashima. 2022. "Explanation of Machine Learning Models Using Shapley Additive Explanation and Application for Real Data in Hospital." *Computer Methods and Programs in Biomedicine* 214: 106584. https://doi.org/10.1016/j.cmpb.2021.106584.

Petetin, H., D. Bowdalo, A. Soret, et al. 2020. "Meteorology-Normalized Impact of the COVID-19 Lockdown Upon NO2 Pollution in Spain." *Atmospheric Chemistry and Physics* 20: 11119–11141. https://doi.org/10.5194/acp-20-11119-2020.

Pirouz, B., S. Shaffiee Haghshenas, B. Pirouz, S. Shaffiee Haghshenas, and P. Piro. 2020b. "Development of an Assessment Method for Investigating the Impact of Climate and Urban Parameters in Confirmed Cases of COVID-19: A New Challenge in Sustainable Development." *International Journal of Environmental Research and Public Health* 17, no. 8: 2801. https://doi.org/10.3390/ijerph17082801.

Pirouz, B., S. Shaffiee Haghshenas, S. Shaffiee Haghshenas, and P. Piro. 2020a. "Investigating a Serious Challenge in the Sustainable Development Process: Analysis of Confirmed Cases of COVID-19 (New Type of Coronavirus) Through a Binary Classification Using Artificial Intelligence and Regression Analysis." *Sustainability* 12, no. 6: 2427. https://doi.org/10.3390/su12062427.

Prata, D. N., W. Rodrigues, and P. H. Bermejo. 2020. "Temperature Significantly Changes COVID-19 Transmission in (Sub)tropical Cities of Brazil." *Science of the Total Environment* 729: 138862. https://doi.org/10.1016/j.scitotenv.2020.138862.

Qi, H., S. Xiao, R. Shi, et al. 2020. "COVID-19 Transmission in Mainland China Is Associated With Temperature and Humidity: A Time-Series Analysis." *Science of the Total Environment* 728: 138778. https://doi.org/10.1016/j.scitotenv.2020.138778.

Regional Environmental Protection Agency (ARPA). 2024. "Climate in Lombardia." (in Italian, https://www.arpalombardia.it/temi-ambientali/meteo-e-clima/clima/il-clima-in-lombardia/).

Regional Environmental Protection Agency (ARPA). 2024. "Climate Atlas of Emilia-Romagna." (in Italian, https://www.arpae.it/it/temi-ambientali/clima/rapporti-e-documenti/atlante-climatico).

Rodriguez-Galiano, V., M. Sanchez-Castillo, M. Chica-Olmo, and M. Chica-Rivas. 2015. "Machine Learning Predictive Models for Mineral Prospectivity: An Evaluation of Neural Networks, Random Forest, Regression Trees and Support Vector Machines." *Ore Geology Reviews* 71: 804–818. https://doi.org/10.1016/j.oregeorev.2015.01.001.

Sahin, M. 2020. "Impact of Weather on COVID-19 Pandemic in Turkey." *Science of the Total Environment* 728: 138810. https://doi.org/10.1016/j.scitotenv.2020.138810.

Shaikhina, T., U. Bhatt, R. Zhang, K. Georgatzis, A. Xiang, and A. Weller. 2021. "Effects of Uncertainty on the Quality of Feature Importance Explanations." In *AAAI Workshop on Explainable Agency in Artificial Intelligence*. AAAI Press. https://umangsbhatt.github.io/reports/AAAI_XAI_QB.pdf.

Shetty, S. H., S. Shetty, C. Singh, and A. Rao. 2022. "Supervised Machine Learning: Algorithms and Applications." In *Fundamentals and Methods of Machine and Deep Learning: Algorithms, Tools, and Applications*, edited by P. Singh, 1–16. Wiley. https://doi.org/10.1002/9781119821908.ch1.

Shi, P., Y. Dong, H. Yan, et al. 2020. "Impact of Temperature on the Dynamics of the COVID-19 Outbreak in China." *Science of the Total Environment* 728: 138890. https://doi.org/10.1016/j.scitotenv.2020.138890.

Tan, L., and D. M. Schultz. 2023. "Weather Effects on the Spread of COVID-19: Characteristics and Critical Analysis of the First and Second Years of Scientific Research." *Bulletin of the American Meteorological Society* 104: E1345–E1371. https://doi.org/10.1175/BAMS-D-23-0071.1.

van Doremalen, N., T. Bushmaker, and V. J. Munster. 2013. "Stability of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) Under Different Environmental Conditions." *Euro Surveillance* 18, no. 38: 20590. https://doi.org/10.2807/1560-7917.ES2013.18.38.20590.

Vinoj, V., N. Gopinath, K. Landu, B. Behera, and B. Mishra. 2020. "The COVID-19 Spread in India and Its Dependence on Temperature and Relative Humidity. Preprints.Org." https://doi.org/10.20944/preprints202007.0082.v1.

Wu, Y., W. Jing, J. Liu, et al. 2020. "Effects of Temperature and Humidity on the Daily New Cases and New Deaths of COVID-19 in 166 Countries." *Science of the Total Environment* 729: 139051. https://doi.org/10.1016/j.scitotenv.2020.139051.

Yuan, J., Y. Wu, W. Jing, et al. 2021. "Association Between Meteorological Factors and Daily New Cases of COVID-19 in 188 Countries: A Time Series Analysis." *Science of the Total Environment* 780: 146538. https://doi.org/10.1016/j.scitotenv.2021.146538.

Zhu, Y., J. Xie, F. Huang, and L. Cao. 2020. "Association Between Short-Term Exposure to Air Pollution and COVID-19 Infection: Evidence From China." *Science of the Total Environment* 727: 138704. https://doi.org/10.1016/j.scitotenv.2020.138704.