

The Method for Optimum Estimation of COVID-19 Variant Type Virus Infection Status Analysis by the Multivariate Analysis Considering the Environmental Variability Impact in Japan

Eiji Toma, Yukinori Kobayashi

National Institute of Technology, Tomakomai College, Tomakomai, Japan

Email: toma-e@tomakomai-ct.ac.jp, kobay@tomakomai-ct.ac.jp

How to cite this paper: Toma, E. and Kobayashi, Y. (2022) The Method for Optimum Estimation of COVID-19 Variant Type Virus Infection Status Analysis by the Multivariate Analysis Considering the Environmental Variability Impact in Japan. *Journal of Applied Mathematics and Physics*, 10, 425-448.

<https://doi.org/10.4236/jamp.2022.102033>

Received: January 2, 2022

Accepted: February 20, 2022

Published: February 23, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Currently, the estimated value of the effective reproduction number (ERN), which is an index for grasping the COVID-19 infection status, is used for important planning and evaluation of infection prevention measures. Since ERN in the Sequential SIR model fluctuates in multiple dimensions due to changes in the surrounding environment, it is difficult to set the appropriate accuracy of the uncertainty region of the estimated data. The challenge in this study is to build a mathematical model of infectious disease according to the characteristics and data characteristics of the infectious disease and select an appropriate estimation method. Highly accurate quantitative research that analyzes the validity of “how infectious diseases prevail” from an academic point of view is the key to prediction and estimation in appropriate infection situation analysis. In this study, we adopted a statistical multivariate analysis method (T method) that enables evaluation and prediction of important factors related to ERN estimation and analysis of phenomena that change in real time (time series analysis). It was clarified that it is possible to estimate with higher accuracy by applying the T method to the estimated value of ERN by the current SIR mathematical model.

Keywords

COVID-19, Sequential SIR Model, Effective Reproduction Number, Multivariate Analysis Method, T-Method, Regression Analysis

1. Introduction

Currently, the infectious disease sequential SIR model, which expresses various phenomena by ordinary differential equations, plays an important role in the ongoing fight against COVID-19 and various other viral infections around the world. The purpose of the sequential SIR model is to eradicate the infection if it occurs in a population that was previously uninfected, whether it causes a sustained increase in the number of infected people. It is an infection situation analysis such as what kind of intervention is effective. An important index for developing infection status analysis is the effective reproductions number. The effective reproductions number is an index showing “how many people are infected by one infected person on average”. The higher this number is, the more rapidly the infection is spreading, and if the period of less than 1 continues, it can be said that the infection is converging. This estimated effective reproductions number is used for important planning and evaluation of infection prevention measures implemented.

The estimated effective reproductions number depends largely on virus infectivity. The coefficient of determination fluctuates in multiple dimensions due to changes in the environment (human intervention such as behavioral restriction, differences in the surrounding environment such as climate change, and an increase in the number of immune carriers as the infection progresses). Therefore, there is an academic problem that it is difficult to set the appropriate accuracy of the uncertainty region of the estimated data. In this study, we propose a new infection situation analysis method by highly accurate quantitative analysis applying multivariate analysis method regarding the validity of how the infectious disease spreads.

2. Basics of Infectious Disease Mathematical Model

In pandemic under emerging infectious diseases such as COVID-19, pandemic data analysis and scenario analysis with infection mathematical model are important evidence that forms the nucleus of the policy decision. The infectious disease mathematical model describes the spread of infectious diseases in a population with mathematical formulas. Mathematical epidemiology of infectious diseases is a research field with a long history dating back to the 18th century, but in recent years, with the development of computer processing power and computational statistics, research methods for its social implementation have dramatically advanced [1]. Although it has been rarely taken up in Japan until now, it is a research field that has attracted attention due to the worldwide pandemic of COVID-19.

In this section, in addition to the epidemiological findings of COVID-19, the basic concept of the mathematical model of infectious diseases is described.

2.1. Sequential SIR Model

The basic mathematical model that captures the pandemic dynamics of infec-

tious diseases that propagate directly from person to person is called “Kermack and McKendrick (1927)”, and is also called the “SIR model” from the variables S , I , and R . Each letter of SIR is an acronym for English, dividing the population into three compartments, susceptibility, infectivity, isolation and recovery, according to the stage of infection, and the temporal changes in the status of infection. It is a bottom-up model [2]. **Figure 1** shows a schematic diagram of the SIR model. The simplest SIR model is described by the ordinary differential equation shown in Equation (1).

$$\begin{aligned}\frac{dS(t)}{dt} &= -\beta S(t)I(t) \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \gamma I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t)\end{aligned}\quad (1)$$

Here, $S(t)$, $I(t)$, and $R(t)$ represent the proportion of susceptible, infected, and recovered/quarantined individuals at a given time in the population. β is a coefficient representing the infection rate per unit time. $\beta I(t)$ gives the force of infection at time t . In other words, when the population is constant, the infectivity that is a hazard is proportional to the number of infected people in the population. γ is the rate of removal by recovery or quarantine per unit time, and the reciprocal γ^{-1} gives the average infectivity period from infection to recovery or quarantine.

Here, by transforming the second Equation (1), the following Equation (2) is obtained.

$$\frac{dI(t)}{dt} = (\beta S(t) - \gamma)I(t) \quad (2)$$

when the number of newly infected people is increasing, $\beta S(t) - \gamma > 0$, which is a condition for infectious disease pandemics. Furthermore, this Equation (2) can be transformed into $\beta S(t)/\gamma > 1$. At time 0, since all members of the population are susceptible populations, $S(0) = 1$ can be set, and β/γ is the threshold for infectious disease pandemics. This value is called the basic reproduction number (R_0), which is cited as an index of the strength of virus infectivity and is given as a reference value at the initial stage of infection.

2.2. Basic and Effective Reproduction Numbers

Basic reproduction number (R_0) is the most basic index of infectious disease in infectious disease epidemiology. Interpreted as the average number of secondary infections that a typical infected person reproduces during the infectious period



Figure 1. Schematic diagram of SIR model.

in a population that is all susceptible to an infectious disease. If the basic reproduction number is greater than 1, a large-scale pandemic can occur, but if it is less than 1, the pandemic disappears spontaneously. Since the value of the basic reproduction number is related to the population density, social structure, and contact mode between individuals, the estimated value varies depending on the situation of the area to be analyzed. The basic reproduction number of COVID-19 in China is estimated to be 1.5 to 3.5 from the data of the early stage of the pandemic [3].

It is known that COVID-19 has a high degree of heterogeneity related to secondary infection, and the distribution of secondary infections reproduced by one infected person varies widely. In other words, most infected people do not produce secondary infections, and some infected people become super spreaders, producing many secondary infections. A study analyzing epidemiological data outside of China estimated that 80% of infected people originated from a small number of infected people (~10%) [4]. In addition, there are many asymptomatic and mildly ill patients, and the occurrence of a large number of secondary infections regardless of their severity is one of the factors that make infection control difficult [5].

On the other hand, as a situation where many secondary infections occur, the environment where they are densely gathered in a closed space has been clarified from the observation data [6]. It is also an infectious disease that can suppress the pandemic by thoroughly avoiding such an environment. The number of reproductions under the implementation of countermeasures is called the effective reproduction number (hereinafter referred to as ERN), and setting the ERN to 1 or less is a guideline for infectious disease control. The basic reproduction number (R_0) can be decomposed and considered as shown on the right side of Equation (3) [7] [8] [9].

$$R_0 = cbd \quad (3)$$

Here, c is the average number of effective contacts (rate) that one person makes effective contact per unit time, b is the probability of infection from infected person to susceptible person per effective contact, and d is the average infectivity period. ERN (R_s) is expressed by the following Equation (4), where p ($0 < p < 1$) is the effect coefficient (decrease rate) of infectivity due to measures against infectious diseases such as wearing a mask, washing hands, and reducing contact.

$$R_s = \frac{S\beta}{\gamma}(1-p) = \frac{S}{N}(1-p)R_0 \quad (4)$$

The estimated value of R_s from Equation (4) largely depends on the virus infectivity R_0 . However, γ , β , and p are unknown coefficients that fluctuate due to changes in the environment (human intervention such as behavioral restriction, differences in the surrounding environment such as climate change, and an increase in immune carriers as infection progresses). In this respect, there is a

practical problem that it is difficult to set the appropriate accuracy of the uncertainty region of the estimated data [10].

Intervention effect Vaccination also can be considered similarly. Let the total population be 1 and the vaccination ratio of the population be x . Assuming that the vaccinated person can be uniformly immunized, ERN (R_S) can be expressed by replacing the right-hand side p of Equation (4) with x . Conditions for infection eradication is to ERN (R_S) is below 1, $(1-x)R_0 < 1$. The ratio $(1-1/R_0)$ when this equation is solved for x is called the critical immunity ratio and is often used as the target value for the vaccination rate [11].

When actually monitoring ERN using observational data in an infectious disease pandemic, various reporting biases need to be considered. A typical example is the delay in the time from infection until the test is actually positive and the person is reported as infected. Since the number of infected persons currently observed is less than that of actual infected persons, the current situation is that the reporting delay is adjusted by statistical estimation [12].

3. Introduction of Multivariate Analysis Method

The purpose of this study is to analyze the validity of “how infectious diseases prevail” from an academic point of view. By building a new infection situation analysis with accurate quantitative analysis, it can create a permanent infection prevention plan. In this study, we propose a new statistical mathematical model of infectious diseases that combines the SIR model in infection situation analysis with the multivariate analysis method that enables appropriate prediction and estimation. Multivariate analysis is a general term for methods that statistically analyze data consisting of multiple variables [13]. In particular, MT systems (hereinafter referred to as MTS) that combine quality engineering and theory based on statistical mathematics are widely applied in the fields of science and technology such as pattern recognition, prediction/estimation, and inspection. It is characterized by being able to evaluate important factors related to prediction and estimation and process multidimensional information.

MTS is a mathematical and method system for pattern recognition and prediction in quality engineering (Taguchi method). In addition to the MT method based on statistical mathematics, it includes a unique feature extraction technology, so it is a system that emphasizes practical use rather than simply multivariate data analysis theory. **Figure 2** shows the method components of MTS [14] [15].

(a) Mathematical pattern recognition: MT method, RT method.

Both MT/RT methods are methods for pattern recognition. As a common point, the normal group is used as a reference (unit space), and the difference in pattern from that is calculated as the distance. A large distance indicates that it is far from the reference pattern. The difference is that the MT method has the highest pattern recognition accuracy and may be regarded as one of AI (Artificial Intelligence). In the RT method, the scale of unit data is 2×2 , regardless of

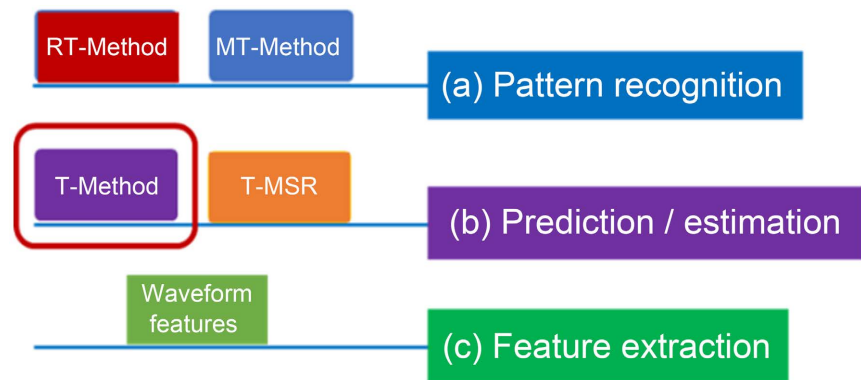


Figure 2. Method components of MTS.

the number of variables. Therefore, it is effective when the number of patterns to be recognized such as character recognition is large [16].

(b) Mathematical prediction/estimation: T method.

Similar to multiple regression analysis, the T method is a means for predicting and estimating output values (objective variables) from multivariate data (explanatory variables). It has the advantage that there is no instability or impossibility of calculation when the number of samples is small and the multi-collinearity problem, which is a weak point of multiple regression analysis.

(c) Feature value extraction.

In pattern recognition for waveforms and images, the success or failure of feature extraction technique determines. MTS includes feature extraction technique called “variation value” and “abundance value” from these patterns. Although it is a simple method, it often has better anomaly detection sensitivity than frequency analysis when targeting waveforms. It is also used as an extension for image inspection, etc., and by using it in combination with the MT method; it is possible to obtain faster and more sensitive results than conventional technology [17].

In this study, we adopt the “T method” which is a method for estimating output values from multivariate data in MTS. The T method is a mathematical method that estimates the output value in the same way as multiple regression analysis. Comprehensive estimation is performed from the relationship between each item and the output value by the method of estimating the output value from multiple item variables [18] [19].

3.1. Computation Formula for the T Method

The T Method defines the Unit Space where the output value is in the medium position and homogeneous (densely populated). The computation procedure of the T Method is explained below [20] [21].

3.2. Definition of the Unit Space and Computation of the Average of Relevant Items and Outputs

Let’s suppose that, as shown in **Table 1**, n number of data have been obtained

Table 1. Data for the unit space and average values of the items and outputs. All the items of the data must be in same dimension as image density or must be no dimension data.

Data No.	Item/variable				Output value
	1	2	...	k	
1	x_{11}	x_{12}	...	x_{1k}	y_1
2	x_{21}	x_{22}	...	x_{2k}	y_2
...
n	x_{n1}	x_{n2}	...	x_{nk}	y_n
Average	$\overline{x_1}$	$\overline{x_2}$...	$\overline{x_k}$	$\overline{y} = M_0$

for the Unit Space. All the items of the data must be in same dimension as image density or must be no dimension data. From the n number of samples in the Unit Space, we find average values $\overline{x_1}, \overline{x_2}, \dots, \overline{x_k}$ and average output value $\overline{y} = M_0$ for all items. Accordingly, the average values work out as follows:

$$\overline{x_j} = \frac{1}{n} (x_{1j} + x_{2j} + \dots + x_{nj}) \quad (j = 1, 2, \dots, k) \quad (5)$$

$$\overline{y} = M_0 = \frac{1}{n} (y_1 + y_2 + \dots + y_n) \quad (6)$$

Table 1 also shows these average values. One of the average values obtained from the n members of the Unit Space is the center of the Unit Space.

3.3. Definition of Signal Data

All data items marked l , left unselected for the Unit Space are treated as Signal data. Signal data is shown in **Table 2**. “Signal data” refers to all data used for finding the proportional coefficient β and SN ratio η .

3.4. Normalization of Signal Data

Signal Data is normalized using the average values of items and the output values of samples in the Unit Space. Normalization is performed by subtracting the average value $\overline{x_j}$ of item j in the Unit Space from value x'_{ij} of item j of the i -th Signal Data.

$$X_{ij} = x'_{ij} - \overline{x_j} \quad (i = 1, 2, \dots, l; j = 1, 2, \dots, k) \quad (4)$$

Likewise, normalization is performed by subtracting average value M_0 of the output from the Unit Space from output value y'_i of the i -th Signal Data.

$$M_i = y'_i - M_0 \quad (i = 1, 2, \dots, l) \quad (8)$$

Normalized Signal data is shown in **Table 3**.

3.5. Computation of Proportional Coefficient β and SN Ratio η (in Duplicate Ratio) for All Items

We will next compute proportional coefficient β and SN ratio η for all items. How the computation is performed is explained with item 1 as an example:

Table 2. Signal data. “Signal data” refers to all data used for finding the proportional coefficient β and SN ratio η .

Item/variable					
Data No.	1	2	...	k	Output value
1	x'_{11}	x'_{12}	...	x'_{1k}	y'_1
2	x'_{21}	x'_{22}	...	x'_{2k}	y'_2
...
l	x'_{l1}	x'_{l2}	...	x'_{lk}	y'_l

Table 3. Normalized signal data. Signal data is normalized using the average values of items and the output values of samples in the unit space.

Item/variable					
Data No.	1	2	...	k	Output value
1	X_{11}	X_{12}	...	X_{1k}	M_1
2	X_{21}	X_{22}	...	X_{2k}	M_2
...
l	X_{l1}	X_{l2}	...	X_{lk}	M_l

Proportional coefficient:

$$\beta_1 = \frac{M_1 X_{11} + M_2 X_{21} + \cdots + M_l X_{l1}}{r} \quad (9)$$

SN ratio:

$$\eta_1 = \begin{cases} \frac{1}{r} (S_{\beta 1} - V_{e1}) & (\text{when } S_{\beta 1} > V_{e1}) \\ 0 & (\text{when } S_{\beta 1} \leq V_{e1}) \end{cases} \quad (10)$$

where:

Effective divider:

$$r = M_1^2 + M_2^2 + \cdots + M_l^2 \quad (11)$$

Total variation:

$$S_{T1} = X_{11}^2 + X_{21}^2 + \cdots + X_{l1}^2 \quad (f = l) \quad (12)$$

f Degree of freedom

Variation of Proportional term:

$$S_{\beta 1} = \frac{(M_1 X_{11} + M_2 X_{21} + \cdots + M_l X_{l1})^2}{r} \quad (f = 1) \quad (13)$$

Error variation:

$$S_{e1} = S_{T1} - S_{\beta 1} \quad (f = l - 1) \quad (14)$$

Error variance:

$$V_{e1} = \frac{S_{e1}}{l-1} \quad (15)$$

From item 2 up to item k , we will likewise find proportional coefficient β and SN ratio η . This operation yields the results that are shown in **Table 4**.

3.6. Computation, Signal by Signal, of Integrated Estimate Value \hat{M} of Output

An item-by-item estimated value is found for each piece of Signal Data using the proportional coefficient β and SN ratio η , item by item. The estimated value of the output of item 1 for the i -th Signal Data is:

$$\hat{M}_{i1} = \frac{X_{i1}}{\beta_1} \quad (16)$$

An estimation is likewise made of item 2 through item l for the i -th Signal Data. And finally an integration of the result is performed by weighting it with $\eta_1, \eta_2, \dots, \eta_k$ which is the estimated measure of precision of each item. Thus, the integrated estimate value \hat{M}_i of the output of the i -th Signal Data becomes:

$$\hat{M}_i = \frac{\eta_1 \times \frac{X_{i1}}{\beta_1} + \eta_2 \times \frac{X_{i2}}{\beta_2} + \dots + \eta_k \times \frac{X_{ik}}{\beta_k}}{\eta_1 + \eta_2 + \dots + \eta_k} \quad (i = 1, 2, \dots, l) \quad (17)$$

Table 5 shows the real values (measured values) of the Signal Data M_1, M_2, \dots, M_l and the integrated estimate values $\hat{M}_1, \hat{M}_2, \dots, \hat{M}_l$.

3.7. Computation of Integrated Estimate SN Ratio

The Integrated Estimate SN Ratio is computed using the following equation based on **Table 5**. The result of the computation will be used in Section 3.8 “Evaluation of the Relative Importance of an Item” comes up as the next step.

Table 4. Proportional coefficient β and SN ratio η , item by item.

	Item/variable			
β, η	1	2	...	k
Proportional β	β_1	β_2	...	β_k
SN ratio η	η_1	η_2	...	η_k

Table 5. Measured values and integrated estimate values of signal data.

Data No.	Measured value	Integrated estimate value
1	M_1	\hat{M}_1
2	M_2	\hat{M}_2
:	:	:
:	:	:
l	M_l	\hat{M}_l

Integrated Estimate SN Ratio:

$$\eta = 10 \log_{10} \left[\frac{\frac{1}{r}(S_\beta - V_e)}{V_e} \right] (\text{db}) \quad (18)$$

where:

Linear equation:

$$L = M_1 \hat{M}_1 + M_2 \hat{M}_2 + \dots + M_l \hat{M}_l \quad (19)$$

Effective divider:

$$r = M_1^2 + M_2^2 + \dots + M_l^2 \quad (20)$$

Total variation:

$$S_T = \hat{M}_1^2 + \hat{M}_2^2 + \dots + \hat{M}_l^2 \quad (f = l) \quad (21)$$

Variation of proportional term:

$$S_\beta = \frac{L^2}{r} \quad (f = 1) \quad (22)$$

Error variation:

$$S_e = S_T - S_\beta \quad (f = l - 1) \quad (23)$$

Error variance:

$$V_{el} = \frac{S_e}{l - 1} \quad (24)$$

3.8. Evaluation of the Relative Importance of an Item

The relative importance of an item is evaluated in terms of the extent to which the Integrated Estimate SN Ratio deteriorates when the item is not used. For the evaluation, a two-level orthogonal array ($4 \times$ prime version of the two-level series is advisable) is used. Use of an orthogonal array allows a comparison to be made of the SN ratio η of the integrated estimate under various conditions [22]. Let's suppose we have 11 items before us, X_1, X_2, \dots, X_{11} . We assign the 11 items to Columns 1 to 11 in **Table 6**. Where the two levels of the orthogonal array mean the following:

Level 1: Item will be used.

Level 2: Item will not be used.

In Test 1, with all Columns, Column 1 through Column 11, being on Level 1, it is shown that items X_1, X_2, \dots, X_{11} will all be used, and that the SN ratio (db) of the integrated estimate works out to η_1 . Note in passing that the SN ratio η (db) of the integrated estimate computed is used for η_1 in **Table 6**. In Test 2, with Columns 1 through 5 being on Level 1, it is shown that five items, X_1, X_2, \dots, X_5 will be used and that the SN ratio (db) of the integrated estimate come out to be η_2 . In the same manner as in Test 12, it is shown that, with Columns 3, 4, 6, 8, and 11 being on Level 1, five items, $X_3, X_4, X_6, X_8, X_{11}$, are

Table 6. Orthogonal array L12 and assignment of items.

No.	Item No.											Integrated estimate SN Ratio η (db)
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	
1	1	1	1	1	1	1	1	1	1	1	1	η_1
2	1	1	1	1	1	2	2	2	2	2	2	η_2
3	1	1	2	2	2	1	1	1	2	2	2	η_3
4	1	2	1	2	2	1	1	1	2	2	2	η_4
5	1	2	2	1	2	2	1	2	1	2	1	η_5
6	1	2	2	2	1	2	2	1	2	1	1	η_6
7	2	1	2	2	1	1	2	2	1	2	1	η_7
8	2	1	2	1	2	2	2	1	1	1	2	η_8
9	2	1	1	2	2	2	1	2	2	1	1	η_9
10	2	2	2	1	1	1	1	2	2	1	2	η_{10}
11	2	2	1	2	1	2	1	1	1	2	2	η_{11}
12	2	2	1	1	2	1	2	1	2	2	1	η_{12}

used, and the SN ratio (db) of the integrated estimate work out to be η_{12} . With regard to the SN ratio of the integrated estimate, we find the difference between the averages of the SN ratio for Level 1 (with the items to be used) and that for Level 2 (with the items not to be used), item by item, and on that basis determine the relative importance of the items.

4. Statistical Analysis of COVID-19 Analysis of Infection Status

Figure 3 shows the transition in the number of new coronavirus infections of Tokyo and Hokkaido in Japan from June 2020 to the end of June 2021. In Tokyo, the waves spread of infection in this period has occurred three times. Especially in the third wave during the winter season (November 2020 to the end of February 2021), the new COVID-19 cases per day exceeded 2000 and the infection spread rapidly. An important index for analyzing the infection status is ERN (R_s) in the sequential SIR model. In this study, we apply the T method to the sequential SIR model in the infection situation analysis and verify the estimation accuracy of ERN and the new COVID-19 cases during the period when the infection is spreading.

In addition, to clarify the important factor and its contribution according to the prediction and estimation of the infection situation, in the fourth wave of the period from April to June 2021, domestic infections due to the influx of variant type virus from overseas are rapidly expanding [23] [24].

4.1. ERN Change Point Analysis by Variant Type Virus

In the fourth wave which is the period from April to June 2021, domestic infections

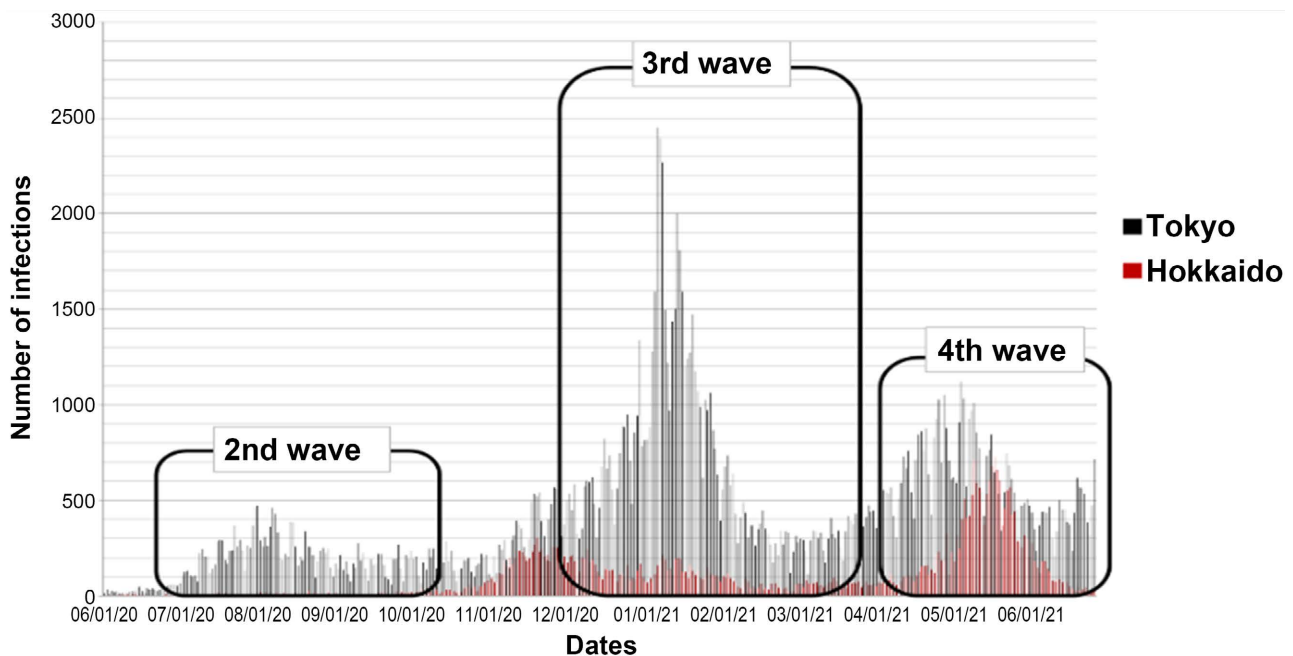


Figure 3. June 2020-June 2021 Tokyo/Hokkaido Transition of the new COVID-19 cases. In Tokyo, the waves spread of infection in this period has occurred three times.

due to the influx of variant type virus from overseas are rapidly expanding. In particular, infection with variant type virus during this period in Hokkaido is spreading rapidly. In this section, we analyze the correlation between important items that affect the spread of infection and ERN (R_T) estimated by the T method, and the change point in the infection status.

Table 7 shows the signal data and items of the 4th wave of Hokkaido from April to June 2021 and the basic data of ERN (R_S in SIR). In estimating the objective variable ERN (R_T in T-Method), the explanatory variables are the number of infected persons (I), recovers (R), PCR test (PCR), needs treatment (Nt), positive rate (P), severely ill (Se), deceased (De) which are characteristic data of infectious diseases and 13 items including 6 items of infection day (W), wind velocity (V), temperature (T), precipitation (Pr), humidity (H), atmospheric pressure (AP) which are environment/weather related data. The basic data is quoted from “Toyo Keizai Online “Coronavirus Disease (COVID-19) Situation Report in Japan” and Japan Meteorological Agency online data.

4.2. ERN (R_S/R_T) Correlation Verification

Table 8 shows the ERN estimation results (measured value R_S /estimated value R_T) by the T method. As shown in the correlation distribution in **Figure 4**, some correlation can be confirmed in the estimation result R_S/R_T . **Table 9** shows the calculation process of the integrated estimate SN ratio η (db).

Table 10 and **Figure 5**, **Figure 6** shows the percentage of contribution and the factor-effect diagram of the items that are valid and that are not valid for the integrated estimate of ERN calculated from the SN ratio difference. In the integrated

Table 7. Data for Hokkaido April-May 2021 items and output values (ERN: R_s in SIR). It shows the signal data and items of the 4th wave of Hokkaido from April to June 2021 and the basic data of ERN (R_s in SIR).

Symbol	Infectious properties items								Environmental items				Output	
	W	I	R	PCR	Nt	P	Se	De	V	T	Pr	H	AP	R_s
Data No.	1	2	3	4	5	6	7	8	9	10	11	12	13	(in SIR)
04/01/21	5	57	96	2951	0	1.9	1	3	3.7	7.4	0	64	6.4	1.02
04/02/21	6	57	64	2258	0	2.5	3	2	3.4	8.4	0	49	5.1	1.01
04/03/21	7	66	33	2435	33	2.7	1	0	3.6	9.1	0.5	75	8.7	1.00
04/04/21	1	53	31	1636	22	3.2	0	0	3.6	6.3	5	72	6.9	0.94
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
04/29/21	5	234	130	5867	102	4.0	1	2	2.4	12.8	0	61	8.8	1.33
04/30/21	6	187	67	2763	116	6.8	0	4	5.9	7.7	33.5	86	9.1	1.28
05/01/21	7	180	158	2493	21	7.2	1	1	6.2	7.2	2	77	7.8	1.24
05/02/21	1	326	69	4408	251	7.4	0	6	6.2	6.1	4.5	77	7.2	1.39
05/03/21	2	114	132	2751	0	4.1	-1	4	6	5.6	9	82	7.5	1.31
05/04/21	3	233	128	2388	100	9.8	0	5	3.9	8.5	0	71	7.8	1.37
05/05/21	4	181	67	2308	112	7.8	1	2	8.2	11.9	0	70	9.8	1.25
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	:	:	:	:	:	:	:	:	:
05/30/21	1	288	298	4470	-25	6.4	0	15	2.8	13	0	79	11.8	0.81
05/31/21	2	279	423	4243	-163	6.6	-1	19	6.2	10.4	0	71	8.9	0.79

Table 8. Measured values (R_s) and integrated estimate values (R_T) of ERN.

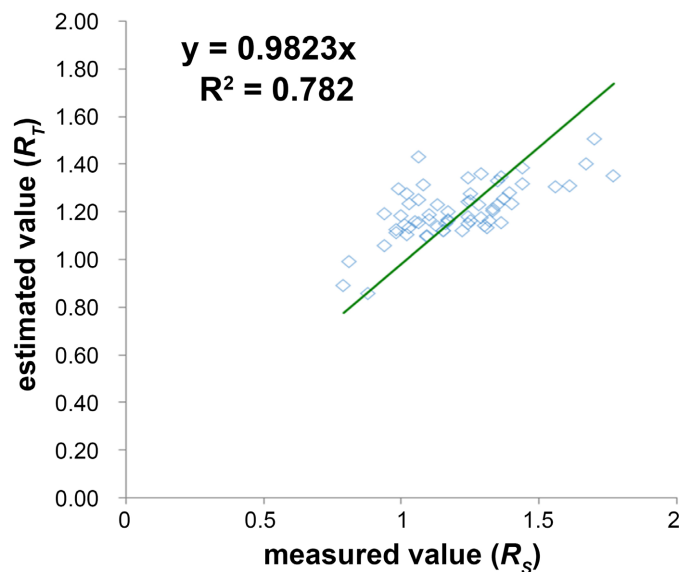
Data No.	Measured value (SIR) R_s	Integrated estimate value (T Method) R_T
04/01/21	1.02	1.01
04/02/21	1.01	1.04
04/03/21	1.00	1.08
:	:	:
:	:	:
04/30/21	1.28	1.23
05/01/21	1.24	1.24
:	:	:
:	:	:
05/30/21	0.81	1.13
05/31/21	0.79	0.89

Table 9. Calculation process of the integrated estimate SN ratio η (db) in estimation of ERN.

Item No.	1	2	3	4	5	6	7	8	9	10	11	12	13
Item Name	W	I	R	PCR	Nt	P	Se	De	V	T	Pr	H	AP
Variation of proportional term S_β	0.7	2.0E+05	9.0E+03	5.0E+06	4.0E+05	98.0	2.0	11.0	1.0	23.0	25.0	87.0	0.0
Error variation S_e	200.0	3.0E+06	1.0E+06	2.0E+08	1.0E+06	475.0	343.0	860.0	165.0	949.0	3925.0	18,946.0	872.0
Error variance V_e	4.0	6.0E+04	3.0E+04	4.0E+06	3.0E+04	9.0	6.5	16.2	3.1	17.9	74.1	357.4	16.5
Proportional coefficient β	-1.0	326.0	-62.0	1464.0	407.0	6.6	0.8	-2.2	0.6	3.2	-3.4	-6.2	0.0
SN ratio η	-0.4	1.4	-0.3	0.1	5.5	4.4	-0.3	-0.1	-0.4	0.1	-0.3	-0.3	-0.5

Table 10. Diagnosis results to the 2021 April over May of ERN. This means that the characteristic items of infectious diseases and environmental and meteorological items are important factors to be considered in estimating ERN.

Item No.	1	2	3	4	5	6	7	8	9	10	11	12	13
Item symbol	W	I	R	PCR	Nt	P	Se	De	V	T	Pr	H	AP
Level 1	5.57	5.53	4.82	4.57	7.72	6.56	5.54	7.32	5.01	6.31	4.71	6.42	4.87
Level 2	5.46	5.51	6.21	6.46	3.31	2.94	5.49	5.22	6.02	4.72	6.32	4.62	6.16
Contribution (db)	0.11	0.02	−1.39	−1.89	4.41	3.62	0.05	2.10	−1.01	1.59	−1.61	1.80	−1.29
Integrated estimate SN ratio	13.7 db	Infectious disease feature items			10.2 db	Contribution of positive rate (P) and needs treatment (Nt)			8.03 db	Contribution of environmental items			3.5 db
		Ratio of infectious disease items to the integrated estimate SN ratio			75%	Ratio of P and Nt items to the integrated estimate SN ratio			59%	Ratio of environmental items to the integrated estimate SN ratio			25%

**Figure 4.** Correlation distribution of measured values (R_s) and integrated estimate values (R_r) of ERN. Some correlation can be confirmed in the estimation result R_r/R_s .

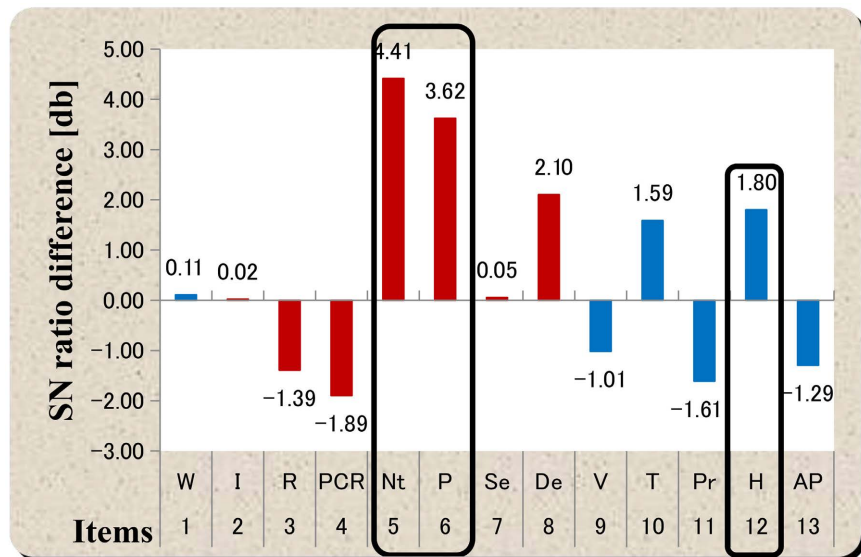


Figure 5. Contribution to estimated value for each item to the 2021 April over May. Items with the contribution contribute to raising the estimated value of ERN.

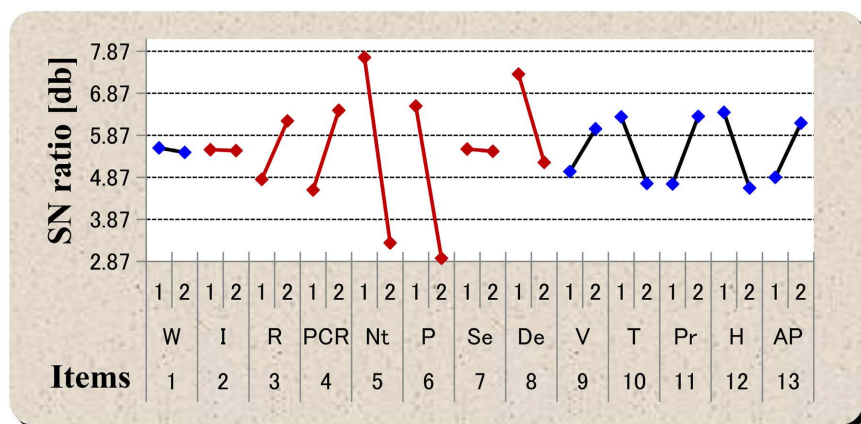


Figure 6. Factor effect diagram.

estimate SN ratio (13.7 db) of ERN (R_T), the contribution of environmental and meteorological items is 3.5 db, and the percentage to the integrated estimate SN ratio is 25%. On the other hand, the contribution of the characteristic items of infectious diseases is 10.2 db, and the percentage to the integrated estimate SN ratio is as high as 75%. This means that the characteristic items of infectious diseases and environmental and meteorological items are important factors to be considered in estimating ERN.

In particular, the total contribution of item No. 5 need treatment (Nt) and No. 6 positive rate (P) are 8.03 db, accounting for 59% of the integrated estimate SN ratio. From this, it can be inferred that the No. 5 need treatment (Nt) and No. 6 positive rate (P) are items that have a strong influence on the spread of infection in the estimation of ERN (R_T).

Figure 7 is a transition graph comparing and verifying the R_S/R_T estimates for Hokkaido (April-May 2021). During this period, both R_S and R_T can be seen as

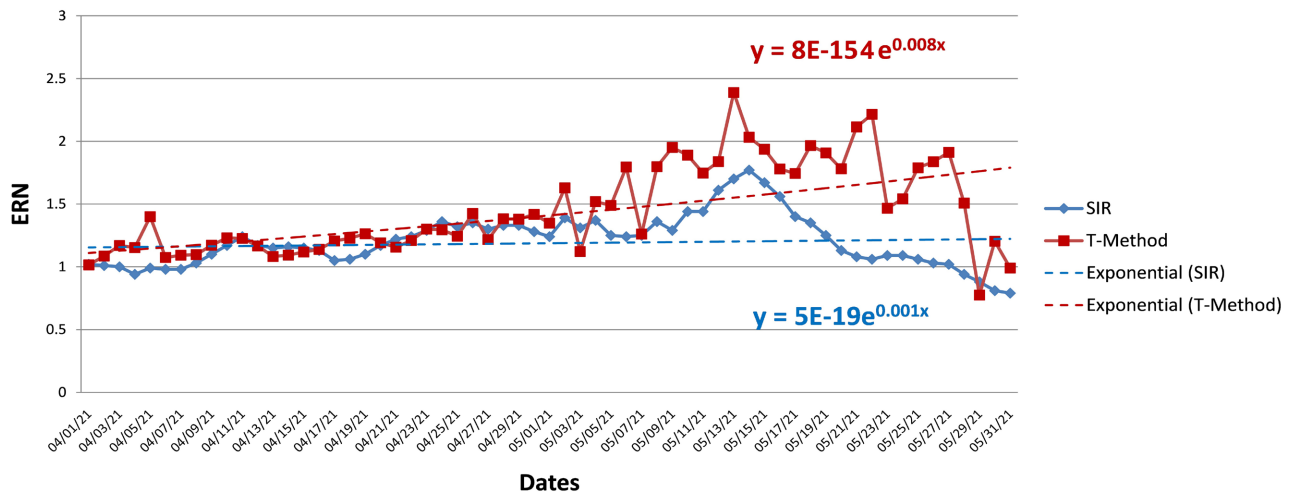


Figure 7. Hokkaido ERN (April-May 2021) R_s/R_T Comparison Transition. During this period, both R_s and R_T can be seen as spreading.

spreading. In particular, since late April, the estimated R_T for R_s has increased by 20%, indicating that the infection has turned to a rapid spread. In Japan, it is a time when infections with variant type virus, which have strong infectivity and are likely to become severe, have spread in place of conventional virus. This means that the estimated value of ERN (R_T) enables more accurate diagnosis in the analysis of infection status.

Figures 8-10 shows the correlation between the change in need treatment (Nt), positive rate (P), humidity (H) and ERN (R_T) during the period from April to the end of May in Hokkaido. The number of need treatment due to the onset in **Figure 8** has increased sharply since the beginning of May, and ERN (R_T) has also been on the increase at the same time, confirming a strong correlation. In addition, the positive rate (%), which indicates the ratio of the number of infected persons to the PCR test in **Figure 9**, and ERN (R_T) tended to increase rapidly at the same time, confirming a strong correlation. Furthermore, the humidity (%) and ERN (R_T) in **Figure 10** are correlated with the 5-days moving average (incubation period). From the results of this analysis, it can be judged that the variant type virus, which has strong infectivity and is likely to become severe, is the main cause of the spread of infection during this period.

4.3. Verification of Estimation Accuracy of the New COVID-19 Cases in R_s/R_T

Figure 11(a), **Figure 11(b)** show the results of comparing the correlation between the estimated/actual number of the new COVID-19 cases in Hokkaido from April to May in ERN (R_s and R_T). These are the analysis results in which the items (12 items) as when estimating ERN by the T method are applied and the objective variable that is the output is the number of new COVID-19 cases. However, the positive rate is the ratio of the number of infected persons to the number of PCR tests, and is excluded from the estimated items of the number of infected persons. From this analysis results, in estimating the new COVID-19

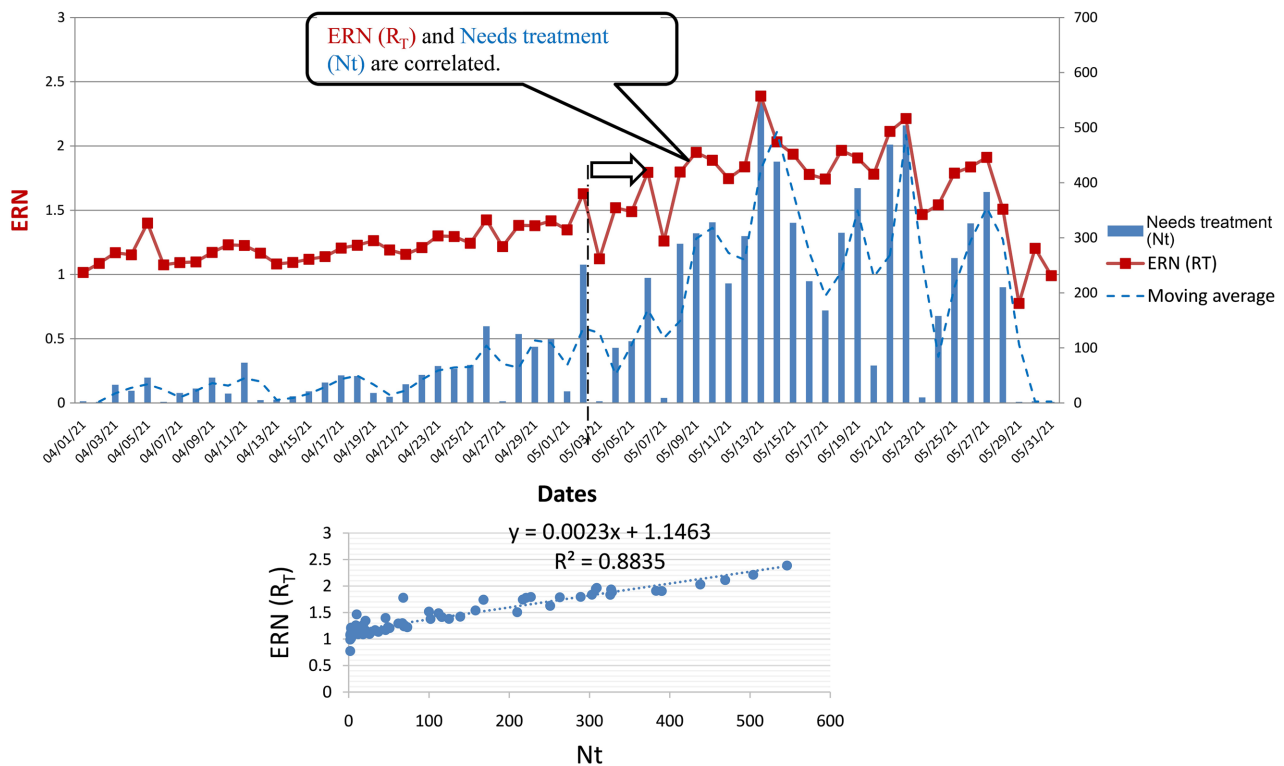


Figure 8. ERN (R_T) and (Nt) correlation diagram from April to May 2021 in Hokkaido. ERN increases in proportion to needs treatment. It is necessary to improve the vaccination rate and strengthen the medical system.

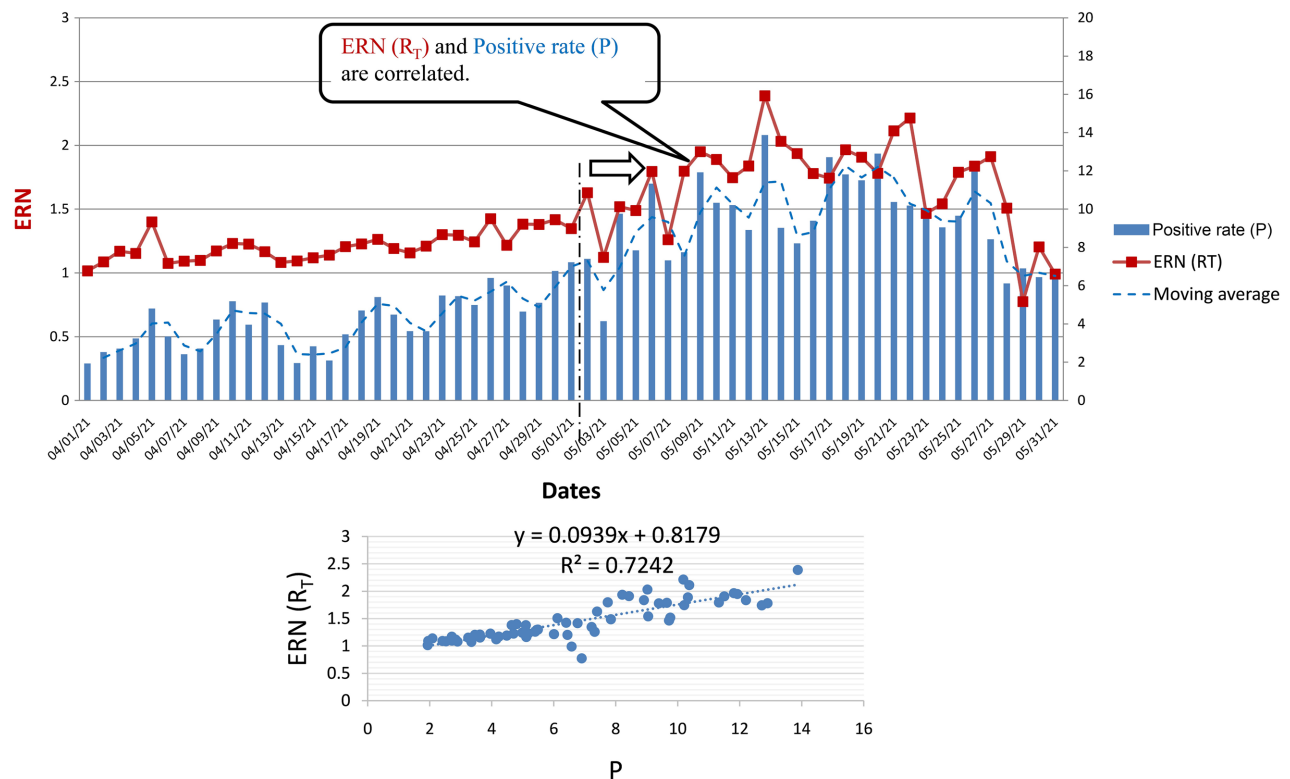


Figure 9. ERN (R_T) and (P) correlation diagram from April to May 2021 in Hokkaido. ERN increases in proportion to the positive rate. Measures to increase the number of PCR tests are required.

cases in ERN (R_s and R_T), the coefficient of determination (R^2) which is an index showing the explanatory power of the estimated value with respect to the measured value of the objective variable (the new COVID-19 cases), R_T is 1.1 times more significant than R_s [$R^2 = 0.8698(R_s) < 0.9670(R_T) = 1.1$].

The results of regression analysis for each item on the estimated number of new COVID-19 cases are shown in Table 11 and Figure 12, Table 12 and Figure 13.

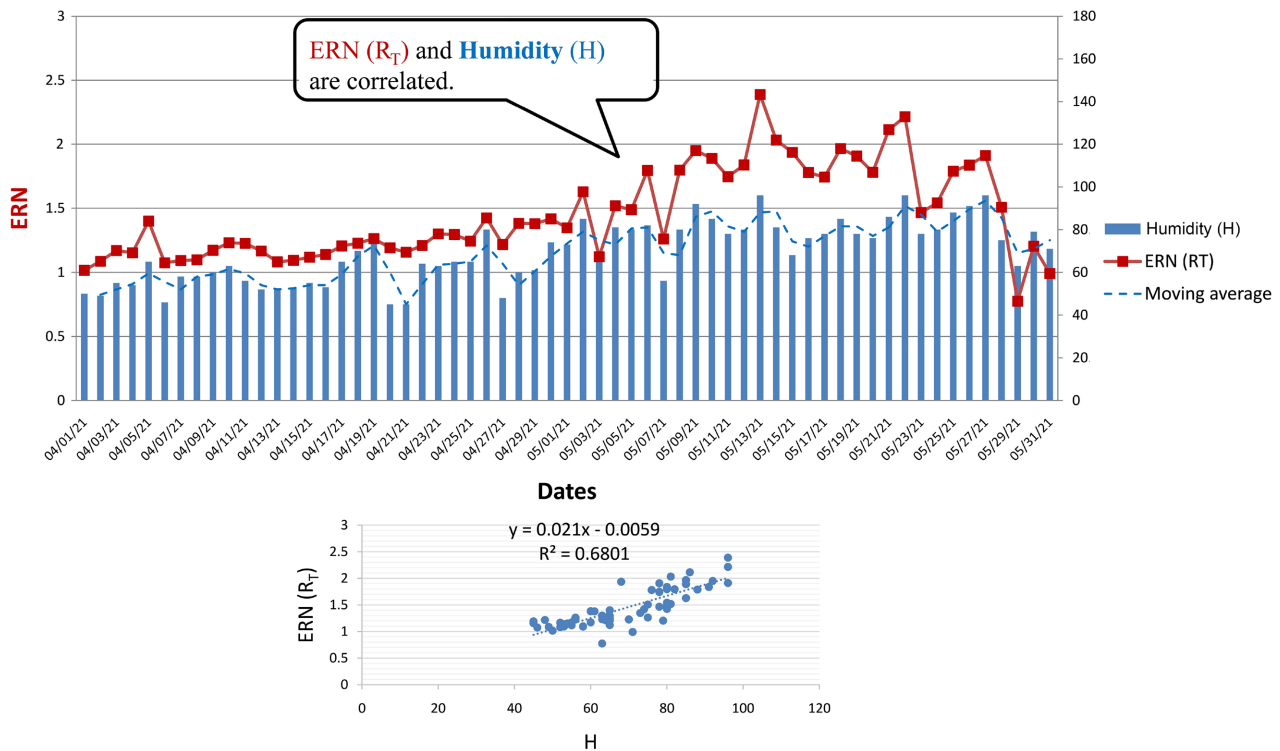
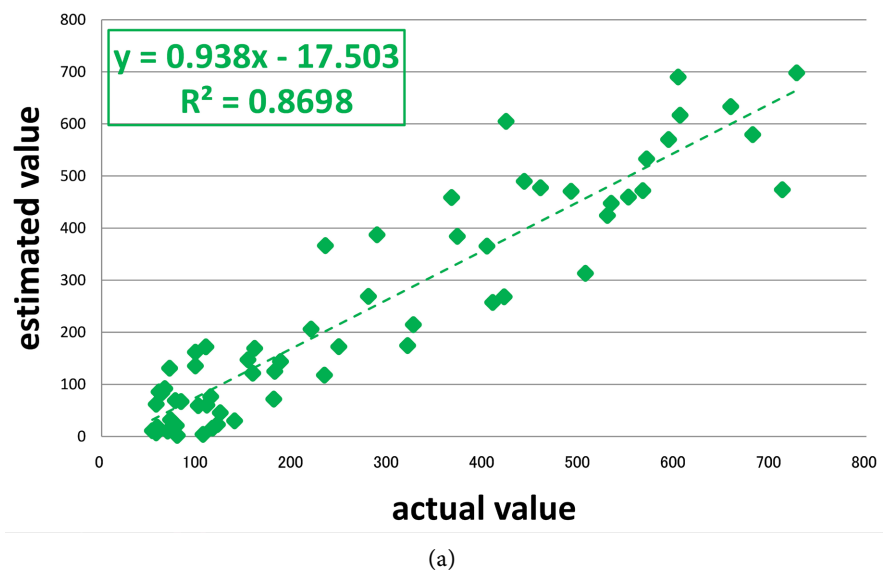


Figure 10. ERN (R_T) and (H) correlation diagram from April to May 2021 in Hokkaido. ERN increases in proportion to humidity. Humid environment affects the spread of infection. It is necessary to take measures while paying attention to fluctuations in humidity rise.



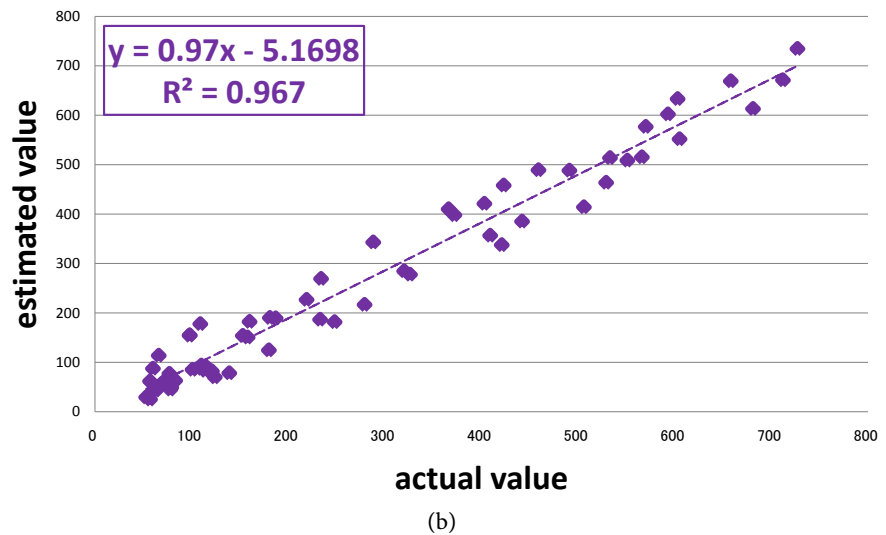


Figure 11. Correlation of estimated/actual the new COVID-19 cases in Hokkaido from April to May in ERN (R_S and R_T). From this analysis results, in estimating the new COVID-19 cases in ERN (R_S and R_T), the coefficient of determination (R^2) which is an index showing the explanatory power of the estimated value with respect to the measured value of the objective variable (the new COVID-19 cases), R_T is 1.1 times more significant than R_S [$R^2 = 0.8698$ (R_S) < 0.9670 (R_T) = 1.1]. (a) In the ERN (R_S ; SIR); (b) In the ERN (R_T ; T-Method).

Table 11. Regression coefficient for the estimated number of the new COVID-19 cases for each item in the R_S April-May.

Item No.	1	2	3	4	5	6	7	8	9	10	11	12	I
Item symbol	W	R_S	R	PCR	Nt	Se	De	V	T	Pr	H	AP	
Average	4.0	1.2	157	3831	119	0.6	6.0	4.3	10.6	3.1	65.4	8.5	278
Standard deviation	2.0	0.2	157	1798	177	2.4	4.2	1.5	4.0	7.7	15.1	3.2	210
Regression coefficient	9.8	327.5	0.7	0.1	0.8	16.7	27.3	-41.3	41.1	-2.0	4.2	50.4	0

Regression statistics		Analysis of variance		Freedom	Variation	Variance
Multiple correlation R		0.32	Regression	1	275,828	275,828
Multiple decision R^2		0.10	Residual	59	2,360,800	40,014
Correction R^2		0.09		Coefficient	Standard error	P-value
Standard error		200.0	Intercept	2.3E-13	25.6	0.56
Number of observations		61	X value 1	327.5	124.7	0.01

Table 12. Regression coefficient for the estimated number of the new COVID-19 cases for each item in the R_T April-May.

Item No.	1	2	3	4	5	6	7	8	9	10	11	12	I
Item symbol	W	R_T	R	PCR	Nt	Se	De	V	T	Pr	H	AP	
Average	4.0	1.5	157	3831	119	0.6	6.0	4.3	10.6	3.1	65.4	8.5	278
Standard deviation	2.0	0.4	157	1798	177	2.4	4.2	1.5	4.0	7.7	15.1	3.2	210
Regression coefficient	9.8	498.7	0.7	0.1	0.8	16.7	27.3	-41.3	41.1	-2.0	4.2	50.4	0

Regression statistics	Analysis of variance		Freedom	Variation	Variance
Multiple correlation R	0.85	Regression	1	1,916,109	1,916,109
Multiple decision R^2	0.73	Residual	59	720,519	12,212
Correction R^2	0.72		Coefficient	Standard error	P-value
Standard error	110.5	Intercept	$-5.9\text{E}-14$	14.2	0.86
Number of observations	61	X value 1	498.7	39.8	$2.9\text{E}-18$

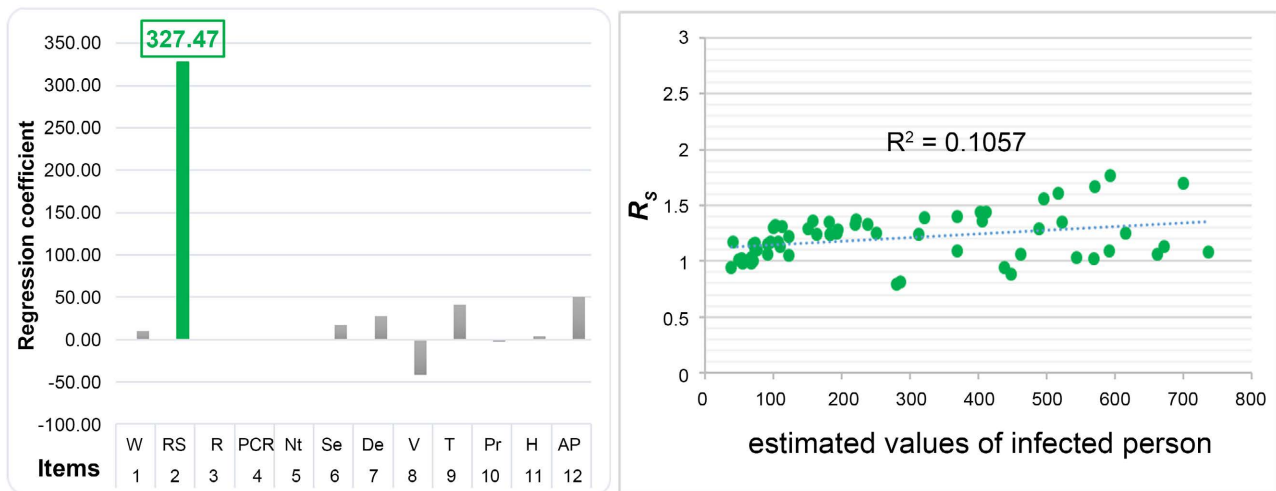


Figure 12. The new COVID-19 cases in the R_S contribution.

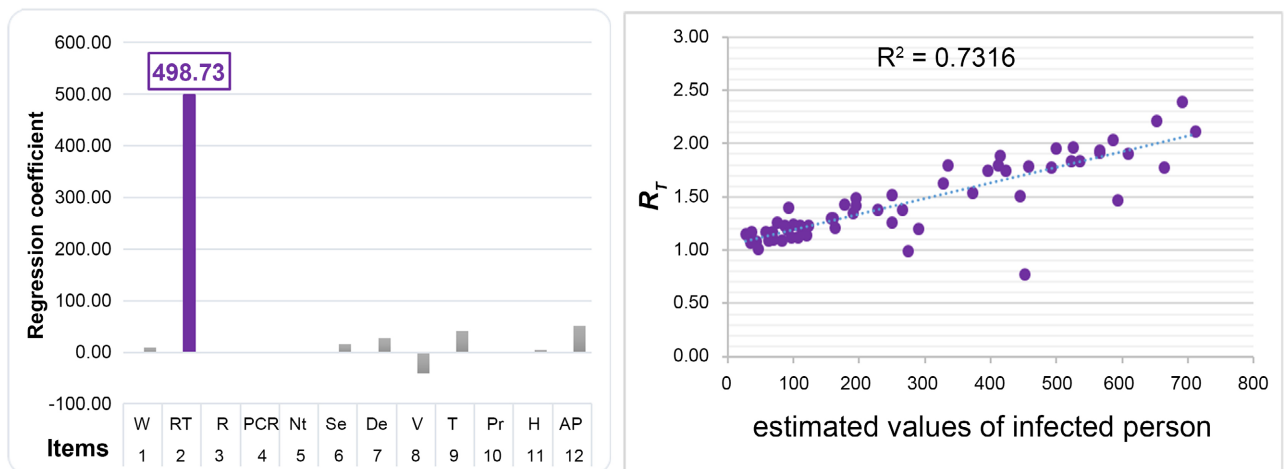


Figure 13. The new COVID-19 cases in the R_T contribution. As a result of comparing and verifying the contribution of R_S and R_T to the estimated number of new COVID-19 cases by the regression coefficient, it can be inferred that R_T has 1.5 times higher estimation accuracy [$R_S = 327.5$, $RT = 498.7 \rightarrow R_T/R_S = 1.5$].

Regression is the application of the model $Y = f(X)$ to data when Y is a continuous value in statistics. In other words, fit the model between the dependent variable (objective variable) Y and the independent variable (explanatory variable) X on the continuous scale. If X is one-dimensional, it is called simple regression, and if X is two-dimensional or more, it is called multiple regression.

Regression analysis is to analyze by regression. The most basic model used in regression is a linear regression of the form $Y = AX + B$. In regression analysis, the formula expressing the relationship between the independent variable and the dependent variable is estimated by a statistical method.

As a result of comparing and verifying the contribution of R_S and R_T to the estimated number of new COVID-19 cases by the regression coefficient (A), it can be inferred that R_T has 1.5 times higher estimation accuracy [$R_S = 327.47$, $R_T = 498.73 \rightarrow R_T/R_S = 1.5$].

This indicates that R_T has a better linearity and a stronger correlation than R_S in estimating the new COVID-19 cases in ERN (R_S and R_T). It also means that the application of the ERN (R_T) estimate by the T method enables a more accurate estimate of the new COVID-19 cases.

5. Conclusions

ERN is considered to be the criterion for the spread of COVID-19 worldwide. However, it is an estimated value that is easily affected by the number of PCR tests, the basic reproduction number, and changes in infectivity due to variant type virus. In the analysis of the infection status by the sequential SIR model, it is possible to analyze and evaluate as a relative value to some extent, but it is difficult to analyze and evaluate as an absolute value.

It is an important index for infection control whether it is possible to identify infected persons (affected persons) at a time when the amount of virus is extremely high and infectivity is high (early onset) and to isolate (including voluntary isolation) them from the primary infection at an early stage. In the sequential SIR model, it is possible to simulate the infection process of infectious diseases by using the population distribution and the quantified characteristics of infectious diseases as input data.

However, in the process of generating ERN estimation data, it depends largely on virus infectivity, and changes the various coefficients related to infectious diseases in multiple dimensions by the changes in the environment (human intervention such as behavioral restriction, differences in the surrounding environment such as climate change, and an increase in immune carriers due to the development of epidemics, etc.). Therefore, there is a problem that it is difficult to set the uncertainty (accuracy) region [25] [26].

In this study, we adopted the multivariate analysis method (T method) for the purpose of further improving the estimation accuracy of ERN in the current sequential SIR model. The T method can evaluate and predict important factors related to estimation and analyze phenomena that change in real time (time series analysis). As a research result, it was clarified that it is possible to estimate with higher accuracy by applying the T method to the estimated value of ERN by the current sequential SIR model. The application of the multivariate analysis method (T method) can be said to be an effective method that plays an essential role in infectious disease prevention measures. In addition, by applying the T

Method, it was clarified that the infectious variables and environmental climate change variables used during ERN estimation have an important effect on the estimation accuracy [27] [28].

As a future research subject, in estimating the number of new COVID-19 cases and ERN, we will examine how effective public health interventions (vaccination, quarantine, contact history survey, etc.) are in controlling infectious diseases. Furthermore, we will approach the construction of a more accurate infectious disease statistical mathematical model by newly introducing various coefficients related to the infectious disease evaluation items listed below [29] [30] [31] [32] [33] [34].

Coefficients related to infectious disease evaluation items:

- Regional coefficient: Virus contact rate in the ratio of regional people flow.
- Onset coefficient: Onset coefficient based on amount of virus contact and antibody.
- Medical collapse coefficient: Correlation coefficient between the new COVID-19 cases accepted and the number of infected peoples [35] [36] [37].

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Grassly, N.C. and Fraser, C. (2008) Mathematical Models of Infectious Disease Transmission. *Nature Reviews Microbiology*, **6**, 477-487.
<https://doi.org/10.1038/nrmicro1845>
- [2] Anderson, R.M. and May, R.M. (1991) Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, New York.
- [3] Imai, N., *et al.* (2020) Report 3: Transmissibility of 2019-nCoV.
<https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-transmissibility-25-01-2020.pdf>
- [4] Endo, A., *et al.* (2020) Estimating the Overdispersion in COVID-19 Transmission Using Outbreak Sizes Outside China. *Wellcome Open Research*, **5**, 67.
<https://doi.org/10.12688/wellcomeopenres.15842.3>
- [5] Hao, X.J., *et al.* (2020) Reconstruction of the Full Transmission Dynamics of COVID-19 in Wuhan. *Nature*, **584**, 420-424. <https://doi.org/10.1038/s41586-020-2554-8>
- [6] Nishiura, H., *et al.* (2020) Serial Interval of Novel Coronavirus (COVID-19) Infections. *International Journal of Infectious Diseases*, **93**, 284-286.
- [7] Lipsitch, M., *et al.* (2003) Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science*, **300**, 1966-1970.
<https://doi.org/10.1126/science.1086616>
- [8] Chandrow, O., *et al.* (2021) Numerical Analysis and Transformative Predictions of Fractional Order Epidemic Model during COVID-19 Pandemic: A Critical Study from Bangladesh. *Journal of Applied Mathematics and Physics*, **9**, 2258-2276.
<https://doi.org/10.4236/jamp.2021.99144>
- [9] Seydou, M. and Tessa, M. (2021) Approximations of Quasi-Stationary Distributions

- of the Stochastic SVIR Model for the Measles. *Journal of Applied Mathematics and Physics*, **9**, 2277-2289. <https://doi.org/10.4236/jamp.2021.99145>
- [10] Nishiura, H. and Inaba, H. (2006) Epidemics of Infectious Diseases: Quantitative Issues in Infectious Disease Mathematical Models. *Statistical Mathematics*, **54**, 461-480.
 - [11] Yan, P. and Chowell, G. (2019) Quantitative Methods for Investigating Infectious Disease Outbreaks. Springer, Cham. <https://doi.org/10.1007/978-3-030-21923-9>
 - [12] Inaba, H. (2020) Mathematical Modelling of Infectious Diseases. Baifukan, Tokyo, 50-265.
 - [13] Wakui, Y. and Wakui, S. (2014) Understanding Multivariate Analysis. Gijutsu-Hyoronsha, Tokyo, 190-200.
 - [14] Tatebayashi, K., *et al.* (2012) Introduction MT System. Union of Japanese Scientists and Engineers, Tokyo, 133-174.
 - [15] Toma, E. and Ito, Y. (2021) The Method for Optimum Design of Water Rocket Flight Stability Performance Conditions Using CAE with T Method and Robust Parameter Design. *Journal of Applied Mathematics and Physics*, **9**, 2669-2697. <https://www.scirp.org/journal/jamp>
<https://doi.org/10.4236/jamp.2021.911172>
 - [16] Suzuki, M. (2012) Introduction to MT System Analysis Method. Nikkan Kogyo Shimbun, Tokyo, 7-101.
 - [17] Taguchi, G. (2005) Purpose and Basic Functions (6)—Comprehensive Prediction by T Method. *Quality Engineering Society*, **13**, 5-10.
 - [18] Taguchi, G. (2006) Purpose and Basic Functions (11)—T Method for Recognition. *Quality Engineering Society*, **14**, 5-9.
 - [19] Inaba, J., *et al.* (2012) Prediction Accuracies of Improved Taguchi's T Methods Compared to Those of Multiple Regression Analysis. *Journal of the Japanese Society for Quality Control*, **42**, 265-277.
 - [20] Hosokawa, T., *et al.* (2015) A Proposal of Development Methodology Integrating Parameter design and T-Method. *Journal of the Japanese Society for Quality Control*, **45**, 194-202.
 - [21] Hirotsu, C., *et al.* (1997) Algorithm for p Value, Power and Sample Size Determination for Max T Method. *Society of Applied Statistics*, **26**, 1-16. <https://doi.org/10.5023/jappstat.26.1>
 - [22] Kawada, H. and Nagata, Y. (2015) Studies on the Item Selection in Taguchi's T-Method. *Journal of the Japanese Society for Quality Control*, **45**, 179-193.
 - [23] Maeda, M. (2017) New Regression Method Based on the Idea of T-Method (1). *Journal of the Japanese Society for Quality Control*, **47**, 185-194.
 - [24] Taguchi, G. (2012) Mahalanobis-Taguchi System in the 21st Century, MTA, TS, and T Method. *Journal of Quality Engineering Forum*, **20**, 261-268. <http://id.ndl.go.jp/bib/000000400206>
 - [25] Soga, M. (2008) A Comparison of Estimate Accuracy with T Method and Multiple Regression Analysis. *QEF Quality Engineering Research Presentation Conference Executive Committee Edition*, **16**, 430-433. <http://id.ndl.go.jp/bib/000009657728>
 - [26] Inao, A., *et al.* (2012) Taguchi's T Method and Its Improved Method and Performance Comparison of Multiple Regression Analysis. *Journal of the Japanese Society for Quality Control*, **42**, 265-277.
 - [27] Das, M., *et al.* (2021) Does Climate Variability Impact COVID-19 Outbreak? An

- Enhanced Semantics-Driven Theory-Guided Model. *SN Computer Science*, **2**, Article No. 452. <https://doi.org/10.1007/s42979-021-00845-9>
- [28] Bashir, M.F., *et al.* (2020) Correlation between Climate Indicators and COVID-19 Pandemic in New York, USA. *Science of the Total Environment*, **728**, Article ID: 138835.
- [29] Shaik, E., Hossain, Q.S. and Forhad Faisal Rony, G.M. (2021) Impact of COVID-19 on Public Transportation and Road Safety in Bangladesh. *SN Computer Science*, **2**, Article No. 453. <https://doi.org/10.1007/s42979-021-00849-5>
- [30] Dey, J., *et al.* (2021) Episode of COVID-19 Telepsychiatry Session Key Origination upon Swarm-Based Metaheuristic and Neural Perceptron Blend. *SN Computer Science*, **2**, Article No. 445. <https://doi.org/10.1007/s42979-021-00831-1>
- [31] Alexopoulos, A.R., *et al.* (2020) The Use of Digital Applications and COVID-19. *Community Mental Health Journal*, **56**, 1202-1203. <https://doi.org/10.1007/s10597-020-00689-2>
- [32] Owusu, P.N. (2020) Digital Technology Applications for Contact Tracing: The New Promise for COVID-19 and beyond? *Global Health Research and Policy*, **5**, Article No. 36. <https://doi.org/10.1186/s41256-020-00164-1>
- [33] Anirudh Chebolu, V.S., Arkajit Datta, N.A.B., Chebolu, S. and Rao, K.R.M. (2021) Pandemic Penetration: Factors for Measurement. *SN Computer Science*, **2**, Article No. 451. <https://doi.org/10.1007/s42979-021-00844-w>
- [34] Paul, S.K., Jana, S. and Bhaumik, P. (2021) Explaining Causal Influence of External Factors on Incidence Rate of Covid-19. *SN Computer Science*, **2**, Article No. 465. <https://doi.org/10.1007/s42979-021-00864-6>
- [35] Rahman, M.M., Islam, M., Manik, M.H., Islam, R. and Al-Rakhmi, M.S. (2021) Machine Learning Approaches for Tackling Novel Coronavirus (COVID-19) Pandemic. *SN Computer Science*, **2**, Article No. 384. <https://doi.org/10.1007/s42979-021-00774-7>
- [36] Shoaib, M., Salahudin, H., Hammad, M., Ahmad, S., Khan, A.A., Khan, M.M., Baig, M.A.I., Ahmad, F. and Ullah, M.K. (2021) Performance Evaluation of Soft Computing Approaches for Forecasting COVID-19 Pandemic Cases. *SN Computer Science*, **2**, Article No. 372. <https://doi.org/10.1007/s42979-021-00764-9>
- [37] Swapnarekha, H., Behera, H.S., Nayak, J., Naik, B. and Kumar, P.S. (2021) Multiplicative Holts Winter Model for Trend Analysis and Forecasting of COVID-19 Spread in India. *SN Computer Science*, **2**, Article No. 416. <https://doi.org/10.1007/s42979-021-00808-0>