Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Journal Pre-proof

Modelling COVID-19 daily cases in Senegal using a generalized Waring regression model

Lucien Gning, Cheikh Ndour, J.M. Tchuenche

PII: S0378-4371(22)00221-7
DOI: https://doi.org/10.1016/j.physa.2022.127245
Reference: PHYSA 127245

To appear in: Physica A

Received date: 29 November 2021
Revised date: 8 March 2022

Please cite this article as: L. Gning, C. Ndour and J.M. Tchuenche, Modelling COVID-19 daily cases in Senegal using a generalized Waring regression model, Physica A (2022), doi: https://doi.org/10.1016/j.physa.2022.127245.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Elsevier B.V. All rights reserved.
Highlights

Modelling COVID-19 daily cases in Senegal using a generalized Waring regression model

Lucien Gning, Cheikh Ndour, J.M. Tchueneche

- The COVID-19 pandemic has triggered substantial economic and social disruptions worldwide.
- The generalized Waring regression model is used to forecast the daily confirmed COVID-19 cases in Senegal.
- The generalized Waring regression has an advantage over other count regression models because it considers factors that could affect the number of daily COVID-19 cases, but cannot be observed or measured.
- The generalized Waring regression model fits the data better and could better explain some of the intrinsic characteristics of the disease dynamics.
Modelling COVID-19 daily cases in Senegal using a generalized Waring regression model

Lucien Gning\textsuperscript{a}, Cheikh Ndour\textsuperscript{b} and J.M. Tchuenche\textsuperscript{c,d}

\textsuperscript{a}Laboratoire d’études et de recherches en statistiques et développement, Université Gaston BERGER, Saint-Louis, Senegal
\textsuperscript{b}Laboratoire de mathématiques et de leurs applications, Université de Pau et des Pays de l’Adour, Pau, France
\textsuperscript{c}School of Computer Science and Applied Mathematics, University of the Witwatersrand, Private Bag 3, Wits 2050, Johannesburg, South Africa
\textsuperscript{d}School of Computational and Communication Sciences and Engineering, Nelson Mandela African Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Coronavirus
Epidemiology
Count Regression Models
Generalized Waring Regression

\textbf{ABSTRACT}

The rapid spread of the COVID-19 pandemic has triggered substantial economic and social disruptions worldwide. The number of infection-induced deaths in Senegal in particular and West Africa in general are minimal when compared with the rest of the world. We use count regression (statistical) models such as the generalized Waring regression model to forecast the daily confirmed COVID-19 cases in Senegal. The generalized Waring regression model has an advantage over other models such as the negative binomial regression model because it considers factors that cannot be observed or measured, but that are known to affect the number of daily COVID-19 cases. Results from this study reveal that the generalized Waring regression model fits the data better than most of the usual count regression models, and could better explain some of the intrinsic characteristics of the disease dynamics.

\section{1. Introduction}

COVID-19 is an acute and severe viral respiratory infection, with high fever, fatigue, cough and shortness of breath as the most common symptoms. The ongoing pandemic has affected more than 184 million people in the world (Figure 1), with severe public health and economic/financial crisis globally (Bubar, K., Kissler, Lipsitch, Cobey, Grad and Larremore (2021)), and countries such as Senegal were constrained to the most restrictive preventive measures such as lockdowns, closure or limited openings of shops and schools (Haug (2020)). As of July 9, 2021, the Word Health Organization’s (WHO) statistics reported that death rate of detected infectious individuals with COVID-19 was on average of 2.16 percent (0.1% - 9.3%) depending on individuals’ age group and other factors such as health status as well as immunity system among the many factors (https://www.who.int/). However, with the persistent issue of under-reporting, asymptomatic or untested cases, one would expect the actual case fatality rate to be (perhaps much) lower. In Senegal, COVID-19 third wave tends to have a higher peak than the first and second waves, with the number of deaths following the same pattern. A powerful tool used to mitigate the spread and potential impact of the pandemic is mathematical modelling to forecasting the long term dynamical spread of the disease (Dietz and Heesterbeek (2002)), and to help support, using data the effective allocation and mobilization of scarce public health resources (Ajbar, Alqahtani and Boumaza (2021), Shan and Zhu (2021)).

From the early days of the epidemic to its current state as a pandemic, numerous mathematical models have been developed. These include (1) compartmental model (Chowell, Hyman, Bettencourt and Castillo-Chavez (2009), Martcheva (2015)) of the SIR (Susceptible-Infected-Recovered) type and its variants (Lopez and Rodo (2021), Saikia, Bora and Bora (2021), Venables and Ripley (2015)), (2) statistical models (Chan, Chu, Zhang and Nadarajah (2021), Chu (2021), Odhiambo, Okungu and Mutuura (2020)), and particularly count data regression models used for example to modelling and forecasting the number of daily COVID-19 cases reported in various countries (Chan et al. (2021), Odhiambo et al. (2020)).

We aim to investigate the spread of COVID-19 by fitting a generalized Waring regression model (GWRM) (Rodriguez-Avi, Conde-Sanchez, Saez-Castillo, Olmo-Jiménez and Martinez-Rodriguez (2009)) to the number of new daily cases in Senegal, and make short time predictions. The application of this model to COVID-19 is to the best of our...
knowledge relatively new. Next, we compare the performance of this model with that of the Poisson regression model (PRM) and the negative binomial regression model (NBRM). An extension of the negative binomial distribution that allows three sources of variation to be distinguished (Rodriguez-Avi et al. (2009)) is the univariate generalized Waring distribution (UGWD) (Harris, Hilbe and Hardin (2014), Irwin (1968), Rodriguez-Avi, Conde-Sanchez, Saez-Castillo, Olmo-Jiminez and Martinez-Rodriguez (2007), Xekalaki (1983)), while the GWRM is a regression model with a UGWD as its underlying distribution. One main advantage of the implementation of the GWRM over other traditional models such as the NBRM is that the GWRM distinguishes individual heterogeneity due to intrinsic (internal) factors, and those due to external factors (e.g., covariates that influence the variability of data but because they cannot be observed or measurable are not included in the model).

The outline of the rest of this paper is as follows. Section 2 describes our dataset. In Section 3, we detail the different models used. Section 4 provides a practical application, with discussion of the results. Conclusion is provided in Section 5.

Figure 1: Total confirmed cases of COVID-19 as of 9th July 2021 (source https://www.worldometers.info/coronavirus/ accessed on 9 July 2021). The unit measurement is one Million (M)

2. Data

Our dataset consists of the historical daily new confirmed COVID-19 cases in Senegal. Cases during the first wave of the outbreak were collected from the Senegalese Ministry of Health website (https://www.sante.gouv.sn/), available at this github repository https://github.com/senegalouvert/COVID-19. The timeline of the observations are from 2nd of March 2020 to the ends of 22nd of October 2020. The daily reported number of new cases and the cumulative number of reported cases are depicted in Figure 2 while the ratio number of reported cases over number of screening tests is depicted in Figure 3. We simulate the possibility of near real-time predictions during an actual outbreak to test the predictive power of each of the models by using 90% of the data for estimation, and the remaining 10% for evaluation. The other variables of the data used are the following:

- the number of daily screening tests
- the number of daily contact COVID-19 cases. A contact case is a COVID-19 case transmission in which the person has been infected by another person with whom he was in contact, whether it is a member of family, a friend, or a neighbour, etc … That is to say the person exactly knows the person who contaminated him/her and where he/she was infected.
- the daily number of community COVID-19 cases. A community case is a COVID-19 case transmission in which the person does not know where he is infected.
• the number of daily COVID-19 imported cases (from foreign countries or from outside the community).
• the number of total recovered cases.
• the number of total deaths due to COVID-19.
• the number of total evacuated cases at foreign countries.
• the number of daily critical cases.

The descriptive statistics of all variables used in this paper are summarized in Table 1.

![Figure 2](image1.png)

**Figure 2:** Top: Daily reported cases. Bottom: Cumulative number of reported cases.

![Figure 3](image2.png)

**Figure 3:** Ratio number of reported cases over number of screening tests.
3. Models Framework

In our framework $Y_t$ denotes the newly confirmed COVID-19 cases on day $t$, $t = 1, \ldots, T$.
In what follows, $\mathcal{F}_{t-\tau}$ denotes the history up to day $t-\tau$, $1 \leq \tau < t$, $x'_t = (1 \ x_1 \ \cdots \ x_p)$ is the vector of covariates on day $t$, and $\beta' = (\beta_0 \ \beta_1 \ \cdots \ \beta_p)$ is the vector of the parameters regression on the covariates.

3.1. Poisson regression model (PRM)

In this model we assume, for the conditional distribution $Y_t$ given $\mathcal{F}_{t-\tau}$, that:

$$Y_t|\mathcal{F}_{t-\tau} \sim \text{Poisson}(\lambda_{x'_t \beta}).$$

(1)

where $\lambda_{x'_t \beta} = \exp(x'_t \beta)$. It’s follow that:

$$\text{Var}(Y_t|\mathcal{F}_{t-\tau}) = E(Y_t|\mathcal{F}_{t-\tau}) = \lambda_{x'_t \beta} = \exp(x'_t \beta).$$

(2)

The conditional mean is then equal to the conditional variance (equal dispersion) for this model that constitutes a major shortcoming for its application on count data.

3.2. Negative binomial regression model (NBRM)

In this model, we make the following distributional assumption for $Y_t$ given $\mathcal{F}_{t-\tau}$:

$$Y_t|\mathcal{F}_{t-\tau}, \lambda_{x'_t \beta} \sim \text{NegativeBinomial}(\lambda_{x'_t \beta}, \phi_{x'_t \beta}).$$

(3)

where $\lambda_{x'_t \beta}$ is a gamma distributed variable:

$$\lambda_{x'_t \beta} \sim \text{Gamma}(\theta, \nu_{x'_t \beta}).$$

(4)

Then, using straightforward calculations we obtain (see, for example, Cameron and Trivedi (2013), Rodriguez-Avi et al. (2009) or Vilchez-Lopez, Saez-Castillo and Olmo-Jimenez (2016)):

$$Y_t|\mathcal{F}_{t-\tau} \sim \text{NegativeBinomial}(\theta, \nu_{x'_t \beta}).$$

(5)

where $\nu_{x'_t \beta} = 1/(1 + \nu_{x'_t \beta})$. Its follows, for the NBRM, that the conditional mean is given by:

$$\mu_{x'_t \beta} = E(Y_t|\mathcal{F}_{t-\tau}) = E(\lambda_{x'_t \beta}) = \theta \nu_{x'_t \beta}.$$  

(6)

Remark that this model is a Poisson-gamma mixture model and, for the gamma distribution, the parameter $\theta$ does not depend on the covariates contrary to $\nu_{x'_t \beta}$. The model parametrized that way is referred as NegbinII model in the literature (see Cameron and Trivedi (2013)). The variance of the model is now given by:

$$\text{Var}(Y_t|\mathcal{F}_{t-\tau}) = E(\lambda_{x'_t \beta}) + \text{Var}(\lambda_{x'_t \beta}) = \mu_{x'_t \beta} + 1/\theta \nu_{x'_t \beta}^2.$$ 

(7)

The decomposition of this variance is interpreted as follow: the first term of the model variance represents the variability due to randomness inherent in any random phenomenon and the second to differences between the days between 2 March 2020 and 22 October 2020 in our practical application. For more details on the interpretation of the variance decomposition see Rodriguez-Avi et al. (2009) or Vilchez-Lopez et al. (2016).

### Table 1

Descriptive Statistics of COVID-19 data in Senegal

| Variable          | Observations | Total  | Mean    | Std.Dev. | Min. | Max. |
|-------------------|--------------|--------|---------|----------|------|------|
| screening tests   | 235          | 207997 | 888.88  | 484.15   | 1    | 1963 |
| confirmed cases   | 235          | 15299  | 65.10   | 49.01    | 0    | 207  |
| contact cases     | 235          | 10912  | 46.43   | 39.66    | 0    | 169  |
| community cases   | 235          | 4046   | 17.22   | 20.02    | 0    | 115  |
| imported cases    | 235          | 348    | 1.48    | 3.04     | 0    | 27   |
3.3. Generalized Waring regression model (GWRM)

The GWMR is an extension of NBRM. Contrary to the last model, it allows to distinguish the variability due to individual differences related to the covariates included in the model, to the one only due to individual differences not related to the model covariates (Rodriguez-Avi et al. (2009), Vilchez-Lopez et al. (2016)). Next, we present a brief background on the genesis of the GWRM and some properties of the model.

In this model we assume that \( Y_i \) given \( \mathcal{F}_{t-r} \) has a univariate generalized Waring distribution:

\[
Y_i | \mathcal{F}_{t-r} \sim \text{UGWD}(\theta_{x_{t-r}}, k, \rho).
\]

This implies that the probability mass function of \( Y_i | \mathcal{F}_{t-r} \) is given by (Irwin (1968)):

\[
f(y|x) = \frac{\Gamma(\theta_x + \rho) \Gamma(k + \rho) (\theta_x + k + \rho)_x} {\Gamma(r_x + k + \rho) (\theta_x + k + \rho)_y} y!, \quad y = 0, 1, 2, \ldots,
\]

where \( \theta_x, k, \rho > 0 \) and \( (\alpha) = \frac{\Gamma(\alpha + 1)} {\Gamma(\alpha)} \) for \( \alpha > 0 \) and \( \Gamma \) is the gamma function defined by \( \Gamma(x) = \int_0^\infty u^{x-1} \exp(-u) \, du \), for \( x > 0 \). Then, the conditional mean is given by:

\[
E(Y_i | \mathcal{F}_{t-r}) = \mu_{x_{t-r}} = \frac{\theta_{x_{t-r}} k}{\rho - 1}, \quad (\rho > 1).
\]

In addition, linking the conditional mean with the explanatory variables as \( \mu_{x_{t-r}} = \exp(\mathbf{x}_{t-r}^T \beta) \), we have:

\[
\theta_{x_{t-r}} = \frac{\mu_{x_{t-r}} (\rho - 1)} {k}.
\]

Furthermore, the variance of the GWMR is given by (Rodriguez-Avi et al. (2009), Vilchez-Lopez et al. (2016)):

\[
\text{Var}(Y_i | \mathcal{F}_{t-r}) = \mu_{x_{t-r}} + \frac{k + 1}{\rho - 2} \mu_{x_{t-r}} + \frac{k + \rho - 1} {\rho - 2} \mu_{x_{t-r}}^2, \quad (\rho > 2).
\]

The decomposition of this variance is interpreted as follow:

- the first term referred as randomness represents the variability inherent to any random phenomenon
- the second one called liability represents the variability due to individual differences related to the explanatory variables
- the third one called proneness is the variability due to individual differences not related to the explanatory variables included in the model

Finally, note that the GWRM is a NBRM-beta mixture model and then a limiting case of the NBRM (Rodriguez-Avi et al. (2009)). For further details of this regression model the reader can refer to Rodriguez-Avi et al. (2009) or Vilchez-Lopez et al. (2016).

3.4. Maximum likelihood estimation

Each of the three models considered above, namely the PRM, the NBRM and the GWRM was fitted using the maximum likelihood estimation. That is by maximizing:

\[
\ell_1(\beta) = \sum_{t=r+1}^T \left( y_i x_{t-r}^T \beta - \exp(x_{t-r}^T \beta) - \log(y_i!) \right),
\]

\[
\ell_2(\beta, \theta) = \sum_{t=r+1}^T \left( y_i \log \left( \frac{\exp(x_{t-r}^T \beta)} {\theta + \exp(x_{t-r}^T \beta)} \right) - \log \left( 1 + \frac{\exp(x_{t-r}^T \beta)} {\theta} \right) + \log \left( \frac{\Gamma(\theta + y_i)} {\Gamma(\theta)} \right) \right),
\]

\[
\ell_3(\beta, \rho, k) = \sum_{t=r+1}^T \left( \log \left( \frac{\Gamma(\theta_{x_{t-r}} + \rho) \Gamma(\rho + 1) \Gamma(\theta_{x_{t-r}} + y_i) \Gamma(k + y_i)} {\Gamma(\rho) \Gamma(\theta_{x_{t-r}}) \Gamma(k) \Gamma(\theta_{x_{t-r}} + k + \rho + y_i)} \right) \right), \quad \text{where} \quad \theta_{x_{t-r}} = \frac{(\rho - 1) \exp(x_{t-r}^T \beta)} {k}.
\]
The numerical calculations of the maximum likelihood estimators of the model parameters are carried out with the open source software R (Team (2020)). Note that, because PRM and NBRM are special cases of the generalized linear model, we estimate their parameters using the iteratively re-weighted least squares (IRLS) algorithm (for more details see Hilbe (2011)). The PRM case is implemented in the \texttt{glm} function included in the R package \texttt{stats}, while the NBRM case is implemented in the \texttt{glm.nb} included in the package MASS. Finally, we use the R package GWRM developed by Vilchez-Lopez et al. (2016) to fit the GWRM. The algorithm used in this package is based on non linear minimization, Nelder-mead optimization and L-BFGS-B methods (see Vilchez-Lopez et al. (2016)).

For adaptation and reproducibility, the source codes of all theses packages are available on the Comprehensive R Archive Network (CRAN) repository (http://cran.r-project.org/web/packages/GWRM).

4. Application

New daily COVID-19 cases is considered as a discrete count variable over-dispersion tendency. With comparison with the Poisson model, several causes might be responsible of this variability. First, there are external observable factors that could significantly influence the number of daily COVID-19 cases, such as:

- $x_1$: the number of daily contact COVID-19 cases.
- $x_2$: the daily number of community COVID-19 cases.
- $x_3$: the number of daily COVID-19 imported cases.

Second, there are missing external covariates which would affect the response variable, and the GWRM could provide more information on the data through the quantification of the effect and sources of variability.

Finally, the variable number of daily screening COVID-19 tests, has been included as an offset in the models. This means that the ratio number of new daily COVID-19 cases / number of daily screening COVID-19 tests is used as response variable for all models.

Choice of $\tau$

The 14-day COVID-19 incubation period from exposure to symptoms onset is denoted by the parameter $\tau$ - the median time is about 5 days (Zaki and Mohamed (2021)). In a machine learning point of view, $\tau$ can be considered as a hyper-parameter, then, in the following, we use a grid search process to automate the tuning of $\tau$. In Figure 4, we plot the Root Mean Square Error (RMSE) of the models as functions of $\tau$, and it can be observed that $\tau = 3$ is the optimal value.

![Figure 4: Relationship between the models RMSE and the hyperparameter $\tau$](image)

4.1. Results and discussion

We fit the models from Section 3 to Senegal data. The computed values of the log-likelihood function shown in Table 2) enables us to evaluate the accuracy of the model fit, the Akaike and the Bayesian information criteria.
Table 2

| Model  | Log-likelihood | AIC    | BIC    | RMSE   | MAE    | ME    |
|--------|----------------|--------|--------|--------|--------|-------|
| PRM    | −1735.0286     | 3478.057 | 3491.170 | 80.47885 | 66.01209 | −66.01209 |
| NBRM   | −919.2790      | 1848.558 | 1864.949 | 80.43183 | 65.92999 | −65.92999 |
| GWRM   | −908.5021      | 1829.004 | 1848.673 | 29.93170 | 22.89277 | 1.33555 |

Table 3

| Covariate | Estimate | Std. Error | z value | P(> |z|) |
|-----------|----------|------------|---------|------|
| (Intercept) | −3.121   | 7.421e−02  | −42.05  | 0.000e+00 |
| contact \(x_1\) | 7.983e−03 | 9.468e−04 | 8.432   | 3.399e−17 |
| community \(x_2\) | 1.876e−03 | 1.466e−03 | 1.280   | 2.007e−01 |
| imported \(x_3\) | −2.258e−02 | 1.664e−02 | 1.357   | 1.748e−01 |

In addition, for each model, using a cross-validation procedure, we calculate in Table 2 the values of RMSE, the Mean Absolute Error (MAE), and the Mean Error (ME). Our results show that the best model out of the three is the Generalized Waring regression model (GWRM).

In Table 3, we display the maximum likelihood parameter estimates for the GWRM (the best fitted model), their respective standard errors, and the associated partial Wald tests (statistics and p-values). The estimates of \(k\) and \(\rho\) are respectively 53.68050 and 12.85094, with standard errors of 46.012067 and 4.668843. Note that the GWRM includes only the daily number of COVID-19 contact cases as a significant covariate at 5% significance level.

Now, we focus on how well the GWRM fits the observed data. We investigate this issue by carrying out a visual residual analysis. Indeed, in Figures 5, 6, and 7, we depict the quantile quantile plots (QQ-plots) with simulated envelope for residuals of type deviance, of type Pearson and of type response (for more details see Atkinson (1985), Vieira, Hinde and Demetrio (2000)), Vilchez-Lopez et al. (2016).

For the deviance and Pearson residuals, all the plotted points fall within the boundaries of the envelope, which indicates a good accuracy of the fitted model. However, for the response residuals, there are some points outside the envelope, a sign of a potential lack of accuracy, and hence no evidence against the adequacy of the fitted model. Thus, the GWRM is appropriate to fit the data. We observe that, on average, the proneness represents 50.18 percent of the variability of the model while the liability represents 41.57 percent and the randomness 8.25. This means about 50% of the variation of the model are due to external factors that are unknown because they could be difficult to observe or to measure.

Figure 8 depicts the relationship between the proportion of each component of the variance and the daily new COVID-19 cases. It can be seen that, on average, the randomness and the liability decrease as the number of new daily COVID-19 cases increase, while the proneness increases on average.

Figure 9 shows the proportion of the variance related to randomness, liability and proneness, in terms of the number of screening tests. In general, the variability in the number of new daily COVID-19 cases due to randomness and liability decreases as the number of screening tests increases, whereas the variability due to proneness increases. So, we can deduce that increasing the number of screening tests emphasizes the role of the daily characteristics as a cause of differences between days in relation to the number of new daily COVID-19 cases, whereas randomness and liability are somewhat less relevant.
The Figures 10, 11 and 12 illustrate the relationships between the model’s covariates and the variance related to randomness, liability and proneness. In general, the variability in the number of daily new COVID-19 infections due to randomness and liability decreases as the number of daily contact or the number of daily community cases increases, whereas the variability due to proneness increases. Thus, an increase of the daily number of cases emphasizes the role of the proneness, whereas randomness and liability become less relevant. In contrast, the variability in the number of new daily infections varies randomly as the number of daily imported cases increases (Figure 12). However the fraction of proneness is globally more important than the other sources of variation. It is important to note that the imported cases represent 2.24% of the total COVID-19 cases in Senegal during the first stage outbreak with a peak at 27 on June 13, 2020, and, the GWRM does not include this variable as a significant covariate at 5% significance level.
Figure 7: Response residuals against the order statistics of the normal distribution from the fitted GWRM.

Figure 8: Relationship between proportion of each component and the new daily cases. The blue line represents the proportion of the variability due to randomness. The red line depicts the proportion of the variability due to liability. The black line represents the proportion of the variability due to proneness.
Figure 9: Proportion of the variance components for the daily number of screening tests. The blue line represents the proportion of the variability due to randomness. The red line depicts the proportion of the variability due to liability. The black line represents the proportion of the variability due to proneness.

Figure 10: Proportion of the variance (randomness, liability, proneness) for the daily number of contact cases. The blue line represents the proportion of the variability due to randomness. The red line depicts the proportion of the variability due to liability. The black line represents the proportion of the variability due to proneness.
5. Conclusion

The rapid spread of the COVID-19 pandemic has triggered substantial economic and social disruptions worldwide. Several mathematical studies have been proposed to investigate and forecast long term dynamics of the disease. We use count regression models such as the generalized Waring regression model to forecast the daily confirmed new COVID-19 cases in Senegal. We fitted the generalized Waring model (GWRM) to the Senegal data (daily number of confirmed new COVID-19 infections). Results from this study reveal that the generalized Waring regression model fits better the data than most of the usual count data such as the negative binomial regression model (NBRM). Therefore the former model can explain specific characteristics of the studied phenomenon. We also analyze the variance decomposition for each day. An advantage of the GWRM is that it accounts for factors that are known to affect the number of daily COVID-19 cases.

This study is not exhaustive, and future work can explore several other avenues such as modeling the daily mortality due to Covid-19 using a GWRM, investigating (1) the impact of testing capacity on the number of detected daily cases (Zhan, Chen and Zhang (2021)), (2) the effect of various prevention and therapeutic intervention measures on COVID-19 dynamics. One can also try to model the daily number of COVID-19 cases using Generalized Waring auto-regressive process in order to understand better the COVID-19 dynamics. A possible issue with daily confirmed cases data is the potential of weekly seasonality effects, few tests performed over weekends, hence less cases and oscillations in data. In our future study, we plan to investigate - if there are any such issues in the Senegal data and how to correct this for further improvements - whether the generalized Waring regression model could capture the disease multiple waves.

Figure 11: Proportion of the variance components for the daily number of community cases. The blue line represents the proportion of the variability due to randomness. The red line depicts the proportion of the variability due to liability. The black line represents the proportion of the variability due to proneness.

Figure 12: Proportion of the variance components for the daily number of imported cases. The blue line represents the proportion of the variability due to randomness. The red line depicts the proportion of the variability due to liability. The black line represents the proportion of the variability due to proneness.
Acknowledgements

The authors would like to thank the editor and the referees for careful reading and comments which improved the paper.

References

Ajbar, A., Alqahtani, R., Boumaza, M., 2021. Dynamics of an SIR-based covid-19 model with linear incidence rate, nonlinear removal rate, and public awareness. Front. Phys. 9, 634251.
Atkinson, A., 1985. Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford Science Publications, London.
Bubar, K., K., R., Kissler, S., Lipsitch, M., Cobey, S., Grad, Y., Larrere, D., 2021. Model-informed covid-19 vaccine prioritization strategies by age and serostatus. Science 371(6532), 916–921.
Cameron, A., Trivedi, P., 2013. Regression Analysis of Count Data. Cambridge University Press, New York.
Chan, S., Chu, J., Zhang, Y., Nadarajah, S., 2021. Count regression models for covid-19. Physica A: Statistical Mechanics and its Applications 563, 125460.
Chowell, G., Hyman, J., Bettencourt, L., Castillo-Chavez, C., 2009. Mathematical and statistical estimation approaches in epidemiology. Springer, New York.
Chu, J., 2021. A statistical analysis of the novel coronavirus (covid-19) in Italy and Spain. PLoS One 16(3), e0249037.
Dietz, K., Heesterbeek, J., 2002. Bernoulli’s epidemiological model revisited. Mathematical Biosciences 180(1-2), 1–21.
Harris, T., Hilbe, J., Hardin, J., 2014. Modelling count data with generalized distributions. The Stata Journal 14(3), 562–579.
Haug, N., 2020. Ranking the effectiveness of worldwide covid-19 government interventions. Nature Human Behaviour 4, 1303–1312.
Hilbe, J., 2011. Modeling Count Data. Springer, Berlin.
Irwin, J., 1968. The generalized waring distribution applied to accident theory. Journal of the Royal Statistical Society. Series A (General) 131(2), 205–225.
Lopez, L., Rodo, 2021. A modified SEIR model to predict the covid-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. Results in Physics 21, 103746.
Martcheva, M., 2015. An introduction to Mathematical Epidemiology. Springer, New York.
Odhiambo, J., Okungu, J., Mutuura, C., 2020. Stochastic modeling and prediction of the covid-19 spread in Kenya. Engineering Mathematics 4, 31–35.
Rodriguez-Avi, J., Conde-Sanchez, A., Saez-Castillo, A., Olmo-Jimenez, M., Martinez-Rodriguez, A., 2007. A new generalization of the waring distribution. Computational Statistics and Data Analysis 51(12), 6138–6150.
Rodriguez-Avi, J., Conde-Sanchez, A., Saez-Castillo, A., Olmo-Jimenez, M., Martinez-Rodriguez, A., 2009. A generalized waring regression model for count data. Computational Statistics and Data Analysis 53(10), 3717–3725.
Saikia, D., Bora, K., Bora, M.P., 2021. Covid-19 outbreak in india: an seir model-based analysis. Nonlinear Dynamics 104, 4727–4751.
Shan, C., Zhu, H., 2021. Bifurcations and complex dynamics of an sir model with the impact of the number of hospital beds. Journal of Differential Equations 257(5), 1662–1688.
Team, R.D.C., 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL: http://www.R-project.org.Vienna.
Venables, W.N., Riplcy, B.D., 2015. Modern Applied Statistics with S. Fourth edition. Springer, New York.
Vieira, A., Hinde, J., Demetrio, C., 2000. Zero-inflated proportion data models applied to a biological control assay. Journal of Applied Statistics 27(3), 373–389.
Vilchez-Lopez, S., Saez-Castillo, A., Olmo-Jimenez, M., 2016. An R package for identifying sources of variation in overdispersed count data. PLoS ONE 11(12), e0167570.
Xekalaki, E., 1983. The univariate generalized waring distribution in relation to accident theory: Proneness, spells or contagion? Biometrics. 39(4), 887–895.
Zaki, N., Mohamed, E., 2021. E.a. the estimations of the covid-19 incubation period: A scoping reviews of the literature. Journal of Infection and Public Health 14(5), 638–646.
Zhan, C., Chen, J., Zhang, H., 2021. An investigation of testing capacity for evaluating and modeling the spread of coronavirus disease. Information Sciences 561, 211–229.
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Lucien GNING Ph.D