A simple statistic approximation to 2019–nCoV confirmed cases in mainland China: A data sight in the early phase of the outbreak

Héctor D. Molina Ruiz

Universidad Autónoma del Estado de Hidalgo, Instituto de Ciencias Básicas e Ingeniería, Abasolo 600, Pachuca de Soto, Hidalgo, México. Tel.: +52 (771) 717 2000, Ext.: 5850

Resumen

Existen algunas enfermedades de importancia mundial, el caso más remembrado es el de la peste bubónica, que provocó una drástica reducción en la población mundial. Los dos últimos síndromes respiratorios o infecciones del sistema respiratorio, que afectaron a la población mundial, fueron el SARS (*Severe Acute Respiratory Syndrome* – Síndrome Respiratorio Agudo Severo) y MERS (*Middle Eastern Respiratory Syndrome* – Síndrome Respiratorio del Medio Oriente). 2019 vuelve a activar las alarmas mundiales con el brote de un nuevo coronavirus, altamente contagioso. En la primera semana de febrero de 2020, el n-CoV 2019, también conocido como *Wuflu*, coronavirus de Wuhan o neumonía de Wuhan, según cifras oficiales, contaba con más de 30,000 casos confirmados en China continental y 28 países con casos confirmados en todo el mundo, aumentando la escala de sobresalto en la población. El presente documento muestra la aplicación de una regresión lineal, que pronostica el período n + i, con respecto de los datos diarios de infección, para el n-CoV 2019.

Abstract

There exist some mayor world diseases around the world, most reminded case, was bubonic pest, which provoke a mayor population disease in world population. Last two well-known respiratory syndromes or infections of the respiratory system, which hardly affect human beings were SARS (Severe Acute Respiratory Syndrome) and MERS (Middle Eastern Respiratory Syndrome). 2019 arises the world's alarms again with the outbreak of a novel highly transmittable coronavirus. In the first week of February 2020, 2019 n-CoV also known as Wuflu, Wuhan coronavirus or Wuhan pneumonia, according to official numbers, had more than 30,000 confirmed cases in main land China and 28 countries with confirmed cases around the world, ascending the scale of warning for global population. Present document shows a linear regression's application, which forecast the n+i period, on 2019 n-CoV infection's daily rates

Palabras clave: 2019 n-CoV, Infecciones diarias, Regresión lineal.
Keywords: 2019 n-CoV, Daily infections, Linear regression.

1. 1. INTRODUCTION

Today, 2019 n-CoV, presents a possible pandemic scenario. There exist some mayor world diseases around the world, most reminded case, was bubonic pest, which provoke a mayor population disease in world population. Last two well-known respiratory syndromes or infections of the respiratory system, which hardly affect human beings were SARS (Severe Acute Respiratory Syndrome) and MERS (Middle Eastern Respiratory Syndrome). Both are kinds of coronaviruses, which outbreak in 2002 (SARS) and 2012 (MERS), killing around 750 and 450 persons, respectively.

There is concern about a new coronavirus, the 2019-nCoV, as a global public health threat (Benvenuto, et al., 2020). 2019 n-CoV has a mayor impact, due to its spreading paths and behavior.

While there are several hypotheses about the origin of 2019-nCoV, the source of this ongoing outbreak remains elusive (Pradhan et al., 2020). Since December 29, 2019, pneumonia infection with 2019-nCoV has rapidly spread out from Wuhan, Hubei Province, China to most other provinces and other counties (Liu et al., 2020). To the

first week of February 2020, the have been infected more than 31,000 persons in mainland China (confirmed cases), with 637 deaths and 1728 recovered, reported in mainland China. In comparison with SARS (8,000 infected) and MERS (1,200 infected), novel 2019 n-CoV official numbers, outrange its predecessors.

In the following table, they are listed the main insight of 2019 n-CoV, from December 31st to January 20th (Table 1).

Table 1. Event highlights from 31 December 2019 to 20 January 2020

• On 31 December 2019, the WHO China Country Office was informed of cases of pneumonia unknown etiology (unknown cause) detected in Wuhan City, Hubei Province of China. From 31 December 2019 through 3 January 2020, a total of 44 case-patients with pneumonia of unknown etiology were reported to WHO by the national authorities in China. During this reported period, the causal agent was not identified.
• On 11 and 12 January 2020, WHO received further detailed information from the National Health Commission China that the outbreak is associated with exposures in one seafood market in Wuhan City.
• The Chinese authorities identified a new type of coronavirus, which was isolated on 7 January 2020.
• On 12 January 2020, China shared the genetic sequence of the novel coronavirus for countries to use in developing specific diagnostic kits.
• On 13 January 2020, the Ministry of Public Health, Thailand reported the first imported case of lab-confirmed novel coronavirus (2019-nCoV) from Wuhan, Hubei Province, China.
• On 15 January 2020, the Ministry of Health, Labour and Welfare, Japan (MHLW) reported an imported case of laboratory-confirmed 2019-novel coronavirus (2019-nCoV) from Wuhan, Hubei Province, China.
• On 20 January 2020, National IHR Focal Point (NFP) for Republic of Korea reported the first case of novel coronavirus in the Republic of Korea.

2. ABOUT THE 2019 N-COV

Coronaviruses (CoVs) are found in a wide variety of animals, in which they can cause respiratory, enteric, hepatic, and neurological diseases of varying severity (Woo et al., 2012). Coronaviruses (CoV) are single-stranded positive-sense RNA viruses that infect animals and humans (Pradhan et al., 2020).

Molecular clock analysis showed that the most recent common ancestor of all coronaviruses was estimated at approximately 8100 BC, and those of Alphacoronavirus, Betacoronavirus, Gammacoronavirus, and Deltacoronavirus were at approximately 2400 BC, 3300 BC, 2800 BC, and 3000 BC, respectively. Now a day, it can be found literature about four types of coronaviruses (CoVs), in the subfamily Coronavirinae (family Coronaviridae), that belong to the order Nidovirales: Alphacoronavirus (Ge et al., 2017; Han et al., 2019), Betacoronavirus (Ge, et al., 2017; Han et al., 2019), Gammacoronavirus (Thor et al., 2011; Torres et al., 2016) and Deltacoronavirus (Dong et al., 2015; Janetanakit et al., 2016; Boley et al., 2020).

The number of 2019-nCoV cases has increased rapidly in China, and cases have been transmitting to other countries (Liu et al., 2020). The transmission patterns of 2019-nCoV is similar to patterns of transmission documented in the previous outbreaks including by bodily or aerosol contact with persons infected with the virus (Pradhan et al., 2020). In Zhao et al. (2020), they estimated the mean R0 of 2019-nCoV ranging from 2.24 (95%CI: 1.96-2.55) to 3.58 (95%CI: 2.89-4.39) if the reporting effort has been increased by a factor of between 8- and 2-fold after the diagnostic protocol released on January 17, 2020 and many medical supplies reached Wuhan.

Ahmed, Quadeer & McKay (2020) provide a screened set of epitopes that can help guide experimental efforts towards the development of vaccines against 2019-nCoV. Gao et al. (2020), showed that the protease inhibitor binding sites of 2019-nCoV and SARS-CoV are almost identical, which means all potential anti-SARS-CoV

chemotherapies are also potential 2019-nCoV drugs. The key enzyme for coronavirus replication is the ~306 aa long main protease (M^{pro}). According to Liu & Wang (2020), drugs for the M^{pro} of SARS-CoV, two HIV-1 protease inhibitors (lopinavir and ritonavir), clinical application on 2019-nCoV patients, appears to be effective, demonstrating the importance of the drug binding site for suppressing 2019-nCoV M^{pro} activity.

3. LINEAR REGRESSION

In this section, they are calculated the linear regression of the daily confirmed infected persons with 2019 n-CoV, reported in WHO (2020) and wuflu.info web page.

Linear regression estimation is adjusted to the following equation.

$$\hat{Y} = a + bX \qquad \ldots \ldots \ldots \ldots \ldots \quad (1)$$

with:

$$a = \bar{y} - b\bar{x} \qquad \ldots \ldots \ldots \ldots \ldots \quad (2)$$

and

$$b = \frac{cov(X,Y)}{V(X)} = \frac{S_{XY}}{S_X^2} \qquad \ldots \ldots \ldots \ldots \quad (3)$$

Error estimation follows the next equations:

Explained variation

$$VE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad \ldots \ldots \ldots \ldots \quad (4)$$

Unexplained variation

$$VNE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad \ldots \ldots \ldots \ldots \quad (5)$$

Total variation

$$VT = \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad \ldots \ldots \ldots \ldots \quad (6)$$

$$VT = VE + VNE = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \ldots \quad (7)$$

In linear regression, determination coefficient is:

$$R^2 = 1 - \frac{VNE}{VT} \qquad \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \quad (8)$$

it indicates the adjustment's quality to regression line. Alternatively, determination coefficient can be calculated by:

$$R^2 = \frac{VE}{VT} \qquad \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \; \cdot \quad (9)$$

The following table contains the daily data report for confirmed cases of 2019 n-CoV (Table 2):

Table 1. Daily data report for confirmed cases of 2019 n-CoV

Day	(Xi)	Historic data (Yi)
01_20	0	278
01_21	1	440
01_22	2	571
01_23	3	830
01_24	4	1287
01_25	5	1975
01_26	6	2744
01_27	7	4515
01_28	8	5874
01_29	9	7711
01_30	10	9692
01_31	11	11791
02_01	12	14380
02_02	13	17236
02_03	14	20471
02_04	15	24363
02_05	16	28060
02_06	17	31211

Source: WHO (2020) & wuflu.info

Based on equations we have:

$$b = \frac{48,846.86}{26.92} = 1,814.744066$$

and,

$$a = 10190.5 - 1,814.744066 * (8.5) = -5,234.824561$$

with parameters estimation, we have:

$$\hat{Y} = -5,234.824561 + 1,814.744066 * X$$

Which means that period $t + 1$ (February 2nd or period 18) is:

$$\hat{y}_{t+1} = \hat{y}_{02\_07} = -5,234.824561 + 1,814.744066 * (18) = 27,430.56863 \; confirmed \; cases$$

With a R^2 equals to 0.9033, it means that adjusted regression fits 90% the data.

In the following table, it is shown the forecast from February 7th to February 29th (Table 3):

Table 2. daily data report for confirmed cases of 2019 n-CoV

Day	(Xi)	Historic data (Yi)
02_07	18	27430.56863
02_08	19	29245.31269
02_09	20	31060.05676
02_10	21	32874.80083
02_11	22	34689.54489
02_12	23	36504.28896
02_13	24	38319.03302
02_14	25	40133.77709
02_15	26	41948.52116
02_16	27	43763.26522
02_17	28	45578.00929
02_18	29	47392.75335
02_19	30	49207.49742
02_20	31	51022.24149
02_21	32	52836.98555
02_22	33	54651.72962
02_23	34	56466.47368
02_24	35	58281.21775
02_25	36	60095.96182
02_26	37	61910.70588
02_27	38	63725.44995
02_28	39	65540.19401
02_29	40	67354.93808

Source: Self elaboration using the linear regression method

4. DISCUSSION

The so called epidemic 2019 n-CoV has risen the global warning on a possible pandemic scenario, due that it is important to generate tools to continuous monitoring on this hot topic.

Despite linear regression (forecast) on period $t + 1$ (February 2nd or period 18), lays under period t (February 1st or period 17), it can be seen that at the end of the month (February 29th), there would be expected more than 67,354 infected persons, based on official data.
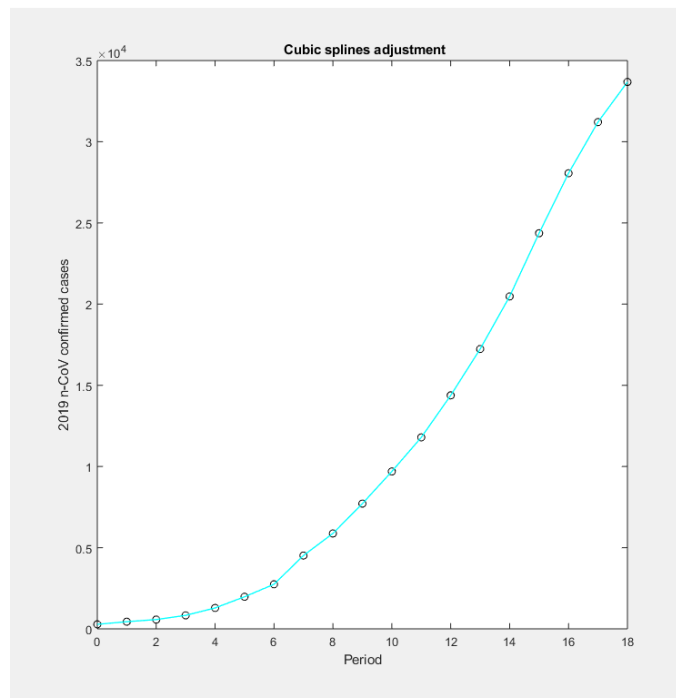
If it would be used an exponential adjustment, in the terms of:

$$y = 317.46e^{0.2825x}$$

That equation would provide a better adjustment, while $R^2 = 0.966$ which means that exponential adjustment fits the data in a 96.6%.

That adjustment would present a scenario in which the confirmed cases would be 51,295.16751, in period $t + 1$ (February 2nd or period 18) and 25'657,637.05 at the end of the month (February 29th).

Using the cubic splines (consult Schoenberg, 1964; Reinsch, 1967; Zernicke, Caldwell & Roberts, 1976; Wood & Jennings, 1979; Roncero-Peña, 2013), in the period $t + 1$ (February 2nd or period 18), the scenario would present a 33,681 confirmed cases picture (Figure 1).

Figure 1. Cubic splines adjustment. Source: Self elaboration with a script compiled in Matlab® 2019b.

[1]   Ahmed, F., Quadeer, A.A. & McKay, R.M. (2020). Preliminary identification of potential vaccine targets for 2019-nCoV based on SARS-CoV immunological studies, DOI: [10.1101/2020.02.03.933226], URL: [https://www.biorxiv.org/content/10.1101/2020.02.03.933226v1.full].

[2]   Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, A. & Ciccozzi, M., (2020). The 2019-new coronavirus epidemic: Evidence for virus evolution, Journal of Medical Virology, DOI: [10.1002/jmv.25688], URL: [https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.25688].

[3]   Boley, P. A., Alhamo, M. A., Lossie, G., Yadav, K., Vasquez-Lee, M., Saif, L. J. & Kenney, S. P. (2020). Porcine Deltacoronavirus Infection and Transmission in Poultry, United States, Emerging Infectious Diseases, 26(2), 255-265, DOI: [10.3201/eid2602.190346], URl: [https://wwwnc.cdc.gov/eid/article/26/2/19-0346_article?utm_source=Omeda&utm_medium=Email&utm_content=NL-Poultry+Future&utm_campaign=NL-Poultry+Future_20200124_0430].

[4]   Dong, N., Fang, L., Zeng, S., Sun, Q., Chen, H. & Xiao, S. (2015). Porcine Deltacoronavirus in Mainland China, Emerging Infectious Diseases, 21(12), 2254–2255, DOI: [10.3201/eid2112.150283], URL: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4672429/].

[5]   Gao, K., Nguyen, D.D., Wang, R. & Wei, W. (2020). Machine intelligence design of 2019-nCoV drugs, DOI: [10.1101/2020.01.30.927889], URL: [https://www.biorxiv.org/content/10.1101/2020.01.30.927889v1.full].

[6]   Ge, X.Y., Yang, W.H., Zhou, J.H., Li, B., Zhang, W., Shi, Z.L. & Zhang, Y.Z. (2017). Detection of alpha- and betacoronaviruses in rodents from Yunnan, China, Virology Journal, 14(98), DOI: [10.1186/s12985-017-0766-9], URL: [https://virologyj.biomedcentral.com/articles/10.1186/s12985-017-0766-9].

[7]   Han, Y., Du, J., Su, H., Zhang, J., Zhu, G., Zhang, S., Wu, Z. & Jin, Q. (2019). Identification of Diverse Bat Alphacoronaviruses and Betacoronaviruses in China Provides New Insights Into the Evolution and Origin of Coronavirus-Related Diseases, Frontiers in Microbiology, DOI: [10.3389/fmicb.2019.01900], URL: [https://www.frontiersin.org/articles/10.3389/fmicb.2019.01900/full].

[8]   Janetanakit, T., Lumyai, M., Bunpapong, N., Boonyapisitsopa, S., Chaiyawong, S., Nonthabenjawan, N., Kesdaengsakonwut, S. & Amonsin, A. (2016). Porcine Deltacoronavirus, Thailand, 2015, Emerging Infectious Diseases, 22(4), 757–759, DOI: [10.3201/eid2204.151852], URL: [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4806967/].

[9]   Liu, T., Hu, J., Kang, M., Lin, L., Zhong, H., Xiao, J., He, G., Song, T., Huang, Q., Rong, Z., Deng, A., Zeng, W., Tan, X., Zeng, W., Zhu, Z., Li, J., Wan, D., Lu, J., Deng, H., He, J. & Ma, W. (2020). Transmission dynamics of 2019 novel coronavirus (2019-nCoV), DOI: [10.1101/2020.01.25.919787], URL: [https://www.biorxiv.org/content/10.1101/2020.01.25.919787v1.full].

[10] Liu, X. & Wang, X.J. (2020). Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines, DOI: [10.1101/2020.01.29.924100], URL: [https://www.biorxiv.org/content/10.1101/2020.01.29.924100v1.full].

[11] Pradhan, P., Pandey, A.K., Mishra, A., Gupta, P., Tripathi, P.K., Menon, M.B., Gomes, J., Vivekanandan, P. & Kundu, B (2020). Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag, DOI: [10.1101/2020.01.30.927871], URL: [www.biorxiv.org/content/10.1101/2020.01.30.927871v1.full#ref-17].

[12] Roncero-Peña, B. (2013). Splines cúbicos suavizantes en el diseño naval, Tesis para la obtención de grado como Ingeniero Naval y Oceánico, Departamento de Matemática Aplicada y Estadística, Escuela Técnica Superior de Ingeniería Naval y Oceánica, Universidad Politécnica de Cartagena, URL: [http://repositorio.upct.es/handle/10317/3641?show=full].

[13] Schoenberg, I. J. (1964). On Interpolation by Spline Functions and its Minimal Properties, En: Approximation Theory / Über Approximationstheorie, 109–129, DOI: [10.1007/978-3-0348-4131-3_12], URL: [https://link.springer.com/chapter/10.1007/978-3-0348-4131-3_12].

[14] Thor, S. W., Hilt, D. A., Kissinger, J. C., Paterson, A. H. & Jackwood, M. W. (2011). Recombination in avian gamma-coronavirus infectious bronchitis virus, Viruses, 3(9), 1777-1799, DOI: [10.3390/v3091777], URL: [https://www.mdpi.com/1999-4915/3/9/1777/pdf].

[15] Torres, C. A., Hora, A. S., Tonietti, P. O., Taniwaki, S. A., Cecchinato, M., Villarreal, L. Y. B., & Brandão, P. E. (2016). Gammacoronavirus and deltacoronavirus in quail, Avian diseases, 60(3), 656-661, DOI: [10.1637/11412-032316-Reg.1], URL: [https://bioone.org/journals/avian-diseases/volume-60/issue-3/11412-032316-Reg.1/Gammacoronavirus-and-Deltacoronavirus-in-Quail/10.1637/11412-032316-Reg.1.pdf].

[16] WHO (2020). Novel Coronavirus (2019-nCoV), Situation Report – 1, World Health Organization, URL: [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4].

[17] Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H. & Yuen, K.-Y. (2012). Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. Journal of Virology, 86(7), 3995–4008, DOI: [10.1128/JVI.06540-11], URL: [www.ncbi.nlm.nih.gov/pubmed/22278237]

[18] Wood, G. A. & Jennings, L. S. (1979). On the use of spline functions for data smoothing. Journal of Biomechanics, 12(6), 477–479, DOI: [10.1016/0021-9290(79)90033-2], URL: [https://www.sciencedirect.com/science/article/abs/pii/0021929079900332?via%3Dihub].

[19] Zernicke, R. F., Caldwell, G. & Roberts, E. M. (1976). Fitting Biomechanical Data with Cubic Spline Functions, Research Quarterly. American Alliance for Health, Physical Education and Recreation, 47(1), 9–19. DOI: [10.1080/10671315.1976.10615335], URL: [https://shapeamerica.tandfonline.com/doi/10.1080/10671315.1976.10615335#.XcwOsDNKjIU].

[20] Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D. & Wang, M.H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak, DOI: [10.1016/j.ijid.2020.01.050], URL: [https://www.ijidonline.com/article/S1201-9712(20)30053-9/fulltext].

* **Correo autor:** *m_en_i_molina_ruiz@engineer.com; hmolina@uaeh.edu.mx*