

Forecasting Novel Corona Positive Cases in India using Truncated Information: A Mathematical Approach

Brijesh P. Singh

Department of Statistics, Institute of Science
Banaras Hindu University, Varanasi-221005 INDIA
Email: brijesh@bhu.ac.in

Abstract

Novel corona virus is declared as pandemic and India is struggling to control this from a massive attack of death and destruction, similar to the other countries like China, Europe, and the United States of America. India reported 2545 cases novel corona confirmed cases as of April 2, 2020 and out of which 191 cases were reported recovered and 72 deaths occurred. The first case of novel corona is reported in India on January 30, 2020. The growth in the initial phase is following exponential. In this study an attempt has been made to model the spread of novel corona infection. For this purpose logistic growth model with minor modification is used and the model is applied on truncated information on novel corona confirmed cases in India. The result is very exiting that till date predicted number of confirmed corona positive cases is very close to observed on. The time of point of inflexion is found in the end of the April, 2020 means after that the increasing growth will start decline and there will be no new case in India by the end of July, 2020.

Introduction

A novel corona virus is responsible for epidemic popularly known as COVID-19 is a new strain that has not been identified previously in humans. WHO declared COVID-19 a pandemic on March 11, 2020.^[1] The virus that caused the incidence of Severe Acute Respiratory Syndrome (SARS) in 2002 in China, Middle East respiratory syndrome (MERS) in 2012 in Saudi Arabia and the virus that causes COVID-19 are genetically related to each other, but the diseases they caused are quite different.^[2] These viruses, in general, are a family of viruses that target and affect mammal's respiratory systems. The SARS corona virus spread to humans via civet cats, while the MERS virus spread via dromedaries. In case of the novel corona virus, typically happens via contact with an infected animal, perhaps the common carriers are bats initial reports from seafood market in central Wuhan, China.

Novel corona virus is spreading throughout the world at alarming speed. Worldwide it has exploded to 1118684 cases and caused 58909 deaths by April 3, 2020.^[3] Developed countries like Italy, Spain, France and United State of America etc. are struggling to overcome from the

pressure created by novel corona virus. India with a huge population about 1.3 billion, amongst majority of the people are living in poor hygienic condition and the medical facilities like number of doctors and hospitals are less in India as compared to developed countries indicates that the situation of India will become very critical but comparatively better public health system and political control in India than the above developed countries. India reported 2545 cases novel corona confirmed cases as of April 2, 2020 and out of which 191 cases were reported recovered and 72 deaths occurred. The first case of novel corona is reported in India on January 30, 2020 when a student returned from Wuhan, China.^[4] The Government of India was quick to launch various levels of travel advisories beginning from February 26, 2020, with restrictions on travel to China and nonessential travel restrictions to Singapore, South Korea, Iran and Italy.^[5] The efforts to control by the Hon'ble Prime Minister Narendra Modi Ji through Janata Curfew (public curfew) on March 22, 2020, can be seen as the beginning of wide-scale public preventive measures. India has launched several social distancing measures and personal hygiene measures during the second week of March.^[6]

Since huge population of about 1.3 billion, thus India has chosen a flexible strategy of large-scale quarantine and limited testing because of less number of testing kits and also the cost of testing is too much. The country is relying on the people power; thousands of health-care workers are working out across the country to trace and quarantine people who might have had contact with those with novel corona. People are typically only tested if they develop symptoms. Countries such as South Korea isolated infected people based on widespread testing, but some scientists say that India's mass surveillance approach could achieve a similar goal, and be relevant for other low and low-middle income countries facing kit shortages. Under the lockdown, people are allowed out for essentials, such as food and medical care, but in most states people under quarantine are closely monitored by social workers and cannot leave their homes in some places. If public health workers do not trace all infected individuals during the lockdown, India will need to continue its period of stringent physical distancing.

For the spread of novel corona virus, when disease dynamics are still unclear, mathematical modeling helps us to estimate the cumulative number of positive cases in the present scenarios. Now India is entering in the mid stages of the epidemic. It is important to predict how the virus is likely to grow amongst the population. A mathematical modeling approach is a suitable tool to understand the dynamics of epidemic. In the study some mathematical approach to understand the dynamics of novel corona virus in India has been discussed.

Methodology

We obtained the truncated information on cumulative number of corona positive confirmed cases in India from March 13 to April 2, 2020 from covid19india.org.^[4] All cases are laboratory confirmed following the case definition by the Govt. of India. Some studies modeled the

epidemic curve obeying the exponential growth.^[7, 8] The nonlinear least square framework is adopted for data fitting and parameter estimation for 2019-nCoV at this early stage. In this study first exponential and then logistic growth curve has been used to model the novel corona epidemic, since epidemics grow exponentially not linearly. But it is surprising that exponential growth curve always provide increasing number of daily new cases. There is no saturation point. Another deterministic model used for understanding the dynamics of epidemic is the SIR model, which has been used to accurately predict incidence like SARS. In the SIR model, we need to know the input parameters first the stats we feed into the model.^[9, 10, 11] The first one is R_0 called the basic reproduction Number. It is essentially the number of new cases a single infected person will cause during their infectious period. It is one of the most important parameters for assessing any epidemic. Corona virus has an $R_0 \sim 2.4$. In contrast, the H₁N₁ virus had an $R_0 \sim 1.5$ in the 2009 swine flu epidemic.^[12] The R_0 will inform us about how many people will get infected with one infected person. Other one is the case fatality rate (CFR), which is the percentage of infected people that will die due to the infection. The CFR for corona virus has been reported between 0.5–4%. The lower values are more appropriate in resource better settings of medical facility. But SIR model assumes that every person is moving and has equal chance of contact with each and every other person among the population irrespective of the space or distance between different people. It is assumed that the transmission rate remains constant throughout the period of pandemic. Also this model considered to have the same transmission rate for who have been diagnosed and are in quarantine or those who have not been quarantined. The harmonic analysis methods and dynamic model estimates show that the number of COVID-19 infected would be 9225 (if there were 10 infected individuals as of March 1, 2020, who was not taking any precautions to spread), 17,986 (if there were 20) and 44,265 (if there were 50).^[13]

Growth Models

A growth curve is an empirical model of the evolution of a quantity over time. Growth curves are widely used in biology for quantities such as population size in population ecology and demography for population growth analysis, individual body height in physiology for growth analysis of individuals. Growth is also a key property of many systems such as an economic expansion, spread of an epidemic, the formation of a crystal, an adolescent's growth and the condensation of a stellar mass.

Linear growth

This is the simplest growth model, in which population grows at a constant rate over time. Linear growth is described by the equation

$$P_{t+1} = P_t + A \tag{1}$$

where P_t represents the numbers or size of the system at time t , P_{t+1} represents the system's numbers or size of the system one time unit later, and A is the system's (linear) growth rate. Many times this model fails to explain natural phenomenon.

Exponential growth (Unlimited population growth)

Another simple model describes exponential growth, in which population grows at a constant proportional rate over time. The relation may be expressed in either of two forms, depending on whether reproduction is assumed to be continuous or periodic.^[14] Exponential growth results in a continuous curve of increase or decrease, whose slope varies in direct relation to the size of the population.

$$P_t = y = P_0 e^{rt} \quad (2)$$

where r is the constant rate of growth, P_0 is the initial population size, and the variables t and P_t respectively represent time and the population at time t (Method 1). Another form of exponential curve is as follows

$$P_t = y = P_0 k^t \quad (3)$$

Where $k = \left(\frac{P_n}{P_0}\right)^{1/n}$ and that therefore the growth rate in (3) does not a constant growth rate. David

A. Swanson, University of California, USA used this type equation for prediction (Method 2).

We have used truncated information i.e. only 30 days information (from March 4 to April 2, 2020) on number of corona confirmed cases for the prediction purpose. We have used two equations of exponential curve given below

- I. $28e^{0.14t}$ up to March 31, 2020
- II. $28e^{0.15t}$ from March 31, 2020 onward (adjusting faster rate of occurrence of corona cases due to Tablighi spread)

With the current incidence of the novel corona virus going on, we hear about exponential growth. In this study, an attempt has been made to understand and analyze the data through exponential growth curve. The reason for using exponential growth curve for studying the pattern of novel corona virus incidence is that epidemiologists have studied these types of happenings and it is well known that the first period of an epidemic follows exponential growth. The exponential growth function is not necessarily the perfect representation of the epidemic. I have tried to fit exponential curve first, and at the next point to study the logistic growth curve because exponential curve is only fit the epidemic at the beginning. At some point, recovered people will not spread the virus anymore and when someone is or has been infected, the growth

will stop. Logistic Growth is characterized by increasing growth in the beginning period, but a decreasing growth after point of inflexion. For example in the corona virus case, the maximum limit would be the total number of exposed people in India because when everybody is infected, the growth will be stopped. After that the increasing rate of curve starts to decline and reach to the minimum.

In the figure 1, predicted values of the cumulative number of novel corona positive cases obtained by method 1 and 2 is drawn along with observed cumulative number of novel corona positive cases. Both the methods provide moderately good estimates but the tendency of both the curves are unlimited increasing. The rate of growth of Method 2 is slightly lesser than the rate of growth of Method 1. The number of total infected cases by April 30, 2020 would be about 144700 (Method 1) and 127700 (Method 2). If we do not adjust the Method 1 for Tablighi spread then the total infected cases by April 30, 2020 would be about 81810. Thus we can obtain the effect of Tablighi spread is about 75 percent.

Logistic growth (Sigmoidal)

The logistic model reveals that the growth rate of the population is determined by its biotic potential and the size of the population as modified by the natural resistance, or, in other words, by all the various effects of inherent characteristics, that are density dependence.^[15] Natural resistance increases as population size gets closer to the carrying capacity. Logistic growth is similar to exponential growth except that it assumes an essential sustainable maximum point. In exponential growth curve, the rate of growth of y per unit of time is directly proportional to y but in practice the rate of growth cannot be in the same proportion always. The logistic curve will continue up to certain level, called the level of saturation, sometimes called the carrying capacity, after reaching carrying capacity it starts declining. The factor y is called the momentum factor which increases with time t and the factor $(k - y)$ is known as the retarding factor which decreases with time. A system far below its carrying capacity will at first grow almost exponentially however, this growth gradually slows as the system expands, finally bringing it to a halt specifically at the carrying capacity.^[14, 15] The logistic relationship can be expressed as

$$P_t = y = \frac{k}{1 + e^{a+bt}} ; b < 0 \quad (4)$$

Logistic curve has a point of inflexion at half of the carrying capacity k . This point is the critical point from where the increasing rate of curve starts to decline. The time of point of inflexion can be estimate as $\frac{-a}{b}$. For the estimation of parameter of logistic curve, method of three selected point has been used. The estimate of the parameters can be obtained with equation given as:

$$k = \frac{y_2^2(y_1 + y_3) - 2y_1y_2y_3}{y_2^2 - y_1y_3} \quad (5)$$

$$b = \frac{1}{t_2 - t_1} \ln \left[\frac{(k - y_2)y_1}{(k - y_1)y_2} \right] \quad (6)$$

$$a = \ln \left[\frac{k - y_1}{y_1} \right] - bt_1 \quad (7)$$

Where y_1, y_2 and y_3 are the number at given time t_1, t_2 and t_3 respectively provided that $t_2 - t_1 = t_3 - t_2$.

You may also estimate the parameter a and b by method of least square after fixing k .

To predict confirmed corona cases on different day, logistic growth curve has been also used and found very exciting results. The truncated information on confirmed cases in India has been taken from March 13 to April 2, 2020. The estimated value of the parameters are as follows $k=18708.28$, $a=5.495$ and $b=-0.174$, with these estimates predicted values has been obtained and found considerably lower values than what we observed. On April 1 and 2, 2020 the number of confirmed corona cases are drastically increasing in some part of India due to some unavoidable circumstances thus there is an earnest need to increase carrying capacity of the model, thus it is increased and considered as 22000 and the other parameters a and b are estimated again which are $a=5.657$ and $b=-0.173$. The predicted cumulative number of cases is very close to the observed cumulative number of cases till date. The time of point of inflexion is obtained as 32.65 i.e. 35 days after beginning. We have taken data from March 13, 2020 so that the time of point of inflexion should be April 14, 2020 and by May 30, 2020 there will be no new cases found in the country. The distribution of the new cases is in the red color in the figure 2, which is quite normal and obvious. As mentioned in the above paragraph Method 1 provided natural estimate of the total infected cases by May 30, 2020 is 192400. This estimate is obtained when no preventive measure would be taken by the Government of India. The testing rate is lower in India than many western countries, so our absolute numbers is low, when government initiate faster testing process then we have observed more number of cases and found this logistic model fail to provide cumulative number of corona confirm cases after April 17, 2020 thus there is a need to modify this model. In order to the modification I have taken natural log of cumulative number of corona confirm cases instead of cumulative number of corona confirm cases as taken in the previous model. This model provides the carrying capacity is about 77000 cases and time of point of inflexion April 30, 2020. The present model provides reasonable estimate of the cumulative number of confirmed cases till date (see Appendix and figure 3) and by the end of July, 2020 there will be no new cases found in the country.

Conclusions

India is in the comfortable zone with a lower growth rate than other countries studied. Our mathematical model shows that, the epidemic is likely to stabilize with 77000 cases by the end of July, 2020. A study advocated of 49 days single lockdown i.e. lockdown up to May 13, 2020 reduce infected cases below 10 but our study contradict it.^[11] The government has adopted a strategy of large scale quarantine and limited testing to flatten the epidemic curve and reduce the death rate. The projections produced by the model and after their validation can be used to determine the scope and scale of measures that government need to initiate. In conclusion, if the current mathematical model results can be validated within the range provided here, then the social distancing and other prevention, treatment policies that the central and various state governments and people are currently implementing should continue until new cases are not seen. The spread from urban to rural and rich to poor populations should be monitor and control is an important point of consideration. Mathematical models have certain limitations that there are many assumptions about homogeneity of population in terms of urban/rural or rich/poor that does not capture variations in population density.

References

- [1] Coronavirus Disease (COVID-19) - events as they happen. Available at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-theyhappen>
- [2] Organization WH. WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China; Available from: <https://www.who.int/china/news/detail/09-01-2020-who-statement-regarding-cluster-of-pneumonia-cases-in-wuhan-china>
- [3] Coronavirus Update (Live): COVID-19 Virus Outbreak - Worldometer. Available at: <https://www.worldometers.info/coronavirus>.
- [4] India Covid-19 Tracker: A Crowdsourced Initiative, Available at: <https://www.covid19india.org>.
- [5] Consolidated Travel advisory in view of COVID-19 (26 February 2020), Government of India, Ministry of Health & FW.
- [6] Indian Council of Medical Research, Department of Health Research, Revised Strategy of COVID19 testing in India (version 3, dated 20 March 2020).
- [7] De Silva U, Warachit J, Waicharoen S, Chittaganpitch M. A preliminary analysis of the epidemiology of influenza A (H1N1) v virus infection in Thailand from early outbreak data, June-July 2009. *Euro surveillance* 2009; 14(31):19292.
- [8] Zhao S, Musa SS, Fu H, He D, Qin J. Simple framework for real-time forecast in a data limited situation: the Zika virus (ZIKV) outbreaks in Brazil from 2015 to 2016 as an example. *Parasites Vectors* 2019; 12(1):344.
- [9] Chatterjee K, Chatterjee K, Kumar A, Shankar S. Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model. *Medical Journal Armed Forces India* 2020; <https://doi.org/10.1016/j.mjafi>.
- [10] Mandal Sandip, Bhatnagar Tarun, Arinaminpathy Nimalan, Agarwal Anup et al. Prudent public health intervention strategies to control the corona virus disease 2019 transmission in

India: A mathematical model-based approach. Indian J Med Res 2020; Epub ahead of print
DOI: 10.4103/ijmr.IJMR_504_20

- [11] Singh Rajesh & Adhikari R. Age-structured impact of social distancing on the COVID-19 epidemic in India 2020. arXiv:2003.12055v1 [q-bio.PE]
- [12] Gupta Mohak Corona virus in India: Make or Break. Available at: https://medium.com/@mohakgupta_55841/coronavirus-in-india-make-or-break-5a13dfb9646d
- [13] Srinivasa Rao Arni S. R., Krantz Steven G., Kurien Thomas, Bhat Ramesh et al. Model-based retrospective estimates for COVID-19 or coronavirus in India: continued efforts required to contain the virus spread. Current Science 2020; 118(7):1023-25.
- [14] Shryock. Henry S. and Jacob S. Siegel. Methods and Materials of Demography. Washington: U.S. Dept. of Commerce, Bureau of the Census. 1973
- [15] Pearl, R., & Reed, L. J. On the rate of growth of the population of the United States since 1790 and its mathematical representation. Proceedings of the National Academy of Sciences of the United States of America, 1920; 6(6), 275.

Figure 1

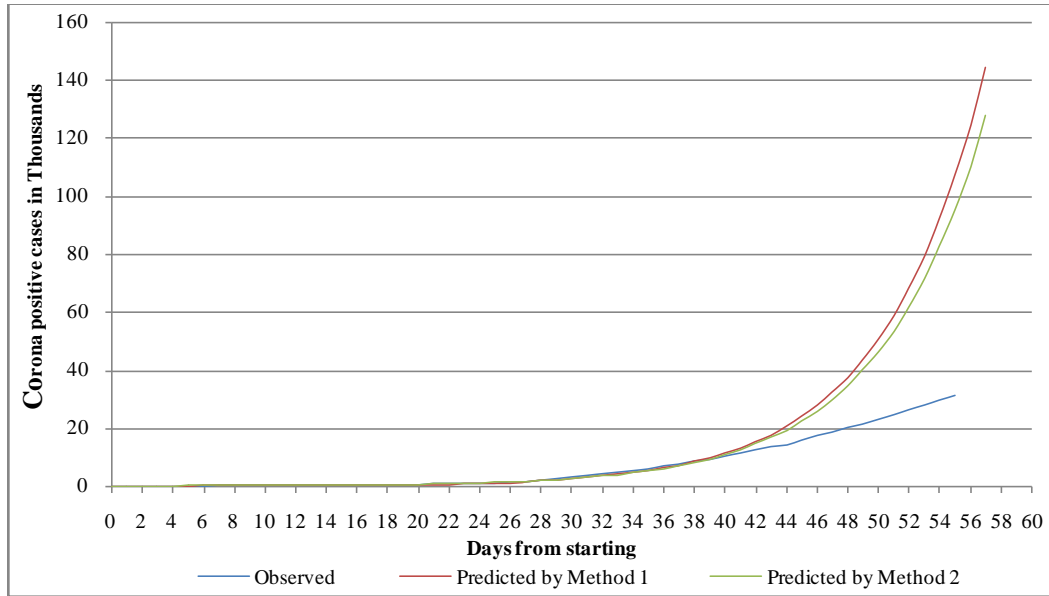


Figure 2

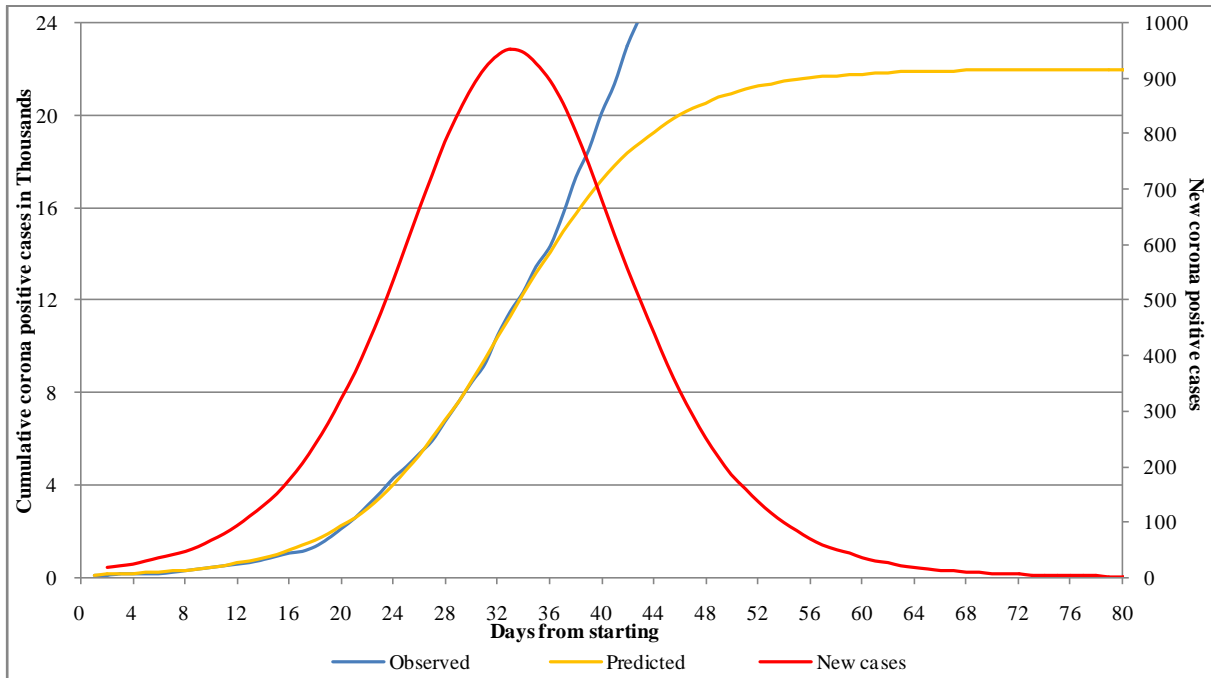
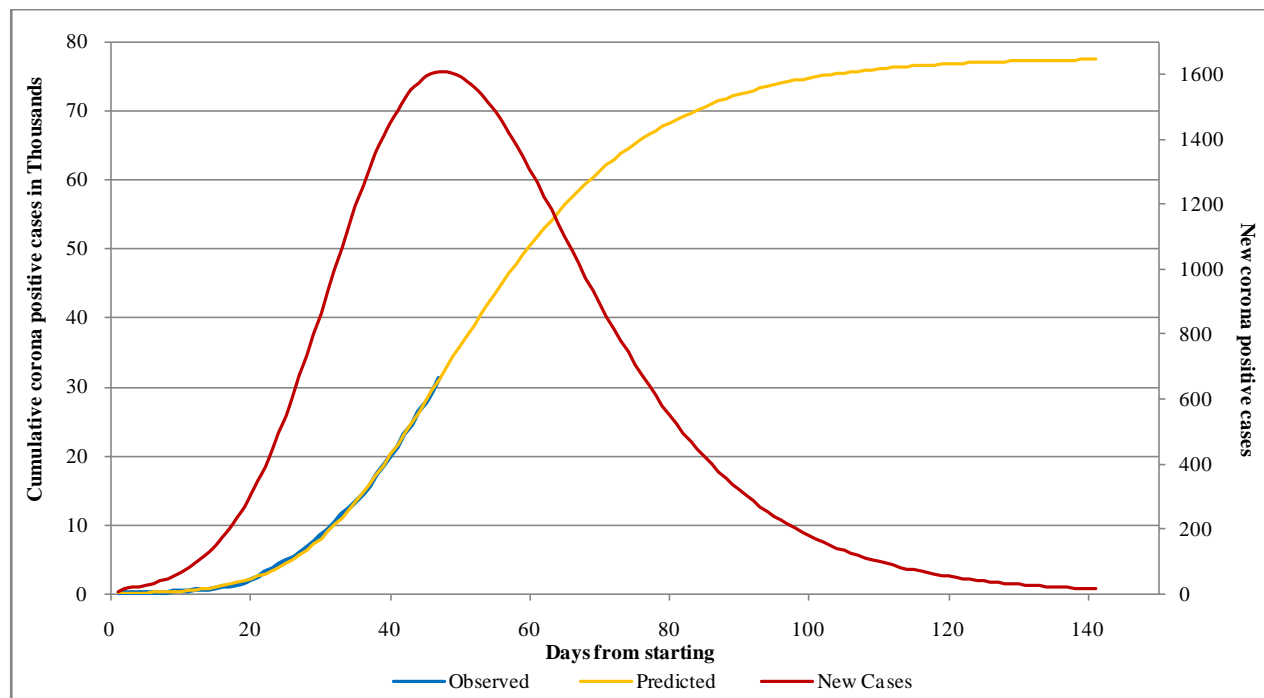


Figure 3



Appendix

Month	Date	Observed	Predicted	95% LCI	95% UCI	New Cases	Residual
March	13	91	91	86	96	9	
	14	107	108	101	114	17	-1
	15	113	127	121	134	20	-14
	16	127	151	144	158	24	-24
	17	146	179	171	187	28	-33
	18	171	213	204	222	34	-42
	19	199	253	243	263	40	-54
	20	258	301	290	312	48	-43
	21	334	358	346	370	57	-24
	22	403	425	411	439	67	-22
	23	505	505	489	521	80	0
	24	571	599	579	619	94	-28
	25	657	709	685	733	110	-52
	26	735	839	809	868	129	-104
	27	886	989	952	1027	151	-103
28	1029	1165	1118	1212	175	-136	
29	1139	1368	1308	1427	203	-229	
30	1347	1602	1527	1676	234	-255	
31	1635	1870	1777	1963	268	-235	
April	1	2059	2176	2061	2291	306	-117
	2	2545	2524	2383	2665	348	21
	3	3105	2917	2746	3089	393	188
	4	3684	3360	3153	3567	442	324
	5	4289	3854	3607	4102	495	435
	6	4778	4404	4111	4698	550	374
	7	5351	5013	4668	5359	609	338
	8	5916	5684	5280	6087	670	232
	9	6729	6417	5950	6884	734	312

	10	7600	7216	6680	7752	799	384
	11	8454	8081	7470	8693	865	373
	12	9212	9013	8321	9705	932	199
	13	10455	10012	9234	10789	999	443
	14	11490	11076	10209	11943	1064	414
	15	12372	12205	11244	13166	1129	167
	16	13434	13396	12339	14454	1191	38
	17	14353	14648	13490	15805	1251	-295
	18	15724	15955	14697	17213	1307	-231
	19	17304	17315	15955	18675	1360	-11
	20	18543	18723	17262	20184	1408	-180
	21	20080	20175	18614	21736	1452	-95
	22	21372	21665	20006	23325	1490	-293
	23	23039	23189	21435	24943	1524	-150
	24	24447	24740	22894	26586	1551	-293
	25	26282	26314	24381	28247	1574	-32
	26	27889	27904	25890	29918	1590	-15
	27	29458	29505	27415	31596	1601	-47
	28	31360	31112	28952	33272	1607	248
	29		32719	30496	34942	1607	
	30		34321	32042	36600	1602	
May	1		35914	33586	38241	1593	
	2		37492	35123	39861	1579	
	3		39053	36650	41456	1560	
	4		40591	38162	43020	1538	
	5		42103	39655	44552	1513	
	6		43587	41127	46048	1484	
	7		45039	42574	47505	1452	
	8		46458	43993	48922	1418	
	9		47840	45384	50297	1382	
	10		49185	46742	51627	1345	
	11		50490	48066	52914	1305	
	12		51755	49356	54154	1265	
	13		52979	50609	55349	1224	
	14		54162	51825	56498	1182	
	15		55302	53003	57602	1141	
June	16		56401	54142	58659	1099	
	17		57457	55242	59672	1057	
	18		58472	56304	60640	1015	
	19		59446	57327	61565	974	
	20		60379	58311	62448	933	
	21		61273	59257	63289	893	
	22		62127	60165	64090	854	
	23		62944	61036	64851	816	
	24		63723	61871	65575	779	
	25		64466	62670	66263	743	
	26		65175	63434	66915	708	
	27		65849	64165	67533	674	
	28		66491	64862	68119	642	
	29		67101	65528	68674	610	
	30		67681	66163	69199	580	
	31		68231	66767	69695	551	
June	1		68754	67343	70164	523	
	2		69250	67892	70608	496	
	3		69720	68413	71026	470	
	4		70165	68909	71421	445	
	5		70587	69380	71794	422	
	6		70987	69828	72146	400	
	7		71365	70253	72477	378	
	8		71723	70656	72789	358	
	9		72061	71038	73084	338	
	10		72381	71401	73361	320	
	11		72683	71745	73621	302	
	12		72969	72071	73867	286	
	13		73239	72379	74098	270	
	14		73493	72672	74315	255	

	15		73734	72949	74520	241	
	16		73961	73210	74712	227	
	17		74175	73458	74893	214	
	18		74378	73693	75062	202	
	19		74568	73915	75222	191	
	20		74748	74124	75372	180	
	21		74918	74323	75513	170	
	22		75078	74510	75645	160	
	23		75228	74687	75770	151	
	24		75370	74854	75886	142	
	25		75504	75013	75996	134	
	26		75630	75162	76099	126	
	27		75749	75303	76195	119	
	28		75861	75436	76286	112	
	29		75967	75562	76371	105	
	30		76066	75681	76451	99	
July	1		76160	75793	76526	94	
	2		76248	75899	76596	88	
	3		76331	75999	76662	83	
	4		76409	76093	76724	78	
	5		76482	76182	76782	74	
	6		76551	76266	76837	69	
	7		76617	76345	76888	65	
	8		76678	76420	76936	61	
	9		76736	76490	76981	58	
	10		76790	76557	77023	54	
	11		76841	76620	77062	51	
	12		76889	76679	77100	48	
	13		76934	76735	77134	45	
	14		76977	76787	77167	43	
	15		77017	76837	77197	40	
	16		77055	76884	77226	38	
	17		77090	76928	77253	35	
	18		77124	76970	77278	33	
	19		77155	77009	77302	31	
	20		77185	77046	77324	30	
	21		77213	77081	77345	28	
	22		77239	77114	77364	26	
	23		77263	77145	77382	25	
	24		77286	77174	77399	23	
	25		77308	77201	77415	22	
	26		77329	77227	77430	20	
	27		77348	77252	77444	19	
	28		77366	77275	77457	18	
	29		77383	77297	77470	17	
	30		77399	77317	77481	16	
	31		77414	77336	77492	15	