2020 IEEE 61st International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)

2020 IEEE 61th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON) | 978-1-7281-9510-0/20/$31.00 ©2020 IEEE | DOI: 10.1109/RTUCON51174.2020.9316598

Using the time varying Kalman filter for prediction of Covid-19 cases in Latvia and Greece

N. Assimakis[1], A. Ktena[1], C. Manasis[1], E. Mele[1], N. Kunicina[2], A. Zabasta[2], T. Juhna[3]

[1] *Energy Systems Laboratory*
National and Kapodistrian University of Athens, Greece
[2] *Institute of Industrial Electronics and Electrical Engineering*
[3] *Water Research and Environmental Biotechnology Laboratory*
Riga Technical University, Latvia

Abstract—**In this work we study applicability of Kalman filters as decision support for early warning and emergency response system for infectious diseases as CoVID-19. Here we use only the actual observations of new cases/deaths from epidemiological survey. We investigated the behavior of various time varying measurement driven models. We implement time varying Kalman filters. Preliminary results from Greece and Latvia showed that Kalman Filters can be used for short term forecasting of CoVID-19 cases. The mean percent absolute error may vary by model; some models give satisfactory results where the mean percent absolute error in new cases is of the order of 2%-5%. The mean absolute error in new deaths is of the order of 1-2 deaths. We propose the use of Kalman Filters for short term forecasting, i.e. next day, which can be a useful tool for improved crisis management at the points of entry to a country or hospitals.**

Keywords—prediction; forecasting; Kalman filters, Covid-19, Internet of Things

I. INTRODUCTION

Due to ongoing outbreak of the CoVID-19 pandemic there is a need for a reliable solutions to assist in the fight against the disease and managing the ensuing crisis in a country's health sector, border control, social well-being, or economy. An important aspect in controlling the spread of infectious diseases is identifying cases, tracing and forecasting of the population flow and the spread in order to be able to locally control it at an early stage.

Well established epidemiological models include variations of the Susceptible – Infectious – Recovered (SIR), such as the Susceptible – Infectious – Recovered – Dead (SIRD) [1] or the most elaborate Susceptible – Exposed – Symptomatic infectious – Asymptomatic infectious – Quarantined – Hospitalized – Recovered cases – Dead (SESAQHRD) model, which mostly focus on the basic reproduction number (R0), and the per day infection mortality and recovery rates [2]. Machine learning techniques have also been proposed in [3], where it is stated that Covid-19 is a virus that cannot be modeled perfectly with any specific traditional model due to the influence of several factors on its dissemination speed: it is difficult to model the behavior of Covid-19 because of the non-linearity of infection data caused by under-notifications and also the lack of effective and constant counter measures, which changes all the time as the infection spreads. A moving average (MA) filter of size 7 was used in order to forecast new daily cases.

The motivation of this work has been the development of a tool that will allow better preparation and management at the administration level and an early warning and emergency response based on ubiquitous data, such as the smart city or smart building data, historical statistical data and data collected at testing points (hospitals, borders, etc). The goal is to identify potential spreaders and epicenters and predict the evolution of the disease in order to contain it as well as efficiently protect the population and manage resources. To this end, the data collected will be modeled and used for forecasting, using a combination of tools. The modeling and simulation results will assist in designing effective measures and assessing their impact.

The technology required to develop such a platform, collect data, model them and offer decision support already exists in a variety of forms. An exhaustive literature review on the topic is beyond the scope of this work, however decision support platforms with IoT powered big data analytics and forecasting tools have already been designed to address other needs [4-9] and in this work we propose that they are redesigned to address the CoVID-19 related objective.

In this preliminary work we will present the basic features of the design of the platform and focus on the functionality of short-term forecasting, i.e.one day-ahead, of expected cases/deaths using Kalman Filtering. The goal is to use the results in order to design action scenarios and anticipate the logistics involved in dealing with the management of the pandemic at the points of control such as borders or hospitals. A safe estimate of the cases to be anticipated in the following day or days may prove very helpful for the local authorities and the health system management in order to effectively protect the health and administration workers and the population. The results presented are based on the publicly available data for Greece and Latvia, two small EU countries with very different climate and population characteristics.

In section II we present the basic functionalities of a system consisting of IoT sensors in combination with a decision support platform with communication and interface modules. In sections III and IV we present the models used with the Kalman Filter (KF) in order to predict Covid-19 new daily cases and the simulation results. The conclusions are summarized in section VI.

II. DECISION SUPPORT SYSTEM

The combination of Internet, objects and mobile services has the potential to offer diffused intelligence and access to services, resource monitoring, interactivity. The Internet of Things (IoT) technologies enable the interconnection of objects that vary in terms of capabilities and characteristics while wireless sensors and sensor networks combined with big data analysis tools enable the extraction of useful information from them. Objects connected in IoT make use of the Internet Protocol Version 6 (IPv6) both because it offers unique network identifiers (due to its abundance of addresses) and because it can provide seamless connectivity with other Internet devices.

On the other hand, intelligent Decision Support Systems (i-DSS), equipped with artificial intelligence (AI) algorithm play an increasingly important role in the automated monitoring, control and operation of various systems [8] and enable the modeling, forecasting and optimization of processes and systems. Such an i-DSS has been developed [4] as a modular and scalable platform to make hourly forecasts at an electrical microgrid level, using and combining various types of historical and real-time data. Its architecture enables implementation, testing, combination and comparison of a number of forecasting and clustering algorithms in order to obtain the optimum forecast for the problem studied. The platform architecture consists of several interconnected well-defined components where each one interacts directly with the other. The database is designed based on the types of data to be collected and allows storing data dynamically. The platform hosts clustering algorithms, such as Self Organizing Maps, and forecasting tools, such as ANNs or KF as well as multi-optimization routines for decision support.

In the following section, we present the models used to formulate the KF, based on historical data of new cases/deaths for Greece and Latvia. The results may be correlated with weather or other sensor, patient-based or user generated data.

III. MODELS AND ALGORITHMS

The general discrete time linear model used to formulate the Kalman Filter (KF) consists of the dynamic and the statistical model. The dynamic model expresses the relationship between state and the measurement and is described by the following state space equations:

$$x(k+1)=F(k+1,k)\,x(k)+w(k) \tag{1}$$

$$z(k+1)=H(k)\,x(k)+v(k) \tag{2}$$

where $x(k)$ is the $n \times 1$ state vector, $z(k)$ is the $m \times 1$ measurement vector, $F(k+1, k)$ is the $n \times n$ transition matrix, $H(k)$ is the $m \times n$ output matrix, $w(k)$ is the $n \times 1$ state noise and $v(k)$ is the $m \times 1$ measurement noise at time k. The statistical model expresses the nature of state and measurements. In this work, the basic underlying assumption is that the state noise, $w(k)$, and the measurement noise, $v(k)$, follow the white noise distribution: $\{w(k)\}$ is a zero mean Gaussian process with known covariance $Q(k)$ of dimension $n \times n$ and $\{v(k)\}$ is a zero mean Gaussian process with known covariance $R(k)$ of dimension $m \times m$.

The following assumptions also hold: (a) the initial value of the state $x(0)$ is a Gaussian random variable with mean $x0$ and covariance $P0$, (b) the noise stochastic processes and the random variable $x(0)$ are independent.

A. Measurements driven Scalar Model (MSM)

Aiming to predict the Covid-19 new cases/deaths, a basic category of models may be developed: the models that take into account only the real observations of new cases/deaths, namely the Measurement driven Scalar Models (MSM) which are presented below.

In the MSM, the measurement vector has one element, namely the observation, which corresponds to new cases or new deaths observed. The state vector has also one element, namely the new cases or new deaths. The state changes in each time step following the ratio of two successive observations:

$$F(k+1,k) = z(k+1)/z(k) \tag{3}$$

The computation of $F(k+1,k) = z(k+1)/z(k)$ requires $z(k) \neq 0$. In the case where $z(k) = 0$, we can use $F(k+1,k) = 1$ or $F(k+1,k) = 0$, or we can use the last nonzero measurement instead of $z(k) = 0$.

A variation of this model can be derived replacing each measurement by the average value obtained from a window of measurements from the beginning till time k:

$$F(k+1,k) = \frac{mean(z(1),\ldots,z(k+1))}{mean(z(1),\ldots,z(k))} \tag{4}$$

Another variation results assuming that the state does not change from step to step, so

$$F(k+1,k) = F = 1 \tag{5}$$

The noisy measurement is

$$H(k) = H = 1 \tag{6}$$

and the linear model takes the form:

$$x(k+1)=F(k+1,k)\,x(k)+w(k) \tag{7}$$

$$z(k+1)= x(k)+v(k) \tag{8}$$

The state and measurement noise variances $Q(k)$ and $R(k)$ are time varying and concern a time period before the prediction time. In fact, $Q(k)$ is the variance of the difference of two succeeded observations $z(k+1)–z(k)$ and $R(k)$ is the variance of the measurements $z(k)$ for a time period before the prediction time.

The noise variances $Q(k)$ and $R(k)$ are time varying and can be computed on line for a fixed backwards time interval before the prediction time, denoted by b (backwards). For example they can be determined from the last 2 weeks (quarantine period). The noise variances $Q(k)$ and $R(k)$ can be computed on line for the time interval from the beginning till the prediction time. This means that $b=k$ is not constant.

The model is clearly time varying.

B. *Prediction Algorithms*

The discrete time Kalman filter [9] is the most well-known algorithm that solves the filtering problem, producing the state estimation $x(k/k)$ and the corresponding estimation covariance matrix $P(k/k)$ as well as the state prediction $x(k+1/k)$ and the corresponding prediction covariance matrix $P(k+1/k)$. The prediction horizon depends on available data. In this paper, it is assumed to be one (1) day.

All the Time Varying Kalman Filters (TVKF) with their parameters derived for MSM are presented in Table I.

The prediction derived by Kalman Filter was corrected through using ceiling function (pessimistic scenario). Simulation scripts were run using Matlab / Octave.

TABLE I. KALMAN FILTERS FOR MSM

Kalman Filter	**F**	**H**	**Q, R**
TVKF-1	if z(k)=0 then F(k+1,k)=1 else F(k+1,k)= $\frac{z(k+1)}{z(k)}$	H=1	total period
TVKF-2	if z(k)=0 then F(k+1,k)=0 else F(k+1,k)= $\frac{z(k+1)}{z(k)}$	H=1	total period
TVKF-3	if z(k)=0 then if z(1)=…=z(k)=0 then F(k+1,k)=0 else F(k+1,k)= $\frac{z(k+1)}{\text{last measurement} \neq 0}$ else F(k+1,k)= $\frac{z(k+1)}{z(k)}$	H=1	total period
TVKF-4	if mean{z(1)…z(k)}=0 then F(k+1,k)=1 else F(k+1,k)= $\frac{\text{mean}\{z(1)\dots z(k+1)\}}{\text{mean}\{z(1)\dots z(k)\}}$	H=1	total period
TVKF-5	if z(k)=0 then F(k+1,k)=1 else F(k+1,k)= $\frac{z(k+1)}{z(k)}$	H=1	quarantine period
TVKF-6	if z(k)=0 then F(k+1,k)=0 else F(k+1,k)= $\frac{z(k+1)}{z(k)}$	H=1	quarantine period
TVKF-7	if z(k)=0 then F(k+1,k)= $\frac{z(k+1)}{\text{last measurement} \neq 0}$ else F(k+1,k)= $\frac{z(k+1)}{z(k)}$	H=1	quarantine period
TVKF-8	if mean{z(1)…z(k)}=0 then F(k+1,k)=1 else F(k+1,k)= $\frac{\text{mean}\{z(1)\dots z(k+1)\}}{\text{mean}\{z(1)\dots z(k)\}}$	H=1	quarantine period
TVKF-9	F=1	H=1	total period
TVKF-10	F=1	H=1	quarantine period

C. *Simulation Results*

For the MSM, the Time Varying Kalman Filter (TVKF) has been implemented.

The noise variances Q(k) and R(k) were a) computed on line for the time interval from the beginning till the prediction time (total period) b) were determined from the last 2 weeks (quarantine period).

Zero initial conditions were assumed for all Kalman Filters. The data used with MSM concern the new cases/deaths in a) Greece for the time interval from Feb 26, onset of the pandemic, till June 14 [10] and b) Latvia for the time interval from Mar 3 till Jul 13 [11].

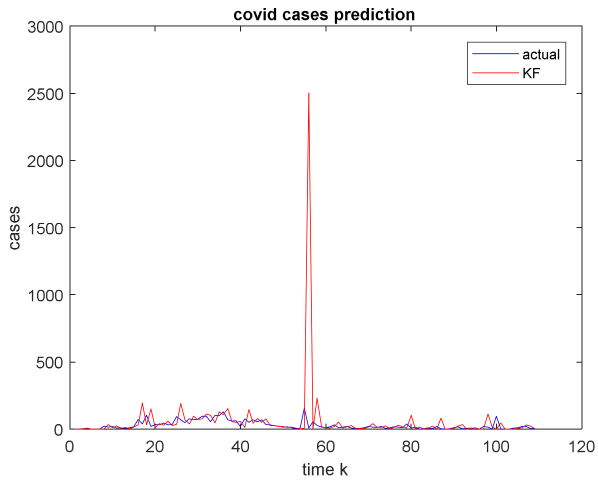

Fig. 1. Greece new cases prediction with TVKF-1.

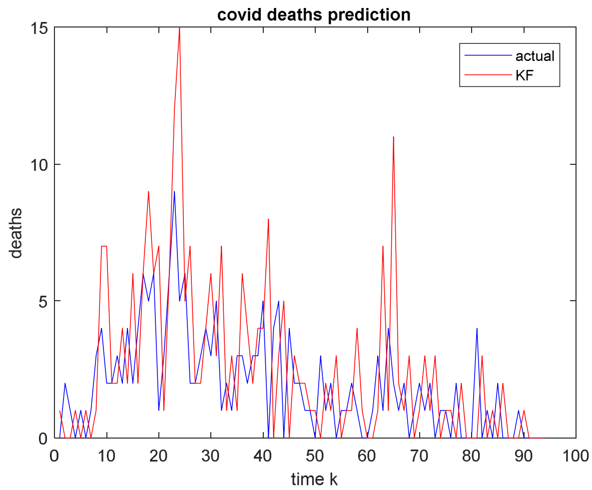

Fig. 2. Greece new deaths prediction with TVKF-1

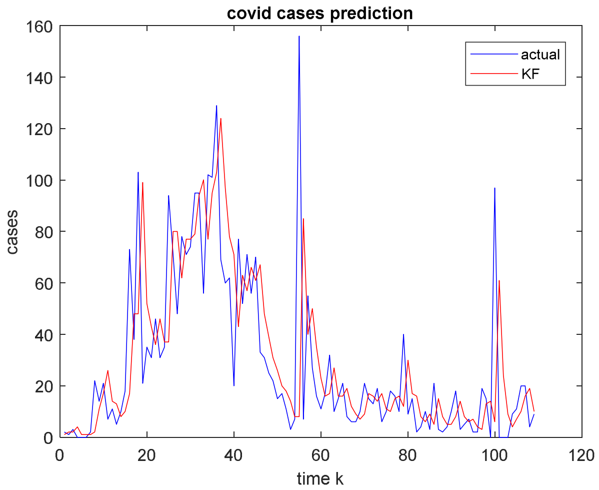

Fig. 3. Greece new cases prediction with TVKF-4

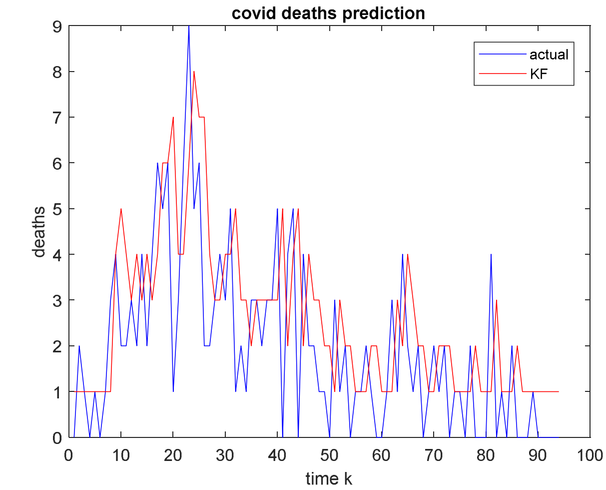

Fig. 4. Greece new deaths prediction with TVKF-4

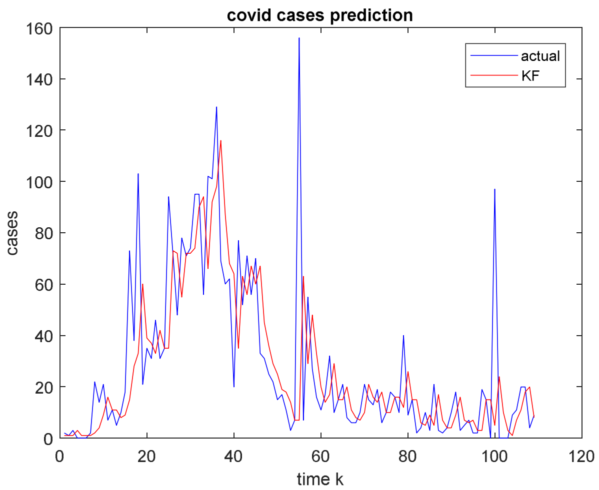

Fig.5 Greece new cases prediction with TVKF-10

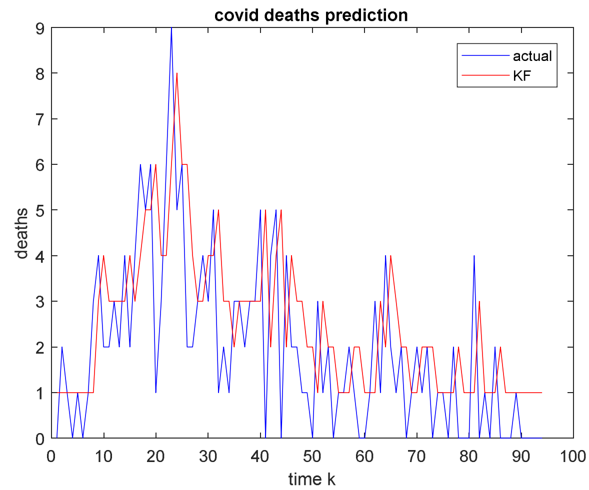

Fig.6.Greece new deaths prediction with TVKF-10

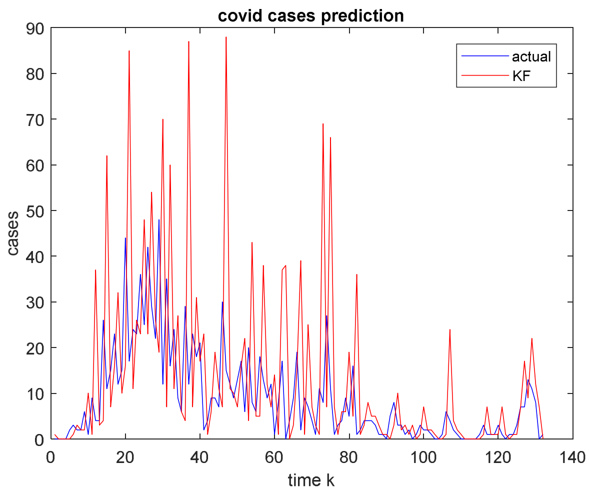

Fig.7 Latvia new cases prediction with TVKF-1

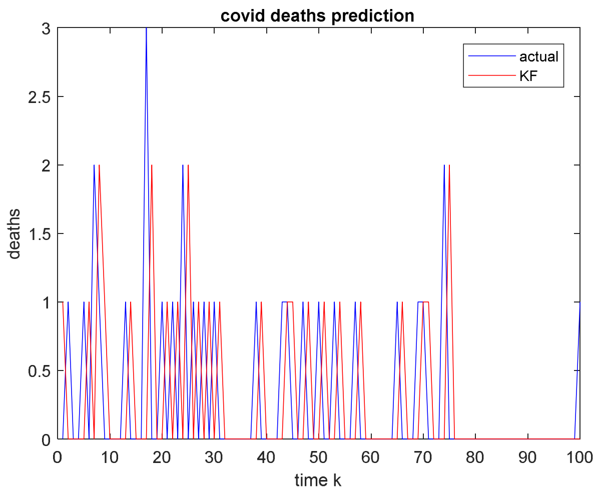

Fig.8 Latvia new deaths prediction with TVKF-1

The difference between two successive measurements affects the prediction: as the difference increases, the prediction becomes worse (see peak).

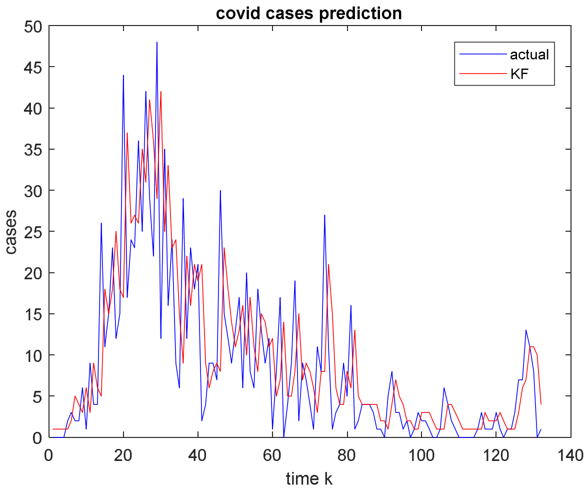

Fig.9 Latvia new cases prediction with TVKF-4

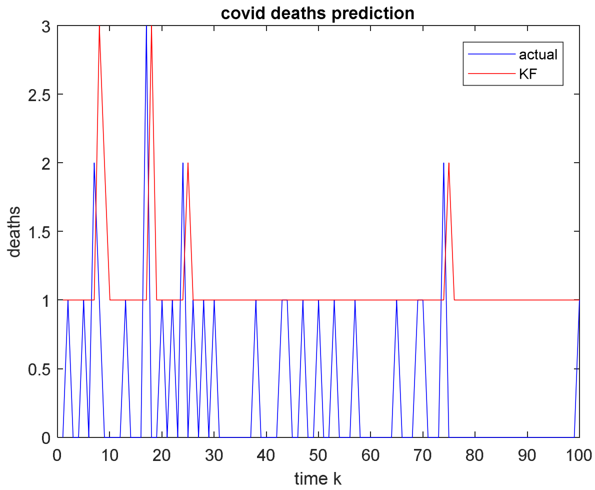

Fig.10 Latvia new deaths prediction with TVKF-4

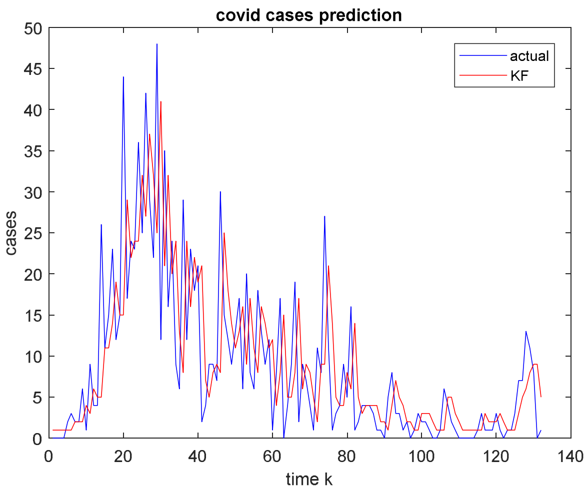

Fig.11Latvia new cases prediction with TVKF-10

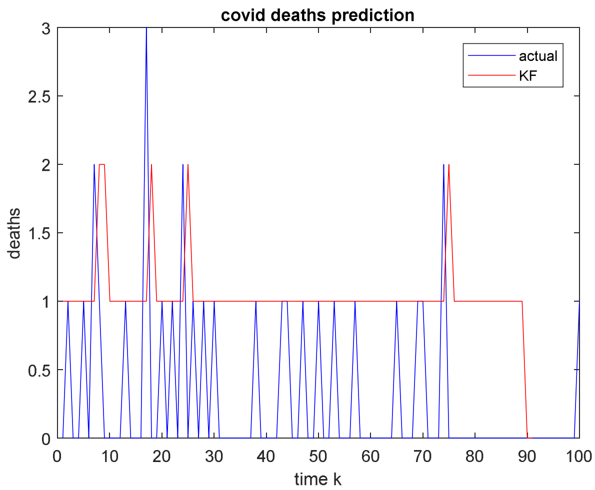

Fig.12Latvia new deaths prediction with TVKF-10

Table II depicts the mean percent absolute error and the mean absolute error in Greece new cases prediction using Kalman Filter.

TABLE II. ERROR IN GREECE NEW CASES PREDICTION USING KF

Kalman Filter	model	mean percent abs error	mean abs error
TVKF-1	MSM	100.8618	48.0734
TVKF-2	MSM	98.0530	47.5780
TVKF-3	MSM	107.8519	50.3211
TVKF-4	MSM	5.8091	15.8716
TVKF-5	MSM	93.8717	46.1376
TVKF-6	MSM	92.0843	45.9358
TVKF-7	MSM	95.8506	46.9633
TVKF-8	MSM	1.3406	15.1927
TVKF-9	MSM	1.8832	15.3119
TVKF-10	MSM	5.8410	14.5413

The main result is that the performance of Kalman Filters may vary a lot: the mean percent absolute error may be of the order of 100% or of the order of 2%.

Table III depicts the mean percent absolute error and the mean absolute error in Greece new deaths prediction using Kalman Filter.

TABLE III. ERROR IN GREECE NEW DEATHS PREDICTION USING KF

Kalman Filter	model	mean percent abs error	mean abs error
TVKF-1	MSM	32.9670	1.9149
TVKF-2	MSM	15.3846	1.8298
TVKF-3	MSM	35.1648	2.0000
TVKF-4	MSM	32.9670	1.3830
TVKF-5	MSM	39.0110	2.0319
TVKF-6	MSM	21.4286	1.9468
TVKF-7	MSM	41.7582	2.0851
TVKF-8	MSM	33.5165	1.3936
TVKF-9	MSM	26.9231	1.3298
TVKF-10	MSM	28.0220	1.3298

The main result is that the mean absolute error in new deaths prediction using TVKF is very low. In fact it is of the order of 1-2 deaths. This leads to the conclusion that TVKF is a reliable tool to predict the new deaths.

Table IV depicts the mean percent absolute error and the mean absolute error in Latvia new cases prediction using Kalman Filter.

TABLE IV. ERROR IN LATVIA NEW CASES PREDICTION USING KF

Kalman Filter	model	mean percent abs error	mean abs error
TVKF-1	MSM	51.6624	11.1061
TVKF-2	MSM	48.9344	11.0909
TVKF-3	MSM	52.2592	11.1439
TVKF-4	MSM	10.2302	5.4545
TVKF-5	MSM	64.1091	12.2121
TVKF-6	MSM	61.6368	12.2045
TVKF-7	MSM	65.0469	12.2803
TVKF-8	MSM	9.1219	5.6894
TVKF-9	MSM	3.4101	5.1667
TVKF-10	MSM	2.5575	5.3939

The main result is that the performance of Kalman Filters may vary: the mean percent absolute error may be of the order of 5%-10%.

Table V depicts the mean percent absolute error and the mean absolute error in Latvia new deaths prediction using Kalman Filter.

TABLE V. ERROR IN LATVIA NEW DEATHS PREDICTION USING KF

Kalman Filter	model	mean percent abs error	mean abs error
TVKF-1	MSM	6.8966	0.5000
TVKF-2	MSM	89.6552	0.3200
TVKF-3	MSM	10.3448	0.5500
TVKF-4	MSM	268.9655	0.8800
TVKF-5	MSM	6.8966	0.5000
TVKF-6	MSM	89.6552	0.3200
TVKF-7	MSM	6.8966	0.5400
TVKF-8	MSM	268.9655	0.8800
TVKF-9	MSM	262.0690	0.8600
TVKF-10	MSM	262.0690	0.8600

The main result is that the mean absolute error in new deaths prediction using TVKF is very low. In fact it is of the order of 1 death. This leads to the conclusion that TVKF is a reliable tool to predict the new deaths.

Overall, Kalman Filter may offer a reliable day-ahead forecast when no big differences occur from day to day. This can lead to a very useful tool in day-today management of cases and deaths such as hospitals or points of entry to a controlled region, e.g. border checkpoints.

IV. CONCLUSIONS

We propose the design of a Decision Support System which will collect and process historical, IoT and user generated data, and execute forecasting and optimization algorithms to assist with CoVID-19 related emergencies and crises.

In this work we have focused on the performance of Kalman Filtering for forecasting the new cases/deaths of Covid-19 based on the Measurement driven Scalar Model. Several KF algorithms have been tested with data from Greece and Latvia.

The results show that KF may assist in short-term decision making as the logistics involved in resources management in hospitals and other administrative units involved in crisis management. The models perform better when the variance of the data is not too high, and especially when the difference between two successive days is not abnormally high.

The choice of the appropriate filter depends on the desired speed in calculations and the assumptions made. In any case, the Decision Support platform is designed in such a way as to allow comparison of several models and algorithms and select the optimum performance.

Future work involves the implementation of Kalman Filters for prediction of Covid-19 cases/deaths for local areas (e.g. one city) and modeling and simulation using statistical data as well to improve the prediction. Moreover it can be used for more better design of epidemiological studies.

REFERENCES

[1] Anastassopoulou C, Russo L, Tsakris A, Siettos C (2020) Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS ONE 15(3): e0230405. https://doi.org/10.1371/journal.pone.0230405

[2] Khondoker NazmoonNabi, Forecasting COVID-19 pandemic: A data-driven analysis, Chaos, Solitons & Fractals, Volume 139, October 2020, 110046, https://doi.org/10.1016/j.chaos.2020.110046

[3] Pereira, I.G.; Guerin, J.M.; Silva Júnior, A.G.; Garcia, G.S.; Piscitelli, P.; Miani, A.; Distante, C.; Gonçalves, L.M.G. Forecasting Covid-19 Dynamics in Brazil: A Data Driven Approach. Int. J. Environ. Res. Public Health 2020, 17, 5115.

[4] E Mele, C Elias, A Ktena, Machine Learning Platform for Profiling and Forecasting at Microgrid Level, Electrical, Control and Communication Engineering 15 (1), 21-29

[5] E Mele, I Tzanis-Kontomichalos, L Sarakis, A Ktena, A Secure, Open Source WSN-based Monitoring System using CM5000 Motes, 2019 8th Mediterranean Conference on Embedded Computing (MECO), 1-4

[6] N Assimakis, C Manasis, A Ktena, Electric load estimation using Kalman and Lainiotis filters, 8th Mediterranean Conference on Embedded Computing (MECO), 1-4, 2019

[7] V Vikias, C Manasis, A Ktena, N Assimakis, Forecasting for Temperature Dependent Power Generation Using Kalman Filtering, IEEE 60th International Scientific Conference on Power and Electrical R.E. Kalman, "A new approach to linear filtering and prediction problems," J. Bas. Eng., Trans. ASME Ser D, vol. 8, no 1, pp. 34-45, 1960.

[8] V. Girones, S. Moret, F. Marechal and D. Favrat, "Strategic energy planning for large-scale energy systems: A modelling framework to aid decision-making," *Energy 2015,* vol. 90, no. 1, pp. 173-186, 2015

[9] B.D.O. Anderson and J.B. Moore, "Optimal filtering," New York: Dover Publications, 2005.

[10] Zhengguo Sheng; Chinmaya Mahapatra; Chunsheng Zhu; Victor C. M. Leung, "Recent Advances in Industrial Wireless Sensor Networks Toward Efficient Management in IoT," IEEE Access, vol. 3, pp. 622-637, 2015.

[11] Daily Overview, Information about Covid-19 in Greece, https://covid19.gov.gr/covid19-live-analytics/. [Accessed on July 1].

[12] Information about Covid-19 in Latvia, Center for Disease Prevention and Control, Latvian Open Data Portal, https://data.gov.lv/dati/lv/dataset/covid-19/resource/d499d2f0-b1ea-4ba2-9600-2c701b03bd4a. [Accessed on July 15].

Nicholas Assimakis received the Diploma and Ph.D. degree from the Department of Computer Engineering and Informatics, University of Patras, Greece. He is currently a Professor with the General Department of the National and Kapodistrian University of Athens, Greece. His research interests are in digital

signal processing, estimation theory, filtering, algorithm design and development.