

Recovering Method of Missing Data Based on Proposed Modified Kalman Filter When Time Series of Mean Data is Known

Kohei Arai 1

Graduate School of Science and Engineering
Saga University
Saga City, Japan

Abstract—Recovering method of missing data based on the proposed modified Kalman filter for the case that the time series of mean data is known is proposed. There are some cases of which although a portion of data is missing, mean value of the time series of data is known. For instance, although coarse resolution of imagery data are acquired every day, fine resolution of imagery data are missing sometimes. In other words, coarse resolution of imaging sensor has wide swath width while fine resolution of imaging sensor has narrow swath, in general. Therefore, coarse resolution of sensor data can be acquired every day while fine resolution of sensor data can be acquired not so frequently. It would be nice to become able to create frequently acquired fine resolution of sensor data (every day) using the previously acquired fine resolution of sensor data together with the coarse resolution of sensor data. The proposed method allows creation of fine resolution sensor data with the aforementioned method based on a modified Kalman filter. As an example of the proposed method, prediction of missing ASTER/VNIR data based on Kalman filter using simultaneously acquired MODIS data as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources. Experimental data shows that 4 to 111% of prediction error reduction can be achieved by the proposed modified Kalman filter in comparison to the conventional Kalman filter. It is found that the reduction rate depends on the mean value accuracy of time series data derived from the other data sources. The experimental results with remote sensing satellite imagery data show a validity of the proposed method

Keywords—Kalman filter; nremote sensing satellite image; time series analysis

I. INTRODUCTION

There are some cases of which although a portion of data is missing, mean value of the time series of data in concern is known. For instance, coarse resolution of imaging sensor has wide swath width while fine resolution of imaging sensor has narrow swath, in general. Therefore, coarse resolution of sensor data can be acquired every day while fine resolution of sensor data can be acquired not so frequently. It would be nice to become able to create frequently acquired fine resolution of sensor data (every day) using the previously acquired fine resolution of sensor data together with the coarse resolution of

sensor data. The proposed method allows creation of fine resolution sensor data with the aforementioned method based on a modified Kalman filter.

Kalman filter is widely used for prediction of missing data [1],[2]. The parameters required for Kalman filter can be basically estimated with the Recursive Least Squares: RLS method. Applicability of the RLS method is discussed [3] and the method for overcoming the problem on the limitation of applicability of the RLS method is also discussed [4]-[6].

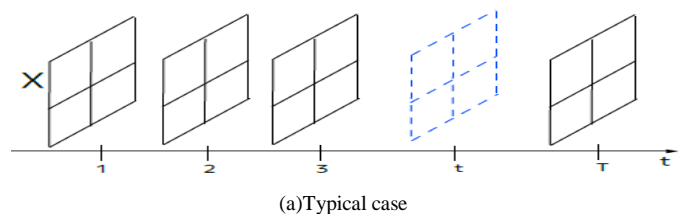
The experimental results with remote sensing satellite imagery data show a validity of the proposed method. As an example of the proposed method, prediction of missing ASTER/VNIR data [7] based on Kalman filter using simultaneously acquired MODIS data [8] as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources.

II. PROPOSED METHOD

A. Kalman Filter

B. Problem Statement

Prediction error of the conventional Kalman filter is not so small in particular for the cases that there are a long time periods of contiguous missing data. There are some cases that there is time series of mean data of the original time series of data which includes missing data as shown in Figure 1. Not only time series of data X, but also its mean of time series data is also available to use for recovering the missing data. It may be possible to improve prediction accuracy by using mean of time series of data in addition to the available data of the original time series of data.



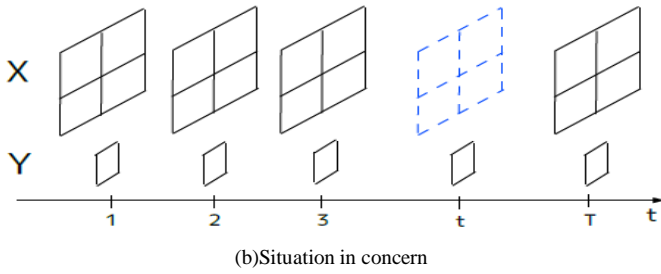


Fig. 1 Prediction of missing data in time series of data, X_t with known time series of data, X_i in the conventional Kalman filter and in the proposed modified Kalman filter.

C. Proposed Modified Kalman Filter

Based on the well known Kalman filter, the missing data can be predicted in accordance with the flow chart of Figure 2. Meanwhile, the flow chart of the proposed modified Kalman filter is shown in Figure 3. Prediction is done by using available data during updating process for estimation of coefficients of the Kalman filter while no update is done when the data is missing for the conventional Kalman filter. Meanwhile, update is done by using mean of the data in concern even for the missing data.

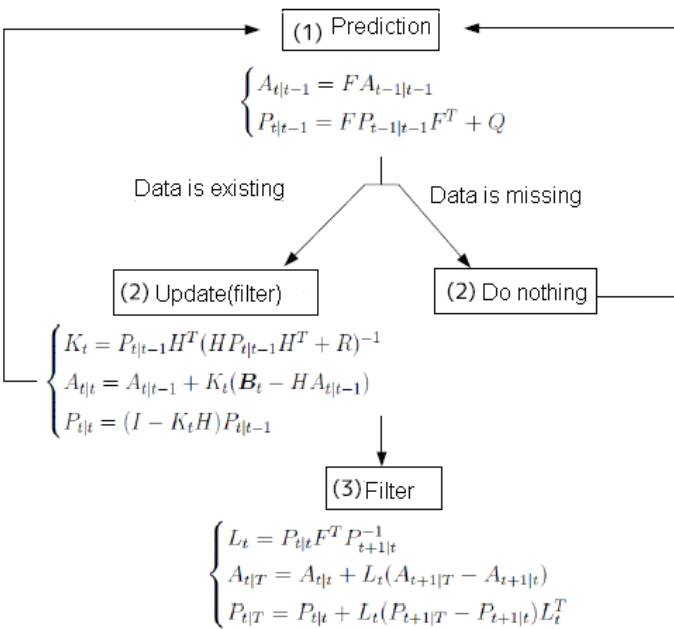


Fig. 2 Prediction of missing data based on the conventional Kalman filter

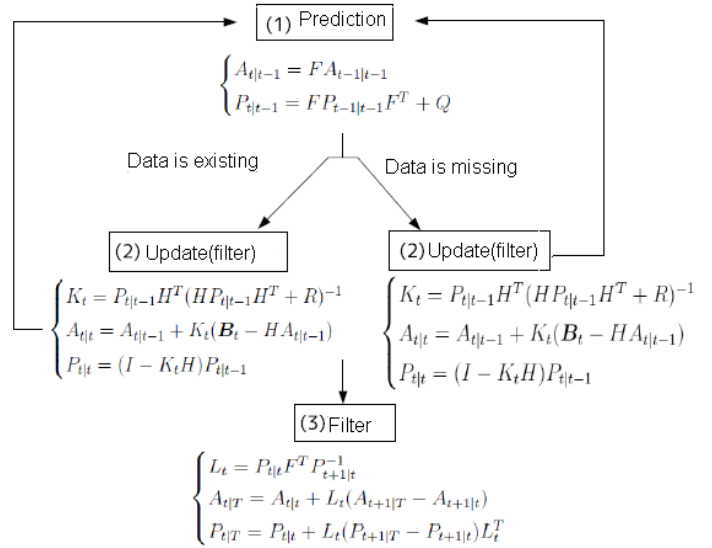


Fig. 3 Process flow of prediction of missing data based on the proposed modified Kalman filter

Where the time series of data is followed by the multi variant auto-regressive model [9] as shown in equation (1)

$$X_t = \sum_{j=1}^p \Phi_j X_{t-j} + N_t \quad \{N_t\} \sim WN(0, \Sigma) \quad (1)$$

where Φ can be estimated with RLS method while Φ_1, \dots, Φ_p is regressive coefficients for X_p . On the other hand, p denotes is order while N denotes residual and is followed by multi variant normal distribution with zero mean and covariance matrix of Σ [10]. These can be rewritten with equation (2).

$$X_t = \begin{bmatrix} x_t(1) \\ \vdots \\ x_t(m) \end{bmatrix}, \Phi_j = \begin{bmatrix} \phi_j(11) & \cdots & \phi_j(1m) \\ \vdots & \ddots & \vdots \\ \phi_j(m1) & \cdots & \phi_j(mm) \end{bmatrix}, N_t = \begin{bmatrix} n_t(1) \\ \vdots \\ n_t(m) \end{bmatrix} \quad (2)$$

Multi variant auto-regressive model of the parameters are estimated as follows,

- 1) Covariance matrix is estimated at first,
- 2) The following Darwin-Levinson algorithm [11] is then used introducing the following conditional mean and covariance,

$$\begin{cases} A_{t|j} \equiv E(A_t|A_j) \\ P_{t|j} \equiv E[(A_t - A_{t|j})(A_t - A_{t|j})^T] \end{cases} \quad (3)$$

A_t is estimated in accordance with equation (4).

$$\begin{array}{l} \text{Prediction} \\ \text{Update} \\ \text{(Filter)} \end{array} \begin{cases} A_{t|t-1} = F_t A_{t-1|t-1} \\ P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t \\ K_t = P_{t|t-1} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} \\ A_{t|t} = A_{t|t-1} + K_t (B_t - H_t A_{t|t-1}) \\ P_{t|t} = (I - K_t H_t) P_{t|t-1} \end{cases} \quad (4)$$

Then the following smoothing filter is applied,

$$\begin{cases} L_t = P_{t|t} F_{t+1}^T P_{t+1|t}^{-1} \\ A_{t|T} = A_{t|t} + L_t (A_{t+1|T} - A_{t+1|t}) \\ P_{t|T} = P_{t|t} + L_t (P_{t+1|T} - P_{t+1|t}) L_t^T \end{cases} \quad (5)$$

State equation of X_n can be written as follows,

$$\begin{aligned} A_{t+1} &= F_t A_t + G_t V_t, \quad \{V_t\} \sim WN(0, \{Q_t\}) \\ B_t &= H_t A_t + W_t, \quad \{W_t\} \sim WN(0, \{R_t\}) \end{aligned} \quad (6)$$

The former equation is called as state equation while the later equations is called as observation equation and are rewritten as follows,

$$A_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}, B_t = X_t \quad (17)$$

$$\begin{bmatrix} X_{t+1} \\ X_t \\ \vdots \\ X_{t-p+3} \\ X_{t-p+2} \end{bmatrix} = \begin{bmatrix} \Phi_p & \Phi_{p-1} & \cdots & \Phi_2 & \Phi_1 \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+2} \\ X_{t-p+1} \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} V_t \quad (7)$$

$$B_t = \begin{bmatrix} I & 0 & \cdots & 0 & 0 \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+2} \\ X_{t-p+1} \end{bmatrix} + W_t \quad (8)$$

Thus equation (1) is represented with the above equation (7) and (8) which is called as state space model. State space model can be updated when the data is available while it cannot be updated when the data is missing.

Prediction of the missing data can be done in accordance with Figure 2 based on the conventional Kalman filter. Meanwhile, B_t can be estimated based on the proposed method assuming $B_t = Y_t$ with equation (9),

$$A_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{bmatrix}, B_t = Y_t \quad (9)$$

then

$$\begin{bmatrix} X_{t+1} \\ X_t \\ \vdots \\ X_{t-p+3} \\ X_{t-p+2} \end{bmatrix} = \begin{bmatrix} \Phi_p & \Phi_{p-1} & \cdots & \Phi_2 & \Phi_1 \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+2} \\ X_{t-p+1} \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} V_t \quad (10)$$

$$B_t = \begin{bmatrix} scalar & 0 & \cdots & 0 & 0 \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{bmatrix} \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+2} \\ X_{t-p+1} \end{bmatrix} + W_t \quad (11)$$

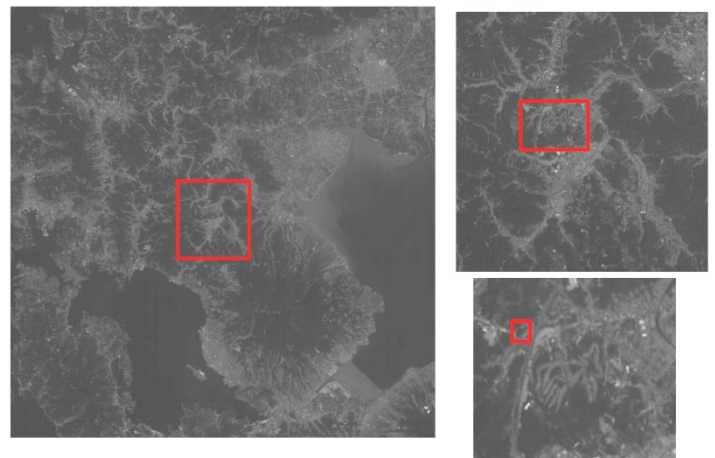
Thus state space model can be updated even for the data is missing. Namely, updating is skipped for the conventional Kalman filter while updating is performed even for the data is missing for the proposed method. Smoothing filter is common for both the conventional and the proposed Kalman filter [12]. In other word, time series of mean data is used as time series of B , mean of A is used as coefficients of H_t of the observation equation. The *scalar* in the equation (11) is averaged vector.

III. EXPERIMENT

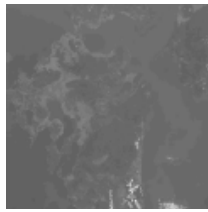
A. Preliminary Simulation

As an example of the proposed method, prediction of missing ASTER/VNIR data based on Kalman filter using simultaneously acquired MODIS data as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources.

An example of the ASTER/VNIR band 2 image used (Red square on the right bottom image shows 16x16 pixels of ASTER/VNIR band 2 which corresponds to 1 pixel of MODIS band 1) is shown in Figure 4.



(a) ASTER/VNIR band 2



(b)Corresponding MODIS band 1

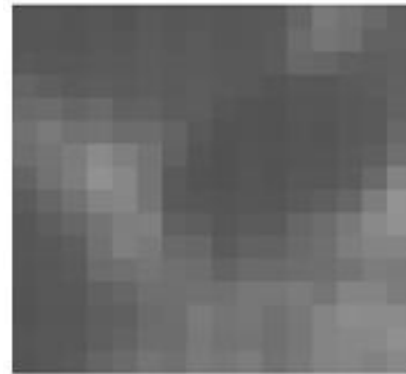
Fig. 4 An example of the ASTER/VNIR band 2 image used (Red square on the right bottom image shows 16x16 pixels of ASTER/VNIR band 2 which corresponds to 1 pixel of MODIS band 1).

Although these ASTER and MODIS sensors are onboard same satellite, Terra, repeat cycle is different each other. Revisit cycle of the ASTER/VNIR is 16 days while that of the MODIS is almost one day due to the fact that swath width of MODIS is wide enough (2000 km) to achieve one day coverage for entire globe. There are missing data when both sensors acquire cloudy and rainy scenes because these are visible to thermal infrared radiometers. Spectral characteristics of ASTER/VNIR band 2 and MODIS band 1 are almost similar. Instantaneous Field of View: IFOV of ASTER/VNIR is 15 m while that of MODIS is 1 km. Therefore, approximately 16 by 16 pixels of ASTER/VNIR are corresponding to one MODIS pixel. Relatively clear scenes of Ureshino city in Kyushu island in Saga prefecture, Japan are acquired with ASTER/VNIR on December 15 2004, September 29 2005, April 12 2007, May 14 2007, August 18 2007, January 9 2008, April 14 2008, May 3 2008, and May 19 2008 during the period from December 2004 and May 2008. Although ASTER/VNIR acquires imagery data every 16 days in the period, just 10 clear scenes are acquired. By using 10 of the acquired ASTER/VNIR band 2 of imagery data, 8 of the imagery data can be predicted based on Kalman Filter: KF without the corresponding MODIS band 1 data. Also 8 of the imagery data can be predicted based on the proposed modified KF with the corresponding MODIS band 1 data. Figure 5 shows example of small portion of the predicted VNIR band 2 images based on the conventional Kalman filter (KF) and the proposed Kalman filter with MODIS band 1 as a mean value.

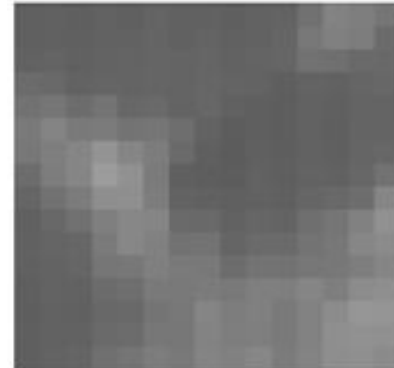
Observation equation in equation (6) requires appropriate white noise with zero mean and appropriate standard deviation. In order to determine appropriate standard deviation, prediction error is evaluated with the different standard deviations then an appropriate standard deviation is determined at which prediction error is minimum value.



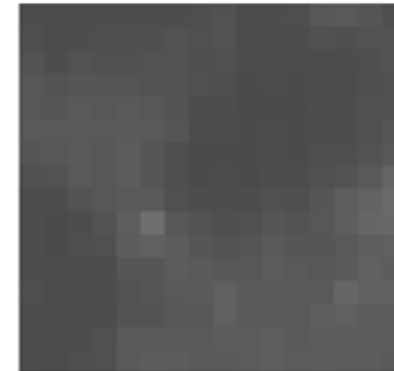
(a) Missing data of 2007/4/12



(b) Conventional KF



(c) Proposed KF



(d) Missing data of 2008/1/9



(e) Conventional KF



(f) Proposed KF

Fig. 5 Examples of the predicted VNIR band 2 images based on the conventional Kalman filter (KF) and the proposed Kalman filter with MODIS band 1 as a mean value

Table 1 shows the difference between averaged pixel value of VNIR band 2 and the corresponding pixel value of MODIS band 1. It is found that the difference between both ranges from -7 to +11%.

TABLE I. Difference between averaged pixel value of VNIR band 2 and the corresponding pixel value of MODIS band 1

Observation Date	VNIR B2	MODIS B1	% Diff.
2004/12/15	65.4	61	6.728
2005/9/29	53.3	50	6.191
2007/4/12	69.1	64	7.381
2007/4/28	65.1	62	4.762
2007/5/14	67.7	60	11.374
2007/8/18	55.9	60	-7.335
2008/1/9	39.9	38	4.762
2008/4/14	64.3	60	6.687
2008/5/3	71.3	64	10.238
2008/5/19	70.9	66	6.911
		Average	5.7699
		St.Dev.	5.0659

Figure 6 shows prediction error as a function of standard deviation for the data of May 14 2007 and August 18 2007. It is found that the most appropriate standard deviation depends on the time series data in concern.

Noise dependency of the proposed modified KF is evaluated with additive noise (zero mean and standard deviation of 5, 10, and 15 % of the mean value). Figure 7 shows influence due to the difference between 16 by 16 pixel average of VNIR band 2 data and the corresponding 1 pixel of MODIS band 1 data on prediction error.

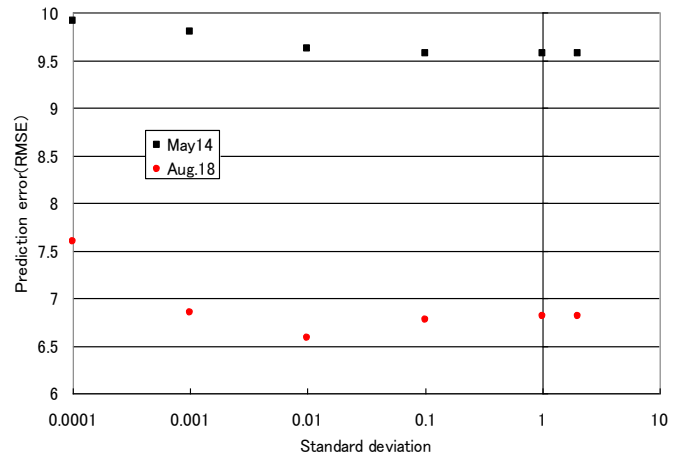


Fig. 6 Experimentally determined standard deviation of R minimizing prediction error (RMSE) for the data of May 14 2007 and August 18 2007.

Table 2 shows prediction errors for the conventional Kalman filter and the proposed modified Kalman filter. Experimental data shows that 4 to 111% of prediction error reduction can be achieved by the proposed modified Kalman filter in comparison to the conventional Kalman filter.

It is found that the reduction rate depends on the mean value accuracy of time series data derived from the other data sources. The experimental results with remote sensing satellite imagery data show a validity of the proposed method

TABLE II. Prediction errors for the conventional Kalman filter and the proposed modified Kalman filter

Missing data	Kalman filter(KF)	Modified KF	%Difference
2005/9/29	21.587	16.67	29.496
2007/4/12	16.78	13.428	24.963
2007/4/28	7.851	7.16	9.651
2007/5/14	9.339	6.289	48.498
2007/8/18	6.443	6.174	4.357
2008/1/9	22.368	10.584	111.338
2008/4/14	14.198	11.509	23.364
2008/5/3	7.901	6.802	16.157
Average	13.308	9.827	33.478
St.Dev.	6.385	3.879	34.214

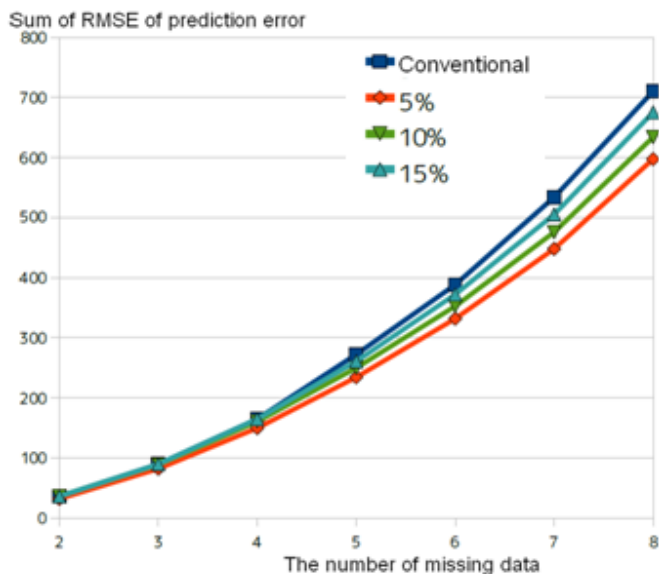


Fig. 7 Influence due to the difference between 16 by 16 pixel average of VNIR band 2 data and the corresponding 1 pixel of MODIS band 1 data on prediction error.

IV. CONCLUSION

Recovering method of missing data based on the proposed modified Kalman filter for the case that the time series of mean data is known is proposed. The proposed method allows creation of fine resolution sensor data with the aforementioned method based on a modified Kalman filter. As an example of the proposed method, prediction of missing ASTER/VNIR data based on Kalman filter using simultaneously acquired MODIS data as a mean value of time series data in revision of filter status is attempted together with a comparative study of prediction errors for both conventional Kalman filter and the proposed modified Kalman filter which utilizes mean value of time series data derived from the other sources. Experimental data shows that 4 to 111% of prediction error reduction can be achieved by the proposed modified Kalman filter in comparison to the conventional Kalman filter. It is found that the reduction rate depends on the mean value accuracy of time series data derived from the other data sources. The experimental results with remote sensing satellite imagery data show a validity of the proposed method

ACKNOWLEDGMENT

The author would like to thank Mr. Tetsuo Yamaguchi for his effort to conduct the experiments.

REFERENCES

- [1] Stratonovich, R.L., *Application of the Markov processes theory to optimal filtering*. Radio Engineering and Electronic Physics, 5:11, pp.1-19, 1960.
- [2] K.Seto and K.Arai, Prediction method for time series data analysis with relatively large missing data based on RLS method, Journal of Japan Society for Photogrammetry and Remote Sensing, 38, 5, 20-27, 1999.
- [3] K.Seto and K.Arai, Applicability of RLS method for parameter estimation of Kalman filter, Journal of Japan Society for Photogrammetry and Remote Sensing, 39, 1, 48-54, 2000.
- [4] K.Seto and K.Arai, Prediction method for time series of data with comparatively large missing data, Journal of Remote Sensing Society of Japan, 20, 1, 43-54, 2000
- [5] K.Seto and K.Arai, Effectiveness of the proposed prediction method for nonlinear and nonstationary time series analysis method with dynamic characteristic extraction, Journal of Japan Society for Photogrammetry and Remote Sensing, 39, 4, 4-12, 2000.
- [6] .Arai K. (1996), Fundamental theory for image processing, Gakujutsu-Tosho Shuppan Publishing Co., Ltd.
- [7] P.Slater, K.Thome, A.Ono, F.Sakuma, K. Arai, F.Palluconi, H.Fujisada, Y.Yamaguchi and H.Kieffer, *Radiometric Calibration of ASTER Data*, Journal of Remote Sensing Society of Japan, Vol.15, No.2, pp.16-23, Jun.1994.
- [8] Barnes, W. L., and V. V. Salomonson, *MODIS: A Global Image Spectroradiometer for the Earth Observing System*, Crit. Rev. Opt. Sci. Technol., CR47, 285-307, 1993.
- [9] George Box and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*, second edition. Oakland, CA: Holden-Day, 1976.
- [10] Hayes, Monson H., Section 9.4: Recursive Least Squares, *Statistical Digital Signal Processing and Modeling*. Wiley. p.541, 1996.
- [11] N.G. van Kampen, *Stochastic processes in physics and chemistry*. New York: North-Holland, 1981.
- [12] Japan Medical Electronics Society of Japan Edt, *Foundamentals of Biological Signal Processing*, ISBN : 978-4-339-07132-0, Corona Publishing Co. Ltd., 2007.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science, and Technology of the University of Tokyo from 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada.

He was appointed professor at Department of Information Science, Saga University in 1990. He was appointed councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.