Vol. 44(2):179–190
http://dx.doi.org/10.4217/OPR.2022009

Ocean and Polar Research

June 2022

Note

Missing Pattern Analysis of the GOCI-I Optical Satellite Image Data

Ho-Kun Jeon[1,2] and Hong Yeon Cho[1,2,3]*

[1]*Marine Bigdata Center, Korea Institute of Ocean Science & Technology, Busan 49111, Korea*
[2]*Major of Applied Ocean Science, University of Science & Technology, Busan 49111, Korea*
[3]*Division of Integrated Ocean Science and Technology, Korea Maritime & Ocean University, Busan 49112, Korea*

Abstract : Data missing in optical satellite images caused by natural variations have been a crucial barrier in observing the status of marine surfaces. Although there have been many attempts to fill the gaps of non-observation, there is little research to analyze the ratio of missing grids to overall sea grids and their seasonal patterns. This report introduces the method of quantifying the distribution of missing points and then shows how the missing points have spatial correlation and seasonal trends. Both temporal and spatial integration methods are compared to assess the effectiveness of reducing missing data. The temporal integration shows more outstanding performance than the spatial integration. Moran's I and K-function with statistical hypothesis testing show that missing grids are clustered and there is a non-random distribution from daily integration. The result of the seasonality test for Moran's I through a periodogram shows dependency on full-year, half-year, and quarter-year periods respectively. These analysis results can be used to deduce appropriate integration periods with permissible estimation errors.

Key words : optical satellite, GOCI, missing ratio, spatial pattern, integration scale

1. Introduction

Satellite image has been one of the outstanding source for marine monitoring and surveillance due to its extensive and regular observation in spatial and temporal aspects more than traditional methods using cruising ships, stationary buoys and stations.

However, optical satellite-based marine observations have significant challenges, data transformation, and missing. First, data transformation depends on developing algorithms to estimate the values of their aimed variables. On the other hand, data missing caused by high aerosol ratio, cloud, and others is a more critical barrier to utilizing optical satellite images. Accordingly, the availability of satellite images depends on its missing ratio and missing pattern.

Most research on utilizing optical satellite images has been focused on developing algorithms to convert satellite-observed radiance into desired variables or on assessing satellite-based products. However, data missing has been a decisive obstacle to the use of satellite-based products (Mercury et al. 2012; Robinson et al. 2019; Stock et al. 2020). Hence, the use of satellite-based products requires spatial-statistics analysis in advance. Mercury et al. (2012) compared cloud geographic features for the exact date of single day image and consecutive 5 and 10 years-composite images for landmass. The more compositing years, the more obvious graphical features apparent. Robinson et al. (2019) investigated cloud cover during earthquakes seasons on a global scale. They then insisted on avoiding dependency on optical satellite images because 40% of the global population is in danger when a cloud covers optical satellite views. Stock et al. (2020) compared spatial interpolation methods such as IDW (inverse distance weight), kriging, DINEOF (data-interpolating empirical orthogonal functions), and machine learning such as ridge regression and random forest. Then he/she found spatial interpolation is preferable

*Corresponding author. E-mail: hycho@kiost.ac.kr

in gap-filling approaches rather than machine learning methods.

The research described above focused on data conversion and filling, not how the missing grids have distributed on temporal and spatial axis. Thus, this report aims to quantitatively evaluate the spatial distribution of missing and give insight into the use of optical satellite images and products. First, we review flags, marked as the reason of missing data, in Geo Stationary Ocean Color Imager I (GOCI-I) chlorophyll concentration product, estimate and analyze the three-years temporal trend of missing ratio by the temporal and spatial integration scale. Then we estimate spatial autocorrelation coefficients and spatial randomness from the distribution of missing grids.

2. Materials and Methods

Materials

GOCI-I is one of the payloads on Communication, Ocean, and Meteorological Satellite (COMS). GOCI-I carried out its mission to monitor the atmosphere and ocean from June 27, 2010, to March 31, 2021, and GOCI-II on satellite GEO-KOMPSAT-2B has succeeded the duty. The GOCI-I image covers East Asia with approximately 2,500 km^2, centered on Korea, consisting of 5,585 by 5,567 pixels in the vertical and horizontal axis, with orthogonal projection. Research areas are divided into two to compare the regional characteristics of the missing ratio. One is overall GOCI-I coverage explained above, and another is East sea of Korea, ranging approximately from 129° to 133° in longitude, from 36° to 40° in latitude (Fig. 1).

In general, eight GOCI-I images are produced per day with hourly intervals from 00:15 to 07:45 UTC, and their spatial resolution is 500 m. In addition, GOCI-I has eight optical bands from between 400 to 900 nm, including NIR (near-infrared); after that, it has contributed to observing chlorophyll, turbidity, and suspended matter in the sea as well as the atmosphere (Kim et al. 2016; Jeon et al. 2020).

The data we used in this report is chlorophyll concentration product at level2a (L2A) estimated from GOCI-I. The period is from March 04, 2018 to March 26, 2021 (approximately three years, total of 1,119 days). Although it has a difference in the number of observations per day due to seasonality, it generally has eight observations per day (Table 1). There are eight days of non-observation, May 8 to 13, 2019(6 days) and June 22, 2019, to January

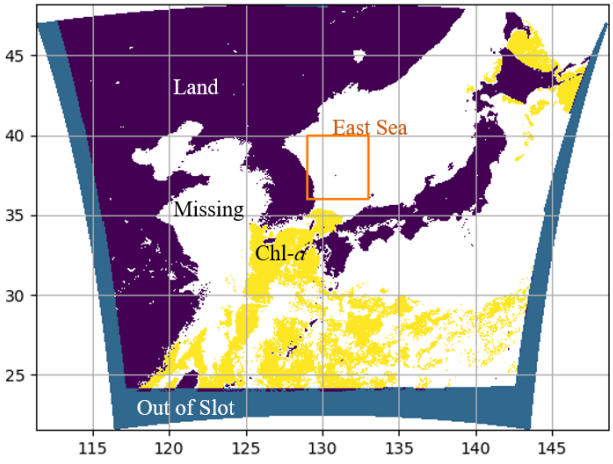

Fig. 1. GOCI-I coverage and flagging of land, out of slot, missing, and chlorophyll-*a* at 2018.04.20. 06:16 UTC

Geo Stationary Ocean Color Imager-I (GOCI-I)

- Launch: June 27, 2010
- Mission period: June 27, 2010, to March 31, 2021
- Spatial resolution: 500 m
- Temporal resolution: hourly (eight times/day, 0 to 7 UTC)
- Coverage: East Asia (2,500 km × 2,500 km), centered on 130°E, 36°N
- Image size: 5,685×5,567 pixels (31,648,395)
- Flag size: sea (17,184,846), land (11,806,138), other (2,657,411)

Table 1. GOCI-1 image data used in this note

Items / No. of images/Day	Days	No. of images	Season*
3	1	3	summer
7	31	217	winter
8(normal)	1,059	8,472	-
9	10	90	-
10	4	40	summer
11	3	33	summer
12	1	12	summer
16	2	32	spring
Total	1,111	8,899	

*Remark shows in which season the observation prevailed. Summer and winter ranges June to August and November to February, respectively. Chlorophyll products were obtained from Korea Ocean Satellite Center, Korea Institute of Ocean Science and Technology, from March 4, 2018, to March 26, 2021

26, 2020, among the period. Consequently, 1,111 days, subtracting eight days from 1,119 days, were used in this report.

The chlorophyll products are available at Korea Ocean Satellite Center (KOSC) website (http://kosc.kiost.ac.kr), operated by Korea Institute of Ocean Science and Technology (KIOST). The chlorophyll products are formatted in HE5, hierarchical data format release 5, which can contain chlorophyll concentration and flag, map projection, radiometric calibration, and others. Its size is approximately 120 MB per image.

GOCI-I defines 28 flags to minimize the error of the estimated the values of products. The specific flag for missed chlorophyll data is summarized using the rayleigh corrected reflectance products at level2c (L2C) and the chlorophyll products at L2A (Table 2). Ten samples were chosen considering minimizing the effect of seasonality and illumination of sunlight, 2018-Apr-20 06:16, 2018-Aug-06 00:16, 2018-Sep-10 07:16, 2018-Dec-14 00:16, 2019-Jun-23 03:16, 2019-Jul-03 03:16, 2020-Feb-24 06:16, 2020-Apr-29 05:16, 2020-Jul-13 03:16 and 2020-Aug-20 03:16.

Table 2. Flags effect on missing of chlorophyll concentration

	FCMR	FCMSR	
Flag of GOCI-I L2C	**10 samples**	**10 samples**	**Apr. 20, 2018**
Ancillary_Data_Warning	100.0%	100.0%	100.0%
Cotm_Pxl_Warn	45.8%	99.9%	99.8%
High_Aerosol_Ratio	99.7%	99.3%	98.5%
Rrs_Spectrum_Shape_Wrong	99.2%	98.3%	83.4%
Negative_Lw	98.3%	96.1%	82.6%
Low_Lw_555	98.1%	95.8%	81.1%
Cloud_or_Ice	67.8%	94.0%	91.9%
Negative_RhoC_NIR	96.7%	91.6%	69.6%
Atmospheric_Correction_Fail	96.5%	91.2%	65.7%
Cloud_Edge	3.8%	8.7%	8.5%
High_Aerosol	2.1%	4.7%	29.5%
HIGH_VZA	3.3%	3.0%	4.3%
Chl_Warn	1.2%	2.7%	10.6%
Chl_Fail	1.2%	2.7%	10.4%
Bright_Pxl_Adj	1.1%	2.5%	2.8%
Turbid_Water	0.3%	0.7%	2.0%
HIGH_SZA	18.0%	0.0%	0.0%
Max_Iteration	0.0%	0.0%	0.2%
Ancillary_Data_Missing	0.0%	0.0%	0.0%
Extremely_Turbid_Water	0.0%	0.0%	0.0%
High_Wind_Speed	0.0%	0.0%	0.0%
Land	42.9%	0.0%	0.0%
Low_Eps	0.0%	0.0%	0.0%
OUT_OF_SLOT	15.7%	0.0%	0.0%
Ozone_Range_Warn	0.0%	0.0%	0.0%
Rrs_Spectrum_Shape_Untrustful	3.7%	0.0%	0.0%
Slot_Info_Missing	0.0%	0.0%	0.0%
Slot_Warn	0.0%	0.0%	0.0%

Flag to Chlorophyll Missing Ratio (FCMR) uses the number of chlorophyll missing pixels on the overall area as the denominator. In contrast, Flag to Chlorophyll Missing at Sea Ratio (FCMSR) uses the number of missing grids at sea as the denominator, determined after subtracting "Land" and "OUT_OF_SLOT" flags from overall grids. In addition, the values on '10 samples' indicate the median of them, whereas 'Apr. 20, 2018' indicates the first sample, 2018-Apr-20 06:16.

Flags are obtained from GOCI-I Rayleigh corrected reflectance products because chlorophyll products do not contain flags themselves. Among the top ten flags, "Ancillary_Data_Warning" indicates an anomaly determined through five-years average trend. "Cotm_Pxl_Warn" is the contaminated pixel from more than two flags. "High_Aerosol_Ratio" is the pixel having aerosol reflectance over three-times water reflectance at any band. "Rrs_Spectrum_Shape_Wrong" is the pixel having unrealistic remote sensing reflectance (Rrs). "Negative_Lw" is the case where water-leaving radiance is less than zero. "Cloud_or_Ice" is the case where $\rho_a(865nm)+\rho_{ra}(865nm)$ exceeds 0.03. "Negative_RhoC_NIR" is the pixel that rayleigh corrected radiance is less than zero. Finally, "Atmospheric_Correction_Fail" is the case where $\rho_a(VIS)+\rho_{ra}(VIS)$ is not retrieved from $\rho_a(NIR)+\rho_{ra}(NIR)$. ρ_a and ρ_{ra} represents multiply scattered aerosol reflectance and reflectance of interactively scattered between aerosol and molecular. VIS and NIR represents the band of visible and near-infrared.

Missing of chlorophyll to sea grids ratio

The missing chlorophyll data to sea grids Ratio (MR) is computed after extracting sea area by subtracting "Land" and "OUT_OF_SLOT." MR is a reference to understand the proportion of the missed chlorophyll concentration data at sea in the image. The missing grids occur for various reasons, as referred to in section 2.1. MR is computed for every hourly image. All the sea grids with chlorophyll concentration are converted into integer one (1) instead of the actual value, and the grids without data are converted into value NaN (-). Thus, we have binary information to determine the grids' observation status between presence (1) and absence (0). The Daily Mode (DM) is generated to reduce the number of missing grids (Fig. 2a).

DM is generated through the temporal integration process. First, hourly images on a day are accumulated on a time axis. Then the individual spatial grid is examined whether it has any missing on the time axis. If there is

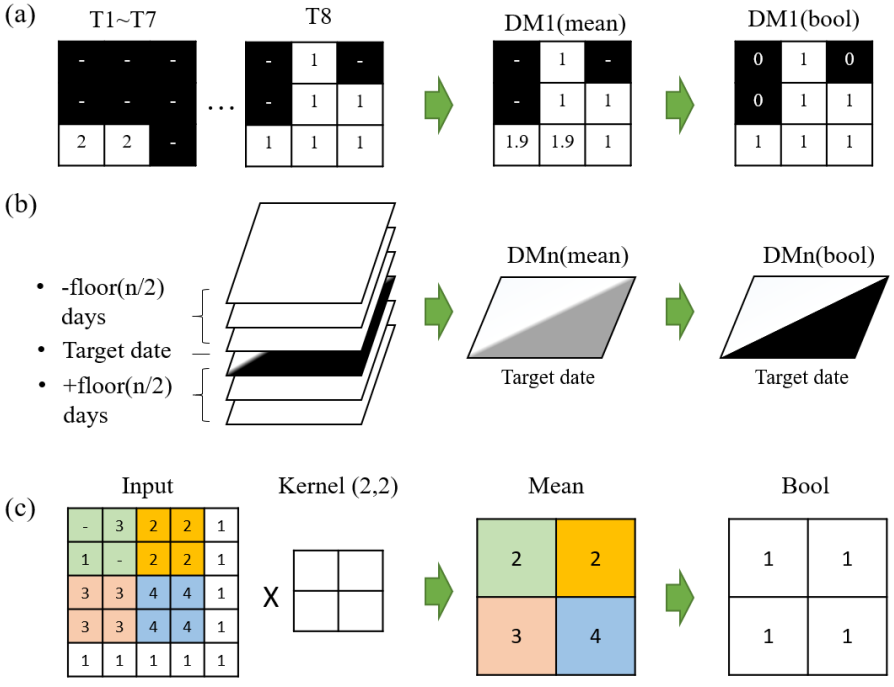

Fig. 2. Integration process. (a) temporal integration to generate daily mode, DM (b) temporal integration to generate multiple days mode, DMn. (c) spatial integration

missing, the grid has integer zero (0), unless otherwise, the grid has integer one (1). Consequently, the increasing number of integration images causes a decrease in missing grids. Then DMn - n represents the days of integration - is made for 3, 5, 7, and 9 days each by using the DM (Fig. 2b). Then the DM and DMn are formatted in NetCDF, array-oriented scientific data, having file extension NC.

The DMn indicates whether it has at least one piece of information for a given grid over a specific period. Like temporal integration bring a missing reduction effect, spatial integration also brings a missing reduction. Thus, $n \times n$ square matrix kernel is used, increasing the value of n, to compare the effect of missing reduction upon the kernel size (Fig. 2c).

Spatial autocorrelation index of missing-grids locations

Before computing spatial autocorrelation of GOCI-I missing grids, DMn is scaled into 100 by 100 pixels, since the original size, 5,585 by 5,567 pixels in the vertical and horizontal axis, leads to too massive computation works.

Both overall coverage and the East sea of the GOCI-I are analyzed through Moran's I, one of the representative spatial autocorrelation indices, to evaluate how the missing grids correlated one to neighbors (Cliff and Ord 1969; Anselin 1995; Jossart et al. 2020). The weights required to compute the respective index are calculated using IDW (inverse distance weight), the inverse of the distance between missing grids (Eq. 1).

$$w_{ij} = 1/d_{ij} \quad \text{where, } d_{ij} = D(p_i, p_j) \qquad \text{Eq. (1)}$$

Moran's I is computed through Eq. 2.

$$\text{Moran's I: } I = \frac{N}{W} \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad \text{Eq. (2)}$$

where, N denotes the number of all grids corresponding to ocean, W denotes the summation of all weights (w_{ij}), (x_i,x_j) denotes whether missing grids is true (1) or false (0). $\bar{x}$ denotes arithmetic mean of every grids expressed 1 or 0, and w_{ij}: weight computed by the function of distance between two grids (x_i,x_j).

Pattern analysis of spatial point distribution

Pattern analysis of spatial point distribution begins with the null hypothesis of complete spatial randomness (CSR), in which spatial grids are homogeneous and isotropic in space. Hence, test whether the distribution of data matches CSR.

To evaluate the randomness or cluster of cloud cover, K-function, one of the distance-based spatial point pattern functions using R package *"spatstat"*(Baddeley and Turner 2005). K value is computed using Eq. 3. The nominator denotes that the number of events where two points (d_{ij}) distance is less than specific distance r while the denominator represents that ratio of the number of every combination of two points in the unit region to the area of unit region (πr^2).

$$\text{K-function: } K(r) = \frac{N(d_{ij} < r)}{\lambda} \qquad \text{Eq. (3)}$$

where, N denotes the number of events where two points distance is less than r, d_{ij} denotes the distance between points i and $j(i \neq j)$, and λ denotes the ratio of N to the area of the unit region (πr^2).

On the other hand, the K value that represents CSR can be expressed as in Eq. 4. Once the distribution of spatial points is independent, that is not affected by the other, and the random variable of the distance between two points (d_{ij}) can be assumed as the spatial randomness and independent. Thus Poisson distribution can be considered here. Given that the density of points (λ) and the variable of distance are constant, the number of combinations d_{ij} within a specific distance r is proportional to the circle are with radius $r(\pi r^2)$.

If $K(r)$ is higher than $K_{CSR}(r)$,

$$K_{CSR}(r) = \frac{E[N(d_{ij} < r]}{\lambda} = \frac{\lambda \pi r^2}{\lambda} = \pi r^2 \qquad \text{Eq. (4)}$$

Seasonality analysis

Although we might assume a seasonality in time-series graphs by glancing, reasonable seasonal periods can be computed from Fisher G-test and Whittle test (Fisher 1929; Hernandez 1999; Akmaev 2003; Cho 2020). The basic is the harmonic analysis that finds trends and periods in time-series data. Specifically, harmonic functions such as

sine and cosine are widely used for spectrum analysis, Fourier transform and fast Fourier transform (FFT) related to the frequency that is inversed period.

The periodogram method is chosen to analyze frequency which will be converted into a period. The periodogram is a diagram that illustrates the magnitude of the coefficient of Fourier transform and is thus useful to detect superior frequency. The magnitude was computed using R package *"TSA"*. Given that the magnitude of frequency is significant, the frequency is adopted as a reasonable one, and then the inversed value of frequency becomes the principal period. F-value and F-critical value were first computed using Eq. 5 and R function *"qf"* to obtain significant frequency candidates through threshold value (Weerahandi 1995).

$$\text{F-value: } F_k = \frac{(n-3)I(w_k)}{2\left\{\sum_{1}^{n/2} I(w_k)\right\} - I(w_k)} \quad \text{Eq. (5)}$$

where, n denotes the number of data, I denotes the magnitude of Fourier transform coefficient, w_k denotes angular frequency.

The statistics for the periodicity test are computed using Eq. 6 & 7. Fisher statistics is used to get g_1 for the primary magnitude of frequency while Whitle statistics is used to get g_m from the second magnitude of frequency.

$$\text{Fisher statistics: } g_1 = \frac{\max\{I(w_k)\}}{\sum_{k=1}^{n/2} I(w_k)} = \frac{I(w_k)_{(1)}}{\sum_{k=1}^{n/2} I(w_k)} \quad \text{Eq. (6)}$$

$$\text{Whittle statistics: } g_2 = \frac{I(w_k)_{(2)}}{\left\{\sum_{k=1}^{n/2} I(w_k)\right\} - I(w_k)_{(1)}} \text{ or } g_m = \frac{I(w_k)_{(m)}}{\left\{\sum_{k=1}^{n/2} I(w_k)\right\}} \quad \text{Eq. (7)}$$

where, subscript (m) denotes rank of the $I(w_k)$ in descending order.

In fact, the exact critical value g_c is computed through a complicated equation in accordance with Fisher test method and thus we use the approximate formula, in which the significant level (α) is 0.05(Eq. 8) Given that g_m is greater than g_c, it means the frequency is significant to stand for periodicity in m-th order. Inversely, the contrast case means the null hypothesis of none seasonality is adopted.

$$\text{Critical value: } g_c = 1-(\alpha/m) \quad \text{Eq. (8)}$$

where, $m = [n/2]$, $P(g > g_c) = \alpha$

3. Results and Discussion

Variation of missing ratio (MR) for different integrating periods

The temporal integration was made from March 4, 2018, to March 26, 2021, through the method described in Fig. 2 above. Hourly image (HR1) shows the highest MR with an average of 80.6%(Fig. 3). Thus HR1 is not proper data to monitor the wide sea area. DM3 shows the most effective in MR reduction rather than other steps.

MR has seasonal trends where the missing is most excessive in winter between November and February and rarely in autumn between September and October.

Figs. 4 and 5 show the variation of the spatial distribution of missing data according to DMn and kernel size, as illustrated in Fig. 2c above, for the same date, March 25, 2018, respectively. The DMn method effectively reduced missing grids by 7.78% at DM9 while the kernel method was ineffective, having 44.09% at the eight-by-eight kernel size, since the individual missing grid was clustered with one another and the kernel size was not enough to obtain a valid grid. This result proves that spatial integration is not preferable to temporal integration in removing missing points.

Fig. 6a and c shows the monthly mean variation of MR of temporal and spatial integration for GOCI-I coverage, respectively. DMn method shows more effectiveness than the kernel method in reducing missing points. The seasonality becomes more evident than the variation of DMn (Fig. 3). On the other hand, the monthly variation of MR i n the East sea is not significantly different from that of GOCI-I coverage in the overall trend (Fig. 6b).

The error of estimation in the missing grids is highly affected by the MR and the pattern of missing-grids spatial distribution. Therefore, a certain level of reducing MR is required to estimate within the allowable error range considering spatial correlation since the higher MR, the higher

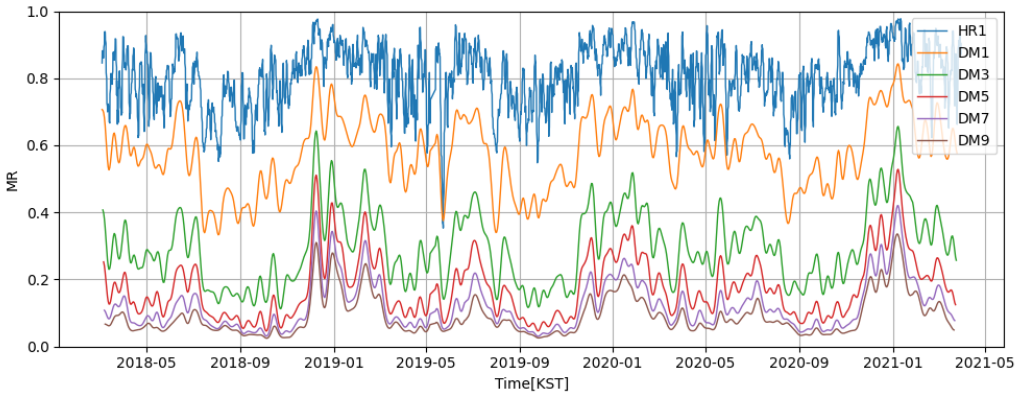

Fig. 3. Daily variation of MR for different integrating periods. HR1 and DMn represents hourly and multiple-days mode, respectively. The MR values are smoothed with 3 days bandwidth

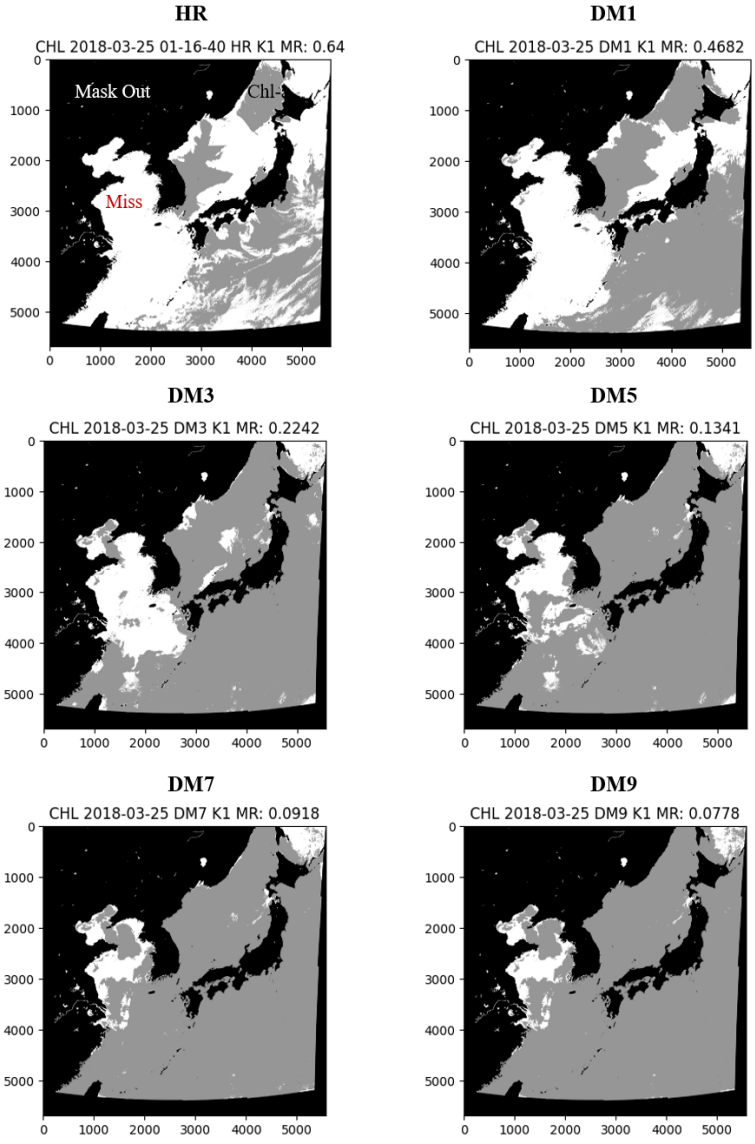

Fig. 4. Distribution of MR according to temporal integration days for March 25, 2018

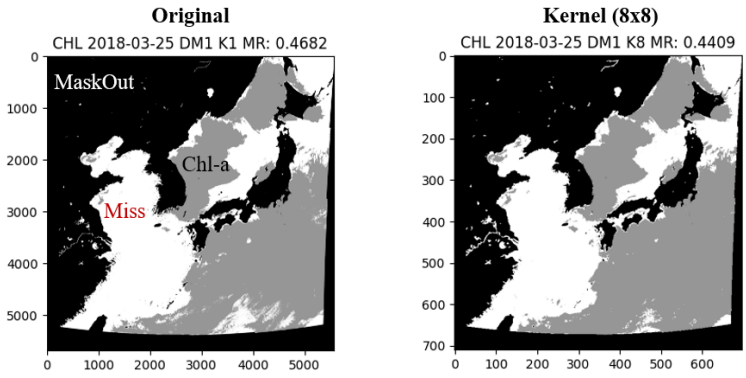

Fig. 5. Distribution of MR according to spatial integration. (left) original DM1 (right) kernel mean of DM1 with 8 by 8 kernel for March 25, 2018

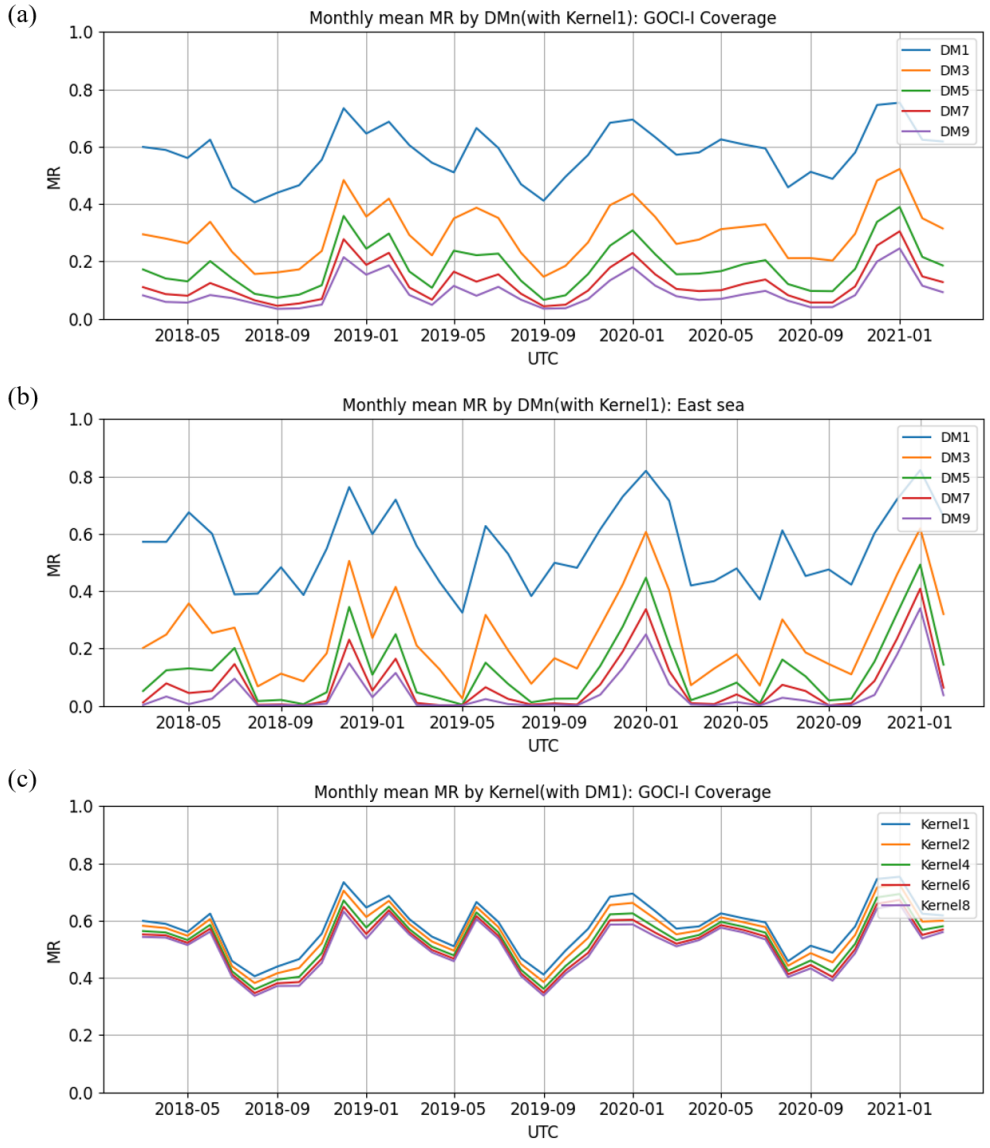

Fig. 6. Monthly missing-grids ratio (MR). (a) GOCI-I coverage varying at DMn. (b) East Sea varying at DMn. (c) GOCI-I coverage varying at kernel size

the estimation error. Hair Jr. et al. (2010) suggests that 10% or less missing is permissible. However, despite the n of DMn increase up to nine, there was no perfect reach to 10% for the whole period (Fig. 3). Thus, it may consume more data for up to a month to meet the MR degrading temporal resolution.

Therefore, it is required to analyze the variation of estimation error according to the reduction of the MR since the reduction of estimation error is a fundamental criterion. In the practical aspect, around 20% MR is appropriate if the temporal resolution is maintained for around five days. The appropriate MR needs to be re-determined considering allowable errors or unavoidable estimation errors because stricter values may show differences depending on the estimated items other than chlorophyll in the missing grids.

This is the trade-off between the decrease in the temporal resolution and the decrease in MR. An optimization issue is balancing benefits and losses from the reduced analytic resolution. The 20% MR and the five-days temporal resolution are presented as the optimal reference conditions in this analysis. By introducing an appropriate cost function, this condition can be changed.

Variation of Moran's I for different integrating periods

Moran's I, one of the spatial autocorrelation indices, indicate that spatial points have autocorrelation when the value is farther from zero. In detail, values over zero indicate that spatial points have a positive autocorrelation where the attribute of points is clustered neighbor to neighbor. In contrast, under zero indicates the negative autocorrelation where the points attribute is opposite neighbor to neighbor like check pattern distribution. On the other hand, values closer to zero indicate that spatial points have a random distribution.

Daily Moran's I was computed for March 4, 2018, to March 26, 2021, and the overall coverage of GOCI-I using 100-by-100 rescaled DMn (Fig. 7a). Overall DMn ranges from 0.03 to 0.26. DM1 shows the highest variation while DM9 shows the lowest variation. The Moran's I values become lower with the increasing number of integrating periods and become stable due to lower variation.

In addition, daily Moran's I for the East sea was also computed (Fig. 7b). Overall DMn ranges from 0.00 to

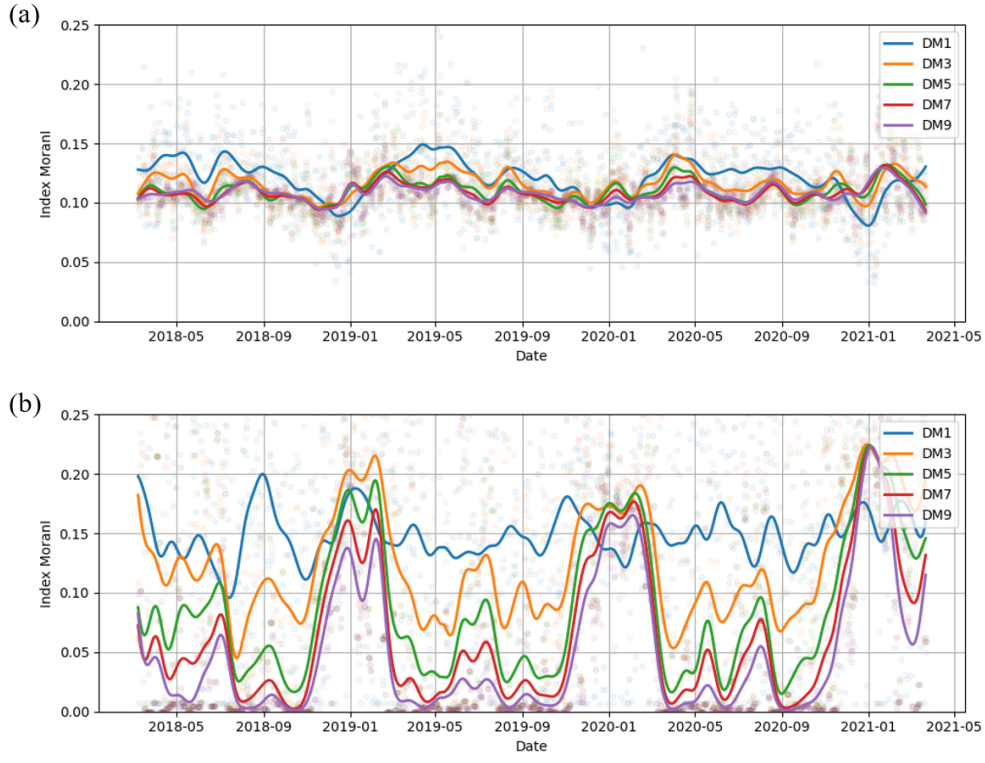

Fig. 7. Variation of Moran's I of missing grids at different integrating periods. Kernel smooth is applied to MR values with 10-days bandwidth using R package *"imputeTS"*. (a) GOCI-I coverage. (b) East sea

0.36. From DM3, the seasonality of Moran's I become apparent, showing the highest in winter and the lowest in autumn.

Distributions of Moran's I for missing grids of daily DMn were compared using a violin plot that illustrates the distribution of data and quantile simultaneously (Fig. 8). The violin-shaped curved area represents the distribution of data. The black-flatten square in the violin indicates the inter-quantile range between 25%(Q1) and 75%(Q3). The white dot in the black square indicates the median value. A small decrease in median between DMn represents spatial autocorrelation is less affected by the temporal integration period. The reduced variance indicates that the spatial autocorrelation is gradually stable.

Variation of K-function for different integrating periods

The comparison of K-function was made between DM1 and DM5 on March 21, 2018. As mentioned in section 2.4, the spatial point distribution analysis is based on the null hypothesis that the individual point in the region of interest is independent of others and has complete spatial randomness.

The observations K of DM1 values are greater than those of theoretic CSR values on every two points' distance r when the missing prevails overall sea (Fig. 9). Thus the spatial distribution is determined as clustered and not in a random state. Likewise, DM5 also shows clustered distribution and none of the randomness.

Seasonality of MR

Periodical trends were found in every time-series plot (Figs. 3, 6 and 7). The Periodogram was used to detect superior frequency and period, representing seasonality. Three frequencies were detected through F-values and F-critical values.

F-test and Whittle-test were made for the three frequencies (Table 3). The seasonal periods, the value of inversed superior frequencies, were a year, half-year, and

Table 3. Seasonality test. Superior frequency, days and their observed g_m

	Freq.	Year	Obs. g_m	Remark
Year	0.0026	≒ 1	0.224	Fisher
Half Year	0.0053	≒ 1/2	0.157	Whittle
Quarter	0.0106	≒ 1/4	0.025	Whittle

*Year, half year, and quarter periodicity have 0.016 as g_c in common

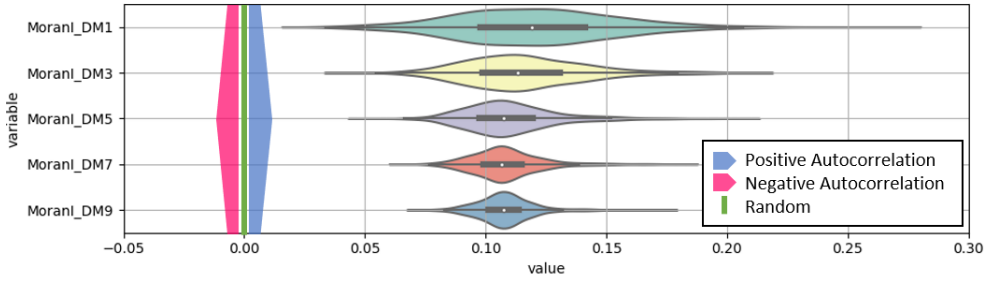

Fig. 8. Distribution of daily Moran's I of missing grids at different temporal integration for GOCI-I coverage

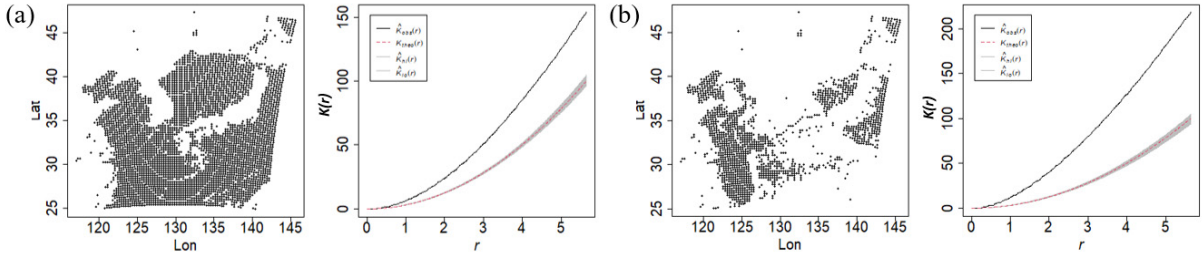

Fig. 9. Spatial point distribution and K-function with DM1&5 for March 21, 2018. Black dots represent the missing grids at sea. A black line and a dashed red line represent the K value of the observation and CSR varying at distance *r*, respectively. (a) DM1. (b) DM5

a quarter. Because every observed g_m of the three periods was above the critical g-value of 0.016, the null hypothesis of none-seasonality is rejected, and they are considered as significant periods which are representative of seasonality.

4. Summary & Conclusions

This note described how missing grids are distributed in amount, spatial pattern, and seasonality aspects. We also answered three questions 1) what is the optimal choice to reduce the missing ratio 2) do the missing grids randomly distributed without pattern 3) Whether the missing ratio has seasonality and what is the interval.

The missing grids of GOCI-1 satellite images are caused by high aerosol ratio, atmospheric correction failure, clouds, and others. The average MR of GOCI-I chlorophyll product was approximate 80±15%. A comparison between temporal and spatial integration was made to reduce the MR. Consequently, temporal integration is revealed as a better method than spatial one. The five-days temporal integration reduced the number of missing data by 20% significantly while spatial integration was not practical where it reduced the missing data by 40 to 60%. The spatial distribution of missing grids was analyzed through Moran's I, a spatial autocorrelation index, and Ripley K, a spatial distribution function. Both methods are based on the null hypothesis of complete spatial random (CSR). The test result shows that the null hypothesis is rejected and reveals that the missing pattern is not random but clustering weakly. The seasonality of MR and its significant period was confirmed through Fisher G-test. The result shows the MR has yearly, half-yearly, and quarter-yearly component exists.

In the future study, we plan to generate the complete chlorophyll product where the missing at sea is filled, considering estimation error, with a week-level resolution and 250m or less spatial resolution using the GOCI-I product based on the result of this report.

Acknowledgments

This study was conducted as part of the 「Research on Sustainable Use of Dokdo」 project of the Ministry of Oceans and Fisheries.

References

Akmaev RA (2003) Comment on "Time series, periodograms, and significance" by G. Hernandez. J Geophys Res **108** (A5):1187. doi:10.1029/2002JA009687

Anselin L (1995) Local indicators of spatial association-LISA. Geogr Anal **27**(2):93–115. doi:10.1111/j.1538-4632.1995.tb00338.x

Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. J Stat Softw **12**(6):1–42. doi:10.18637/jss.v012.i06

Cho HY (2020) Periodicity test of the time-series data. KIDS Rep **2**(1):8–18

Cliff AD, Ord JK (1969) The problem of spatial autocorrelation. In: Scott A (ed) London papers in regional science. Pion, London, pp 25–55

Fisher RA (1929) Tests of significance in harmonic analysis. P R Soc A **125**(796):54–59. doi:10.1098/rspa.1929.0151

Hair Jr. JF, Black WC, Babin BJ, Anderson RE (2010) Multivariate data analysis: a global perspective. Pearson Education, London, 44 p

Hernandez G (1999) Time series, periodograms, and significance. J Geophys Res **104**(A5):10355–10368. doi:10.1029/1999JA900026

Jeon H-K, Kim S, Edwin J, Yang C-S (2020) Sea fog identification from GOCI images using CNN transfer learning models. Electronics **9**(2):311. doi:10.3390/electronics9020311

Jossart J, Theuerkauf SJ, Wickliffe LC, Morris Jr. JA (2020) Applications of spatial autocorrelation analyses for marine aquaculture siting. Front Mar Sci **6**:806. doi:10.3389/fmars.2019.00806

Kim W, Moon JE, Park YJ, Ishizaka J (2016) Evaluation of chlorophyll retrievals from geostationary ocean color imager (GOCI) for the north-east Asian region. Remote Sens Environ **184**:482–495. doi:10.1016/j.rse.2016.07.031

Mercury M, Green R, Hook S, Oaida B, Wu W, Gunderson A, Chodas M (2012) Global cloud cover for assessment of optical satellite observation opportunities: a HyspIRI case study. Remote Sens Environ **126**:62–71. doi:10.1016/j.rse.2012.08.007

Robinson RT, Rosser N, Waters RJ (2019) The spatial and temporal influence of cloud cover on satellite-based emergency mapping of earthquake disasters. Sci Rep **9**:12455. doi:10.1038/s41598-019-49008-0

Stock A, Subramaniam A, Van Dijken GL, Wedding LM, Arrigo KR, Mills MM, Micheli F (2020) Comparison of cloud-filling algorithms for marine satellite data. Remote

Sensing **12**(20):3313. doi:10.3390/rs12203313
Weerahandi S (1995) ANOVA under unequal error variances. Biometics **51**(2):589–599. doi:10.2307/2532947

Received Jan. 14, 2022
Revised Mar. 15, 2022
Accepted Mar. 29, 2022

Author's Information

Ho-Kun Jeon
Student Researcher, Korea Institute of Ocean Science & Technology

Hong Yeon Cho
Principal Research Scientist, Korea Institute of Ocean Science & Technology
Professor, University of Science & Technology
Professor, Korea Maritime & Ocean University