JARQ 56 (1), 77-94 (2022) https://www.jircas.go.jp

Time Series Lengths for the Accurate Isolation of Major Tidal Components by Simple Fourier Analysis

Katsushi SHIRAHATA*, Shuhei YOSHIMOTO, Takeo TSUCHIHARA, Hiroomi NAKAZATO and Satoshi ISHIDA

Institute for Rural Engineering, National Agriculture and Food Research Organization, Tsukuba, Japan

Abstract

Fourier analysis applied to tidally fluctuating groundwater observation time series data can extract sinusoidal tidal components and help in the precise use of a tidal response method to estimate aquifer hydraulic parameters. To accurately isolate one tidal component with a specific period or accurately determine the amplitude and initial phase of the sinusoidal tidal component, an appropriate length for the time series to be analyzed must be selected. Errors are inevitable when Fourier analysis is used to isolate a tidal component from observation data due to the finite length of the analyzed time series. The errors stemming from two sources in the tidal-component isolation were extensively investigated using Fourier analyses of artificial time series with varying lengths that contained a major or relatively major tidal component. Based on the investigation, the recommended criteria for selecting time series lengths for the accurate isolation of major semidiurnal and diurnal tidal components were organized. The criteria, presented with their derivations, provide a practical guide for applying Fourier analysis to observation time series data to accurately isolate major tidal components.

Discipline: Agricultural Engineering
Additional key words: discrete Fourier transform, finite-length time series, Fourier series expansion, spectral leakage, tidal constituent

Introduction

As rivers are absent in hydraulically permeable geologic settings, groundwater is often a major source of water for domestic and agricultural use on oceanic coral or limestone islands. Many such islands belong to developing countries, such as most of those that comprise the Small Island Developing States (White & Falkland 2010, Chattopadhyay & Singh 2013, Ishida et al. 2015, Werner et al. 2017, Dahan 2018, Yoshimoto et al. 2020). To sustainably develop agriculture and improve the living environment in these areas, strategies integrated with an appropriate plan for the development and management of groundwater resources based on the understanding of the properties of groundwater and aquifers are required.

Shirahata et al. (2019a, 2019b) used tidal response methods to estimate the hydraulic parameters of aquifers beneath an island where appropriate groundwater development for agricultural use was desired. Accurate aquifer parameters are needed to construct and develop a groundwater numerical model used for predictions regarding limited groundwater resources in a freshwater lens on the island (Shirahata et al. 2019a, 2019b). In their tidal response methods, Fourier analysis of groundwater-level time series data enabled the isolation of known major tidal components and the determination of the amplitudes and initial phases of the sinusoidal components of specific tidal periods. These amplitudes and initial phases were used in the calculation of aquifer hydraulic parameters.

Fourier analysis, which in this paper denotes the analysis of real-valued finite-length discrete-time time series data using the discrete Fourier transform or based on the Fourier series expansion formula, is widely used in hydrology and hydrogeology to decompose observation time series data containing tidal and other components. For example, Fourier analysis has been used to describe the frequency composition of data, investigate the factors causing fluctuations, detect specific frequencies such as those of major tidal components, and extract or isolate

*Corresponding author: shirahatak@affrc.go.jp
Received 10 September 2020; accepted 30 March 2021.

sinusoidal tidal components (Lanyon et al. 1982, Koizumi 1991, Ishitobi et al. 1993, Evans 2004, Möller et al. 2007, Alcolea et al. 2009, Aichi et al. 2011, Rahi & Halihan 2013, Acworth et al. 2015, Dong et al. 2015, Burgess et al. 2017, Fuentes-Arreazola et al. 2018, Shirahata et al. 2018). For this last purpose, Fourier analysis has the advantage of being simply performed using standard spreadsheet software with built-in functions (Whitford et al. 2001, Shirahata et al. 2018). Such software is available to people who do not have specialized software or programming expertise, such as a local administrative officer in charge of the development and management of limited groundwater resources on a remote island.

In the Fourier analyses of Shirahata et al. (2019a, 2019b), the analyzed time series lengths were selected so that the output amplitude and initial phase of the desired isolated tidal component would have only small errors, following the recommendation of Shirahata et al. (2017). Shirahata et al. (2017) investigated tentatively selected four time series lengths (708 h, 3,279 h, 4,380 h, and 8,856 h) and examined the errors in the Fourier analysis outputs, which were the amplitudes and initial phases of major tidal constituents M2, K1, S2, O1, and P1. The four time series lengths were selected in light of a basic constraint in the Fourier analysis that the analyzed time series length has to be an integer multiple of the period of the isolated component. For instance, to accurately isolate M2, with a period of 12.420601 h, the time series must closely approximate a multiple of the M2 period, such as 708 h (approximately 57 times the M2 period). When Fourier analysis is used to isolate a tidal constituent of a fixed period from a data series of arbitrary length, the output tidal constituent period, given by a quotient of the analyzed data length divided by an integer, generally deviates from the exact tidal constituent period. The deviation, hereafter called the period approximation deviation (PAD), is a basic source of error in the isolation of a tidal constituent. This error due to the PAD of the desired isolated tidal constituent is referred to in this paper as first-category error. It appears as a reduction of the output amplitude of the desired tidal constituent accompanied by the leakage of the amplitude to adjacent frequencies in the frequency spectrum (de Levie 2004, Thomson & Emery 2014, Shirahata et al. 2017), referred to here as spectral leakage. Trefry & Bekele (2004) recognized this error in their study using Fourier analysis to determine the amplitudes of tidal components, and they reproduced the amplitude of a major tidal constituent from three output amplitudes at three consecutive frequencies.

Another category of error was described by Shirahata et al. (2017), who demonstrated that a large PAD of a nearby tidal constituent, that is, a constituent with a frequency or period close to that of the desired constituent, can affect the isolation of the desired constituent and result in a large error in the output when the true amplitude of the nearby constituent is comparable with that of the desired constituent. This error is unavoidable in the Fourier analysis of observation data containing many tidal constituents with different periods. This second-category error is relevant to the spectral leakage of nearby tidal constituents, and it becomes non-negligible when the PAD of a nearby constituent is significant (Blair 1979, Shirahata et al. 2017).

Tanaka et al. (1967) studied errors in the amplitudes and phases of eight major tidal constituents obtained by Fourier analyses of earth-tide time series records. However, the analyses used only two time series lengths of 8,784 h and 8,760 h. Blair (1979) theoretically studied the second-category error in the amplitudes of tidal constituents determined by Fourier analyses of finite-length time series of varying lengths. The study was limited to the isolations of one semidiurnal constituent M2 and one diurnal constituent O1 (period: 25.819342 h). In addition, Blair's discussion laid emphasis on the total possible error owing to several nearby tidal constituents (five semidiurnal constituents for the isolation of M2 and five diurnal constituents for the isolation of O1), not on separate errors owing to each constituent. Furthermore, the derived error prediction formula presupposed that the output frequency was exactly the frequency of the desired tidal constituent M2 or O1 or, equivalently, that the analyzed time series length was an exact multiple of the period of the desired constituent. Naturally, the study also ruled out the possibility of a first-category error. Nowadays, as available observation time series data usually consist of exact whole numbers of time units (e.g., 1 h), it rarely happens that the data length is equal to a multiple of a tidal constituent period; therefore, Fourier analyses of the type treated here should always consider the first-category error. For these reasons, it has been difficult to widely apply the results of Blair (1979) to the Fourier analysis of real observation data. In the study of Shirahata et al. (2017), the first-category error was avoided beforehand in the tentative selection of the time series lengths, and thus was not fully examined. Second-category errors were clearly exemplified in the discussion of Shirahata et al. (2017), but the examples presented were limited to the combination of two major tidal constituents, the desired isolated S2 (period: exactly 12 h) and the affecting nearby M2, and they only considered 20 representative time series lengths up to 8,856 h.

The present study aimed to organize a set of criteria for selecting time series lengths appropriate or recommendable for their use in Fourier analysis to isolate major tidal constituents. We extensively investigated the first- and second-category errors involved in the isolation of eight major semidiurnal and diurnal tidal constituents. The investigation did not use an approximate analytical solution of an equation for predicting errors such as Blair (1979) derived but instead relied on the concrete calculations of the amplitudes and initial phases of the major tidal constituents by simple Fourier analysis of artificial time series in lengths of whole hours between 600 h and 30,000 h. The results of the Fourier analyses, the magnitudes of the amplitude reductions and the spectral leakages, were investigated in relation to the PAD of the tidal constituent and the analyzed time series length. From this effort, we derived empirical formulas that predict the errors for time series even longer than 30,000 h. The formulas, combined with representative relative amplitudes of tidal constituents expected in nature, provide a set of PAD criteria for selecting time series lengths for the accurate isolation of the eight major tidal constituents. This paper presents the investigations of the errors in the tidal constituent isolation by Fourier analysis and discusses the derivation of the criteria. An example list of the time series lengths that pass the criteria is given, and a brief demonstration using real groundwater observation data is provided.

Methods and notes

1. Investigation of errors in the isolation of major tidal constituents by simple Fourier analysis

In this study, the isolation of tidal constituents from time series, that is, the calculation of the amplitude and initial phase of sinusoidal tidal components, followed the Fourier analysis technique described by Shirahata et al. (2014, 2017), with a slight modification. The calculations simply used formulas for Fourier series expansion as described in these studies. The one difference is that in the present study, the origin of the time, where the time $t = 0$, of the analyzed hourly sampled time series data was placed at the center of the analyzed length if it had an odd length (i.e., with an odd number of data points) and at 0.5 h from the center of the analyzed length if it had an even length. In previous studies, in contrast, the time origin was always placed at the start of the analyzed time series. Because the initial phase calculated by the simple Fourier analysis technique is the phase at the time origin, its setting may affect the error in the output initial phase. The effect of the setting of the time origin on the output initial-phase error is briefly described in the Results section.

The analyzed time series data were composed of a sinusoidal component that had the exact period of a known tidal constituent. The values of the time series data $f(t)$ for hourly interval time t (h) were calculated as follows:

$$f(t) = A \cdot \sin(2\pi \cdot T^{-1} \cdot t + IP), \quad (1)$$

where T is the period (in h) of the tidal constituent, A is the amplitude, and IP is the initial phase (in radians or rad). Figure 1 shows an example of the time series analyzed for the investigation in this study.

Table 1 lists the 15 tidal constituents used in this study with their angular frequencies and periods. The constituents and their frequencies were excerpted from the list given by Kudryavtsev (2004), and the periods were calculated from the frequencies in the list. These are the top major 15 constituents in tide-generating potential among the semidiurnal and diurnal constituents listed by Kudryavtsev (2004). These 15 tidal constituents include all the constituents covered by Blair (1979), who was not concerned with the 2N2, L2, or T2 constituents. In this article, the abbreviated notations written in parentheses in Table 1 are occasionally used to enumerate the eight constituents M2, K1, S2, O1, P1, N2, K2, and Q1 or “MKSOpnkq” in decreasing order of tidal potential, which are often referred to as the eight principal or major tidal constituents (Tanaka et al. 1967, Dushaw et al. 1995, Kowalik & Polyakov 1998, Foreman et al. 2009, Siddig

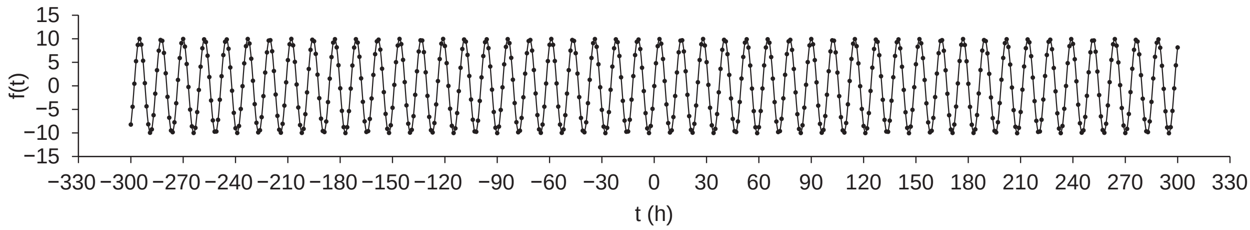

Fig. 1. Example of the hourly sampled time series data subjected to the Fourier analysis for the investigation in this study
Time series data composed of one tidal constituent M2 (period: 12.420601 h), with an amplitude of 10 and initial phase of zero rad, with a finite length of 601 h

et al. 2019). They were focused on in this study as the desired isolated constituents. Among them, the first four (“MKSO”) are referred to as the four major tidal constituents (Nishida 1980, Kowalik & Polyakov 1998, Karang et al. 2010, Siddig et al. 2019).

The length of the time series subjected to Fourier analysis was varied from 600 to 30,000 h with an hourly interval. The Fourier analysis technique applied to a finite-length time series can output the amplitude and initial phase for every discrete frequency (e.g., in cycles per hour [cph]) that is a multiple of the reciprocal of the analyzed time series length (e.g., in h). This study focused on the outputs for the frequencies that represent the eight major tidal constituents. Among the discrete frequencies for which outputs are given, the frequency that represents a tidal constituent was defined as the one closest to the exact frequency of the tidal constituent. In this paper, the Fourier analysis output (amplitude and initial phase) and the output frequency in the spectrum for a tidal constituent are denoted by terms such as “M2 output” and “M2 (frequency) site”; however, they do not necessarily indicate the output for the exact frequency of the tidal constituent. To put it differently in terms of the period, the Fourier analysis of a finite-length time series isolates sinusoidal components of a period that is a quotient of the analyzed time series length divided by an integer. One of the isolated components has a period that is the reciprocal of the output frequency closest to the exact tidal constituent frequency. The period approximates but not necessarily equals the exact period of the tidal constituent. The PAD of a tidal constituent is defined in this paper as follows:

$$D = |T_{APX} / T - 1|, \tag{2}$$

where D is the PAD, T_{APX} is the output approximate tidal constituent period as a quotient of the analyzed time series length, and T is the exact tidal constituent period. For example, the PADs of the S2 constituent for time series lengths 1,197 h and 1,203 h are both 0.250%. The PAD varies with the tidal constituent concerned and with the analyzed time series length.

The first-category error for each of the eight major tidal constituents was measured by performing Fourier analyses of time series composed of a single sinusoidal component of the exact tidal period with an amplitude of 10 and an initial phase of zero rad. The errors in the output amplitude and initial phase for the frequency that represents the tidal constituent (not necessarily the exact frequency of the tidal constituent) were calculated by comparison with the true amplitude and initial phase given in the generation of the time series. In the Results section, the magnitudes of the amplitude and initial-phase errors are expressed in relative and absolute terms, in percent and radians, respectively.

Fourier analyses of time series composed of one of the 15 tidal constituents, with an amplitude of 10 and an initial phase of zero rad, were performed, and the amplitude leakages of the contained constituent to the frequency that represents another desired constituent were computed. This leakage is the source of the second-category error in the isolation of the desired tidal constituent. The analysis results for time series lengths that caused the two relevant tidal constituents (the affecting nearby constituent contained in the time series and the affected desired constituent) to share the same Fourier analysis output frequency were excluded from the investigation. Such results of the analysis are not included in the description below. Such cases are typical for short time series when the frequencies of the two constituents are especially close. For instance, for the combination of K1 and P1 constituents, among 1,000 time series lengths of 600 h-1,599 h, this was the case for 748 lengths.

The “amplitude leakage” or “amplitude leak” quantified in this paper always refers to the ratio between the output amplitude at a frequency site and the true amplitude given in the analyzed time series. To give an example, the Fourier analysis of a pure-M2 hourly sampled time series with a length of 3,660 h, generated by Eq. (1) with T = 12.420601, A = 10, and IP = 0, yields an amplitude of approximately 0.27 at the S2 frequency site (at a frequency of 305/3660 cph). Alternatively, the

Table 1. Major and relatively major diurnal and semidiurnal tidal constituents used in this study

Tidal constituents	Frequency (°/h)	Period (h)
Q1 (q)[a]	13.398661	26.868357
O1 (O)[a]	13.943036	25.819342
M1	14.496694	24.833248
P1 (p)[a]	14.958931	24.065890
K1 (K)[a]	15.041069	23.934470
J1	15.585443	23.098477
2N2	27.895355	12.905374
μ2	27.968208	12.871758
N2 (n)[a]	28.439730	12.658348
ν2	28.512583	12.626004
M2 (M)[a]	28.984104	12.420601
L2	29.528479	12.191620
T2	29.958933	12.016449
S2 (S)[a]	30.000000	12.000000
K2 (k)[a]	30.082137	11.967235

[a] Abbreviations in this article for the eight major constituents

S2 output amplitude from the pure-M2 time series is approximately 0.27. The amplitude leakage of the M2 constituent to the S2 site is approximately 2.7% in this case. If the analyzed time series also contains an S2 constituent with an amplitude of 5.0 and the desired isolated constituent is S2, the M2 leakage to the S2 frequency site is the source of the second-category error in the isolation of S2, leading to a possible maximum error in the amplitude of approximately 5.4%. The subsequent section explains why the magnitude of the leakage of the nearby tidal constituent indicates the "possible maximum" error in the output amplitude of the desired isolated constituent.

2. Relationships among amplitude reduction, amplitude leakage, and two categories of error

The Fourier analysis results are often displayed as the frequency spectra of the output amplitudes for consecutive discrete frequencies and the spectra of the output initial phases. Figure 2 shows examples of six sets of such amplitude and initial-phase spectra. The analyzed hourly sampled time series to derive the spectra shown in Figures 2a-d were composed of the two tidal constituents, M2 and S2, with amplitudes of 12 and 10 and initial phases of 1 and 2 rad, respectively. The following formula was used for generating these four time series:

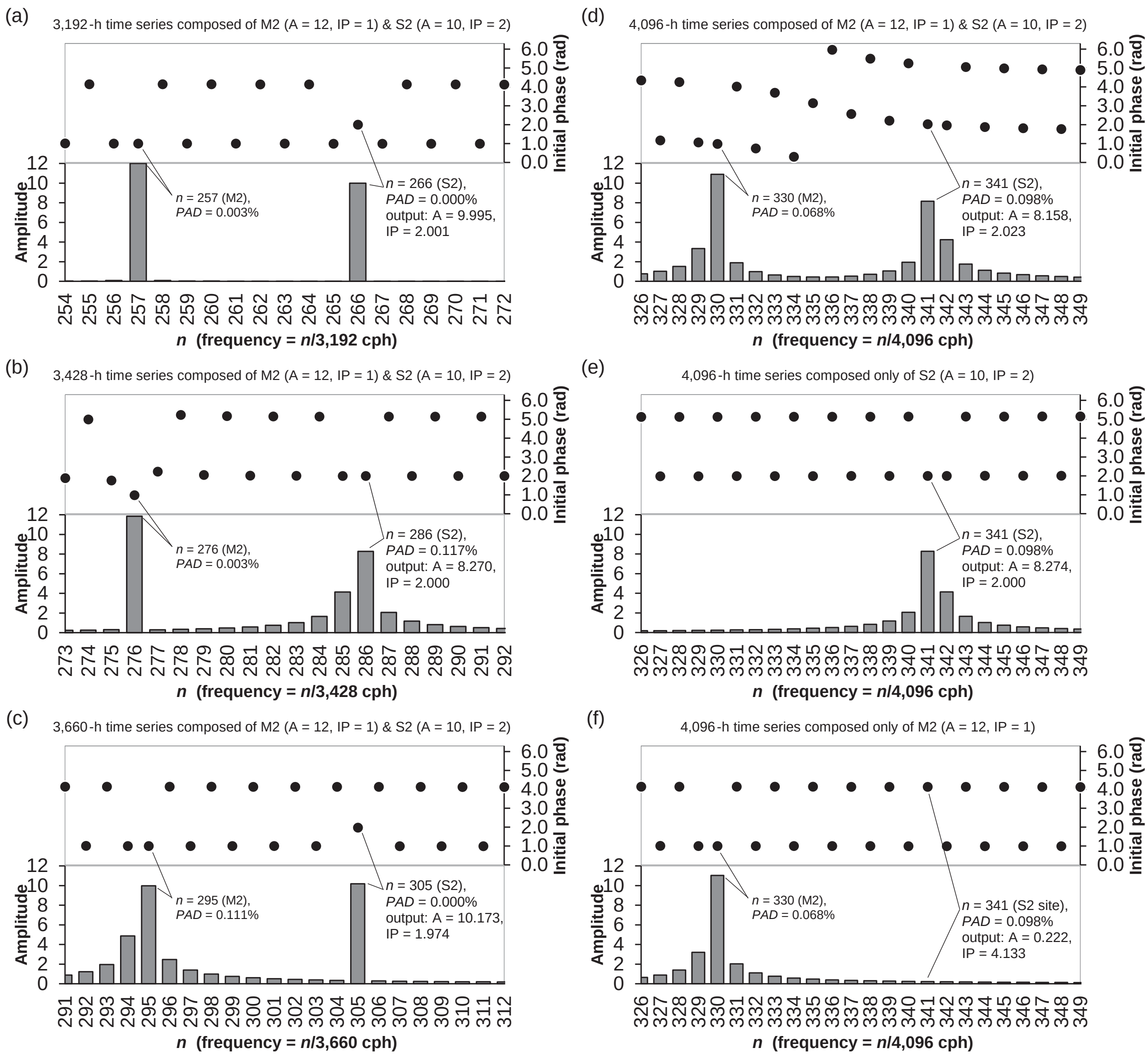

Fig. 2. Pairs of frequency spectra showing Fourier analysis output amplitudes and initial phases

Outputs from (a) 3,192-h, (b) 3,428-h, (c) 3,660-h, and (d) 4,096-h time series composed of M2 and S2; (e) 4,096-h time series composed of S2; and (f) 4,096-h time series composed of M2. The true amplitude and initial phase of each component in the analyzed time series are noted on the top of each pair of spectra.

A: amplitude; *IP*: initial phase; n: frequency indicator or the integer used in the Fourier analysis calculation; *PAD*: period approximation deviation of the tidal constituent

$$f(t) = 12\cdot\sin(2\pi\cdot 12.420601^{-1}\cdot t + 1) + 10\cdot\sin(2\pi\cdot 12^{-1}\cdot t + 2). \quad (3)$$

It is premised here that the desired isolated constituent is S2. Figure 2a shows the Fourier analysis spectra of the time series with an analyzed length of 3,192 h. For this time series length, the PAD for S2 is zero, and the PAD for M2 is also fairly small (0.003%). The outputs for S2 are an amplitude of 9.995 and initial phase of 2.001 rad with small error magnitudes (0.05% and 0.001 rad, respectively). Figure 2b presents the outputs of the Fourier analysis of a 3,428-h-long time series and exemplifies the first-category error in the S2 output. For this time series length, the PAD of S2 is somewhat large (0.117%). Because of the large PAD, the reduction of the output amplitude at the S2 frequency site (i.e., at 286/3,428 cph) accompanied with the spectral leakage of S2 to adjacent frequencies is not negligible. The S2 output amplitude is 8.270 with an error of 17.30%. Figure 2c exemplifies the second-category error in the isolation of S2. The PAD of the S2 constituent for this 3,660-h time series is zero, but the PAD of M2 is slightly large (0.111%). Affected by the leakage component of the M2 constituent at the S2 frequency site (305/3,660 cph), the errors of the S2 output are 1.73% in the amplitude and 0.026 rad in the initial phase. For the Fourier analysis of a 4,096-h time series (Fig. 2d), both categories of error are involved in the S2 output. The two sources of error can be separately represented as in Figures 2e and 2f, which show the results of the Fourier analyses of two time series of the same length 4,096 h with each single constituent S2 or M2. It is noted that the total S2 amplitude error of 18.42% (Fig. 2d) is not the simple addition or subtraction of the amplitude errors observed in Figures 2e and 2f, 17.26% and 2.22%, respectively.

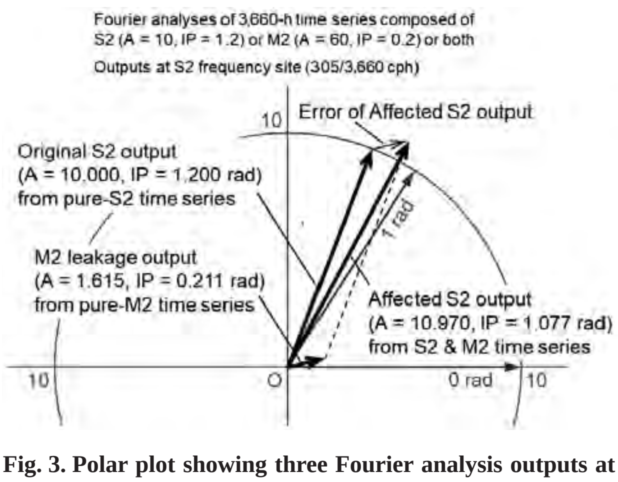

Fig. 3. Polar plot showing three Fourier analysis outputs at the S2 frequency site

Outputs from 3,660-h time series composed of S2 (amplitude: 10, initial phase: 1.2 rad), M2 (amplitude: 60, initial phase: 0.2 rad), or both

A: amplitude; *IP*: initial phase

As our Fourier analysis output for one frequency is composed of the amplitude and initial phase, the output and its error can be represented by vectors in a polar (circular) plot. Figure 3 shows the Fourier analysis outputs of three 3,660-h-long time series at the S2 frequency site. The “original S2 output” vector in Figure 3 represents the Fourier analysis output derived from time series composed only of the S2 constituent with an amplitude of 10 and initial phase of 1.2 rad. This output shows no error. The “affected S2 output” in Figure 3 presents the output from time series composed of the two constituents S2 and M2, the former with the same amplitude and initial phase as in the “original” pure-S2 time series and the latter with an amplitude of 60 and initial phase of 0.2 rad. The subtraction of these two vectors demonstrates the second-category error in the S2 isolation occurring in the Fourier analysis of the time series that include the M2 constituent. The error vector is equivalent to the vector of the output of the other time series composed only of the M2 constituent with an amplitude of 60 and initial phase of 0.2 rad or the leakage component of the M2 constituent at the S2 frequency site (“M2 leakage output” in Fig.3).

This demonstration on a polar plot indicates that the effects of the second-category error on the output amplitude and initial phase of a desired tidal constituent vary with the difference between the initial phase of the (original) desired constituent and that of the leakage component of a nearby affecting constituent. The error of the output amplitude of the affected desired constituent should be maximal when the difference between the two initial phases is either zero or π, with the maximum amplitude error being the ratio between the amplitude of the leakage component of the nearby constituent and the true amplitude of the desired constituent. Given that the amplitude of the leakage component of a nearby constituent is much smaller than the amplitude of a desired constituent at the frequency site of the desired constituent, the affected amplitude error will be nearly zero, and the affected initial-phase error will be near its maximum when the difference between the two initial phases is close to $+\pi/2$ or $-\pi/2$. The expected value of the magnitude of the affected error will be calculated as approximately 0.637 ($=2/\pi$) times the possible maximum error, the multiplier given by the integral from zero to π of the sine function. For example, if the amplitude of the leakage component of a nearby tidal constituent is 1.5% of the amplitude of the (original) desired tidal constituent, the expected value of the error will be a little less than 1.0% for the amplitude and a little less than 0.010 rad for the initial phase. For a specified time series length and a specified nearby affecting constituent, the maximum possible magnitude of the second-category errors in the outputs (amplitude and initial phase) of an affected desired constituent is determined by the magnitude of the leakage from the frequency of the nearby constituent to the frequency site of the desired constituent.

Results

1. Errors due to period approximation deviation of the desired isolated constituent itself

Fourier analyses of hourly sampled time series, each composed of one of the eight major tidal constituents, were performed for varying time series lengths from 600 to 30,000 h. The analysis results demonstrate first-category errors in the outputs of the contained tidal constituent due to the PAD of the same constituent.

Figure 4 shows part of the results: output amplitude errors for six major constituents, O1, P1, K1, N2, M2, and S2, in increasing order of frequency. The magnitudes of the errors are plotted against the PAD of the tidal constituent. The outputs for time series longer than 12,000 h are not shown for a clear presentation.

In Figure 4, for a specified constituent and a limited range of analyzed time series lengths, the data points form a belt shape with approximately straight side edges on a double-logarithmic graph. The slope of the belt is positive (with a value of approximately two), and there is a rough trend that the error magnitude increases as the PAD of the constituent increases. It can also be seen that the error magnitude is generally larger for longer time series for a specified constituent and a fixed PAD. In addition, a comparison of the plots for the six constituents

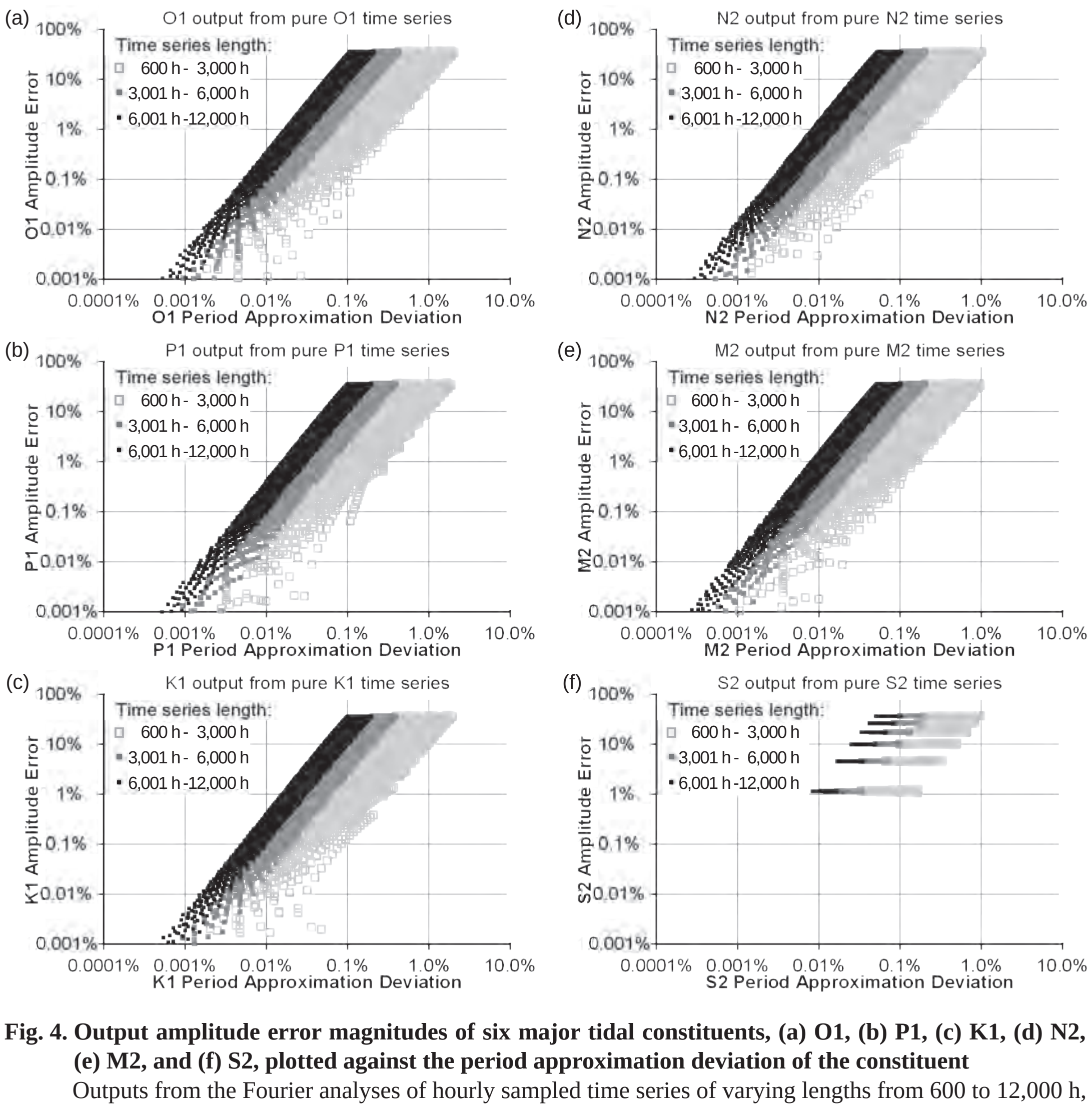

Fig. 4. Output amplitude error magnitudes of six major tidal constituents, (a) O1, (b) P1, (c) K1, (d) N2, (e) M2, and (f) S2, plotted against the period approximation deviation of the constituent

Outputs from the Fourier analyses of hourly sampled time series of varying lengths from 600 to 12,000 h, each composed of a single tidal constituent reveals that generally for a fixed PAD and fixed range of time series lengths, the error magnitude is larger for a constituent of higher frequency or shorter period. These three general tendencies were true for the eight major tidal constituents, including those not shown. In summary, in general, the magnitude of the first-category error in amplitude is positively related to the PAD, the analyzed time series length, and the frequency of the tidal constituent.

After some trials, we found the following calculation formula that yields mostly similar values:

$$E_1 \cdot D^{-2} \cdot L^{-2} \cdot F^{-2} \text{ or} \tag{4}$$

$$E_1 \cdot D^{-2} \cdot L^{-2} \cdot T^{2}, \tag{5}$$

where E_1 is the (first-category) error magnitude of the output amplitude, D is the PAD of the tidal constituent, L is the analyzed time series length (h), F is the exact frequency of the tidal constituent (cph), and T is the exact period of the tidal constituent (h). In cases where the PAD was zero, which occurred for the S2 constituent, the output amplitude error was always zero, and the calculation of the above formula was not performed. With those exceptions, the calculations were made for all outputs derived from time series with lengths 600 h-30,000 h and for the eight major tidal constituents. Figure 5 shows part of the results for lengths up to 12,000 h and for the four constituents O1, K1, M2, and S2.

As shown in Figure 5, most of the calculation results remained close to the value of 1.6 regardless of the PAD, the analyzed time series length, or the constituent period, although the results for small PADs (therefore with small output amplitude errors) became outliers. This concentration and approximate constancy of the values calculated using Eq. (4) or (5) were confirmed for the entire investigated range of the time series lengths from 600 to 30,000 h and for all eight constituents. The first quartiles, medians, and third quartiles (for lengths 600 h-30,000 h) of the values for the eight constituents were in the ranges 1.510-1.533, 1.588-1.594, and 1.623-1.628, respectively. More than 99.5% of the values were < 2.0. We infer that the approximate constancy of the values of Eq. (4) or (5) also holds for time series longer than 30,000 h.

Figure 6 shows examples of output initial-phase errors. The error magnitudes for the P1 and S2 constituents are plotted against the PADs. For both

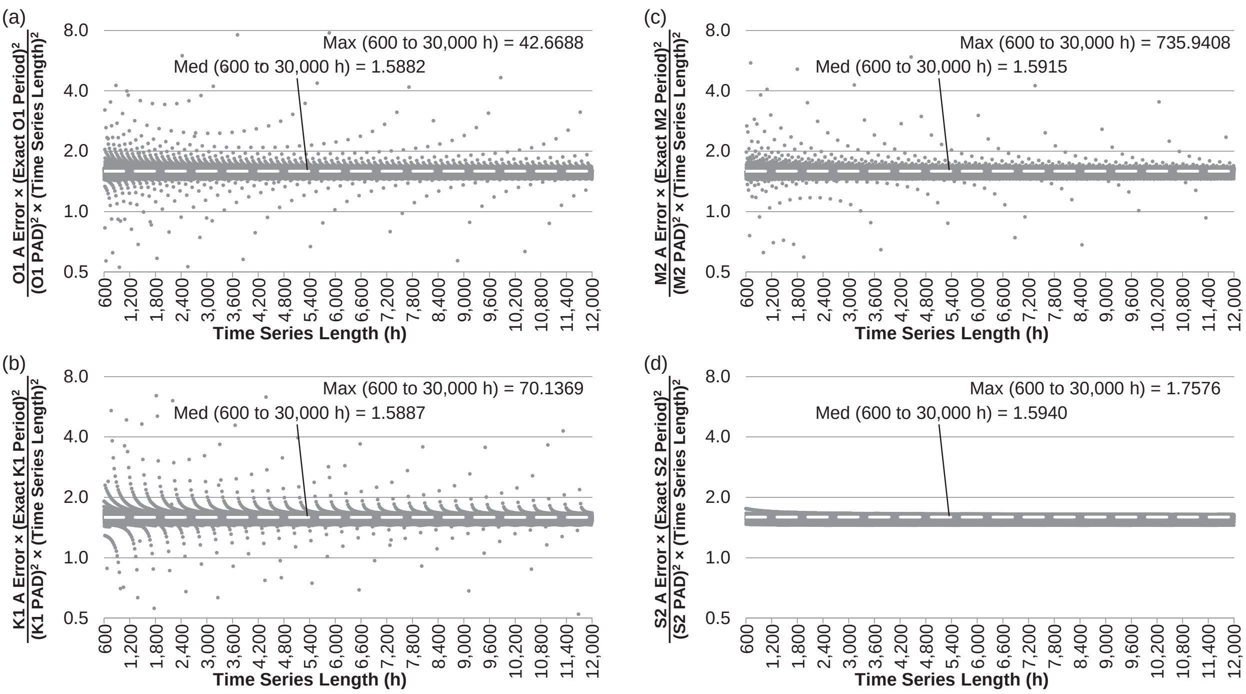

Fig. 5. Values calculated using Eq. (4) or (5) plotted against the analyzed time series length for the (a) O1, (b) K1, (c) M2, and (d) S2 constituents

Values derived from the outputs of the Fourier analyses of hourly sampled time series of varying lengths from 600 to 12,000 h, each composed of a single tidal constituent. Calculations were made up to 30,000 h, but only a portion is shown. *A*: amplitude; *PAD*: period approximation deviation; *Med*: median; Max: maximum constituents, the errors are split into two parts; in one part, they are directly proportional to the PAD, and in the other, they remain zero regardless of the PAD. The errors in the former part were outputs of the time series of an even length and those in the latter part of an odd length. These features of the initial-phase errors were common to all eight major constituents, including those not shown. This shows that the deviation of the time origin of the analyzed time series from its center, when accompanied with a nonzero PAD, led to an initial-phase error. Nevertheless, these initial-phase errors derived from the setting of the time origin in the present Fourier analysis procedure were small for the desired tidal constituents and time series lengths of interest in this study (less than roughly 0.01 rad for the eight major constituents and time series 600 h and longer). The initial-phase errors are not further addressed here.

2. Amplitude leakages due to the period approximation deviation of nearby tidal constituents

To investigate the second-category errors in outputs for one of the eight major constituents, hourly sampled time series, each composed of one of the other 14 major or relatively major tidal constituents, were subjected to Fourier analyses, with analyzed lengths of 600 h-30,000 h.

Figures 7a-c show the amplitude leakages of the Q1, P1, and M2 constituents contained in the analyzed time series to the K1 frequency site, plotted against the PADs of Q1, P1, and M2, respectively. The leakages are the sources of second-category errors in the isolation of the K1 constituent. In each of the three graphs, the leakage is generally positively related to the PAD of the constituent contained in the time series. For each constituent, the slopes of the lines (not shown) connecting the plotted data points and the origin are of course all positive, and their ranges narrow as the time series length becomes longer. The upper limit of a slope, for the range of the time series lengths shown (600 h-12,000 h), is larger for leakages of a constituent that has a frequency closer to the affected constituent K1. For instance, the P1 constituent, the frequency of which is closest to that of K1, has the largest slope upper limit (Fig. 7b).

Figures 7d-f show other examples of amplitude leakages, namely, the leakages of O1, L2, and S2, respectively, as affecting nearby constituents to the frequency site of M2 as the affected constituent, demonstrating the sources of second-category errors in the isolation of the M2 constituent. The same features as described above for the leakages to the K1 site are observed for the leakages to the M2 site.

The ratios between the output amplitude leakage and the PAD of the constituent contained in the time series, which are represented by the slopes of the data points in Figure 7, are plotted in Figure 8 against the analyzed time series length. Figures 8a-d plot the ratios for the leakages of M1, P1, J1, and N2, respectively, to the frequency site of K1. The medians, third quartiles, and maximums for the range from 600 to 30,000 h are depicted and noted in the plot areas. To compute 601-h-wide centered moving averages and other quantities for time series lengths down to 600 h, Fourier analyses with time series lengths as small as 300 h were additionally performed, although the analysis outputs themselves are not included in the plot range of Figure 8. Moving averages and medians were computed to investigate the general trend of the change of the plotted ratio with the time series length. Examples are shown in Figures 8a and b, respectively. In addition, 601-h-wide "moving third quartiles," defined as the third quartiles of the values within the range from 300 h shorter to 300 h longer than the length where the

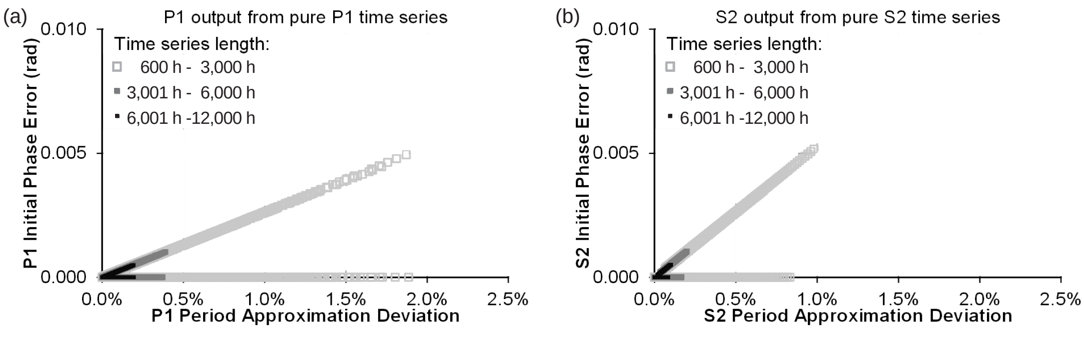

Fig. 6. Output initial-phase error magnitudes of the (a) P1 and (b) S2 constituents, plotted against the period approximation deviation of the constituent

Outputs of the Fourier analyses of hourly sampled time series of varying lengths from 600 to 12,000 h, each composed of a single tidal constituent quartile is defined, were also computed (Fig. 8c). They are used to infer the trend of the third quartile for time series lengths even longer than 30,000 h.

A comparison of Figures 8a-c shows that when the two relevant constituents have closer frequencies, the plotted ratios are generally larger and the maximum of the upper envelope of the plotted ratios occurs with a larger value at a longer time series. Regarding the N2 constituent affecting the K1 site (Fig. 8d), as the frequencies of the two constituents differ much more, the actual maximum occurs at a short time series length outside the plot range with a value larger than the (local) maximum between 600 h and 30,000 h. Among all the 112 combinations of the two relevant constituents, the T2 constituent affecting the S2 frequency site showed the largest maximum of the plotted ratio, with a value of approximately 1,050.3, at the longest time series with a length of approximately 11,500 h. For all combinations of the two relevant constituents, at least within the longer half of the investigated range of lengths 600 h-30,000 h, as the analyzed time series were lengthened, the upper envelope of the plot data decreased and then became nearly constant,

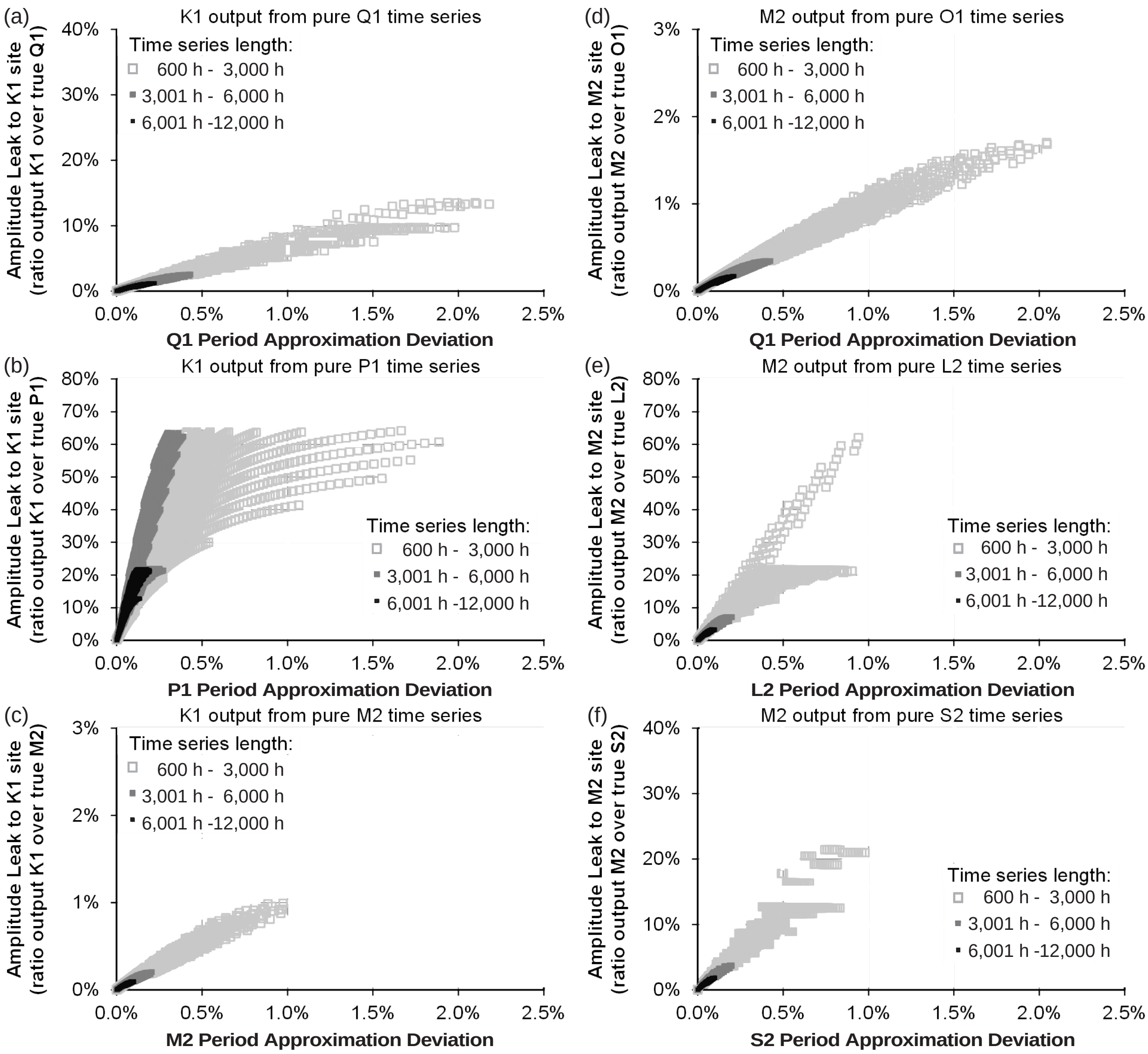

Fig. 7. Amplitude leakages for six combinations of the tidal constituent contained in the analyzed time series and the tidal constituent frequency site affected by the leakage, plotted against the PAD of the contained constituent

The amplitude leakage is the ratio of the output amplitude at the affected frequency site over the true amplitude of the contained constituent given in the analyzed time series. The leakages are for (a) Q1 constituent to K1 site, (b) P1 to K1, (c) M2 to K1, (d) O1 to M2, (e) L2 to M2, and (f) S2 to M2. These are derived from the outputs of the Fourier analyses of hourly sampled time series of varying lengths from 600 to 12,000 h, each composed of one tidal constituent.

and the lower envelope increased and then became nearly constant. For a longer part of the investigated range of time series lengths, the moving third quartile eventually became nearly constant with values almost the same as the third quartile for the entire investigated range. The moving median and moving average for a longer part of the range also became nearly constant with values almost the same as the median and average for the entire investigated range. In addition, we infer from the above investigation that for a range of time series lengths longer than 30,000 h, the maximum of the plotted ratio should be less than the maximum for the investigated range of 600 h-30,000 h, the third quartile should be almost equal to or slightly less than that for the investigated range, and the median and average should be almost equal to those for the investigated range.

Table 2 lists the representative values of the ratios of the amplitude leakage over the PAD for the range of time series lengths 600 h-30,000 h for all combinations of the affecting and affected tidal constituents. Note that the listed values were obtained from the investigations of time series that have a length that is a multiple of 1 h. Investigations of time series with a length that is a multiple of 10 min, for example, would lead to different results at least for the maximums.

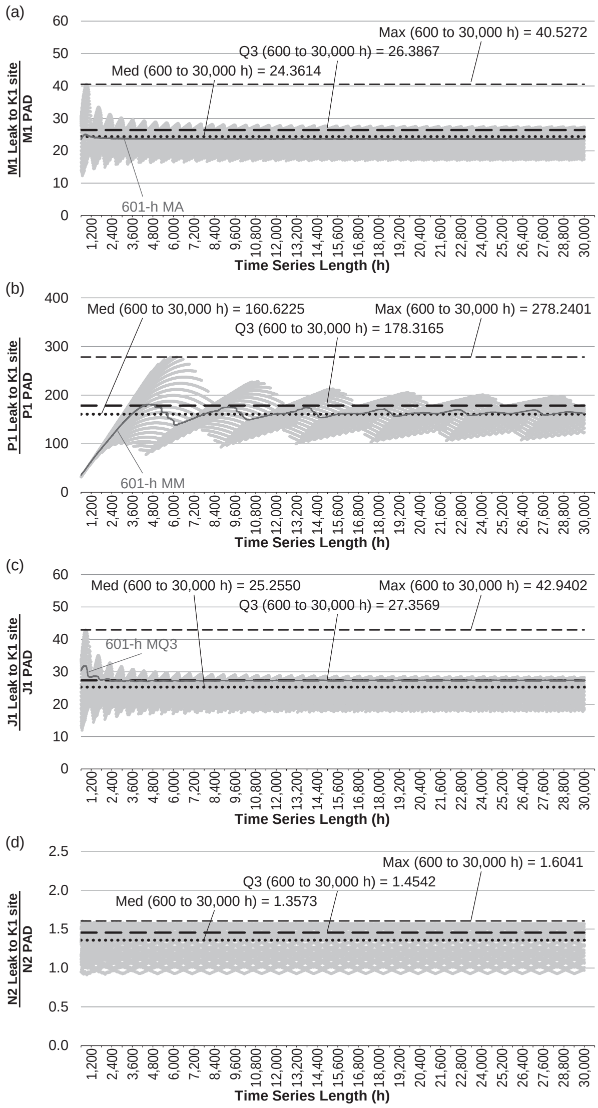

Fig. 8. Ratios of output amplitude leakage over the PAD of the tidal constituent contained in the time series plotted against the analyzed time series length, for leakages of (a) M1 constituent to K1 site, (b) P1 to K1, (c) J1 to K1, and (d) N2 to K1

The ratios are derived from the outputs of the Fourier analyses of hourly sampled time series of varying lengths from 600 to 30,000 h, each composed of one tidal constituent.

PAD: period approximation deviation; *Med*: median; *Q3*: third quartile; *Max*: maximum; *MA*: moving average; *MM*: moving median; *MQ3*: moving third quartile

Discussion

1. Approach to the organization of criteria for PAD of the desired tidal constituent itself

The approximate constancy of the value of Eq. (4) or (5) (exemplified in Fig. 5) can be used to approximate the first-category error in the Fourier analysis output amplitude. The error is approximately predicted as follows:

$$E_{1M} = D^2 \cdot L^2 \cdot T^{-2} \times 1.6 \tag{6}$$

and will very rarely exceed the value calculated by:

$$E_{1X} = D^2 \cdot L^2 \cdot T^{-2} \times 2.0, \tag{7}$$

where E_{1M} and E_{1X} are the approximate prediction of the error and the practical predicted maximum of the error, respectively, defined here by these formulas; D is the PAD of the tidal constituent; L is the analyzed time series length (h); and T is the exact period of the desired tidal constituent (h). For instance, when Fourier analysis is to be performed on a time series containing the M2 constituent with an analyzed length of 720 h, the PAD of the M2 constituent is 0.055%. The amplitude error in the M2 isolation is approximately predicted as 0.16% (= $0.00055^2 \times 720^2 \times 12.420601^{-2} \times 1.6$). The actual output error we already have for this length is 0.141%. For a time series length of 4,096 h, with an M2 PAD of 0.068%, the error is approximately predicted as 8.0%. The actual error is 8.114%.

According to Eqs. (6) and (7) that predict the first-category error, if the PAD of the desired tidal constituent is appropriately limited, the first-category error in the output amplitude can be almost limited. More specifically, if the PAD is limited to less than the value of the formula as:

$$D_{C1} = T \cdot L^{-1} \cdot \alpha, \tag{8}$$

the errors will be less than approximately:

$$E_{1M} = \alpha^2 \times 1.6 \tag{9}$$

and, in almost all cases, will be less than:

$$E_{1X} = \alpha^2 \times 2.0, \tag{10}$$

where D_{C1} is the PAD criterion as a function of T and L, and α is a constant to be determined.

Setting the value of α requires a compromise between allowable error and available data length. If α is set to a tiny value, the Fourier analysis of a time series with a length selected by the criterion will provide isolation of a tidal constituent with a tiny (first-category) error, but the selected time series lengths may be restricted to too few to be practical.

2. Approach to the organization of criteria for the PADs of nearby tidal constituents

The values in Table 2 provide rough predictions of the spectral leakages from the frequencies of the 15 constituents ([A] in Table 2) to the output frequency sites of the eight major constituents ([B] in Table 2). For instance, if Fourier analysis is intended to be performed on an hourly sampled time series that contains the O1 constituent with an amplitude of 10 and if the PAD of the O1 constituent for the analyzed length is 0.050% (e.g., for lengths 2,400 h, 4,800 h, 9,600 h, and 19,200 h), the output leakage component at the Q1 frequency site will have an amplitude of roughly 0.11 (= 10 × 0.050% × 22.540) and will never exceed 0.194 (= 10 × 0.050% × 38.738). The actual output amplitudes at the Q1 site for the above four time series lengths we already have are 0.115, 0.130, 0.116, and 0.100, respectively.

The setting of the PAD criterion for the limitation of the second-category error requires a compromise for a different reason than the reason for the first-category error. In Table 2, three values that can be used to relate the PAD of an affecting nearby constituent to the second-category error in a desired constituent are given. To refer to the maximums in the setting of the PAD criteria is the safest way. However, this choice would make the criteria excessively severe for long time series in the case where

Table 2. Maximum, third quartile, and median of the ratios of the amplitude leakage over the period approximation deviation for the Fourier analyses of hourly sampled time series with lengths of every whole hour from 600 to 30,000 h

		[A] Tidal constituent contained in time series														
		Q1	O1	M1	P1	K1	J1	2N2	μ2	N2	ν2	M2	L2	T2	S2	K2
	Period (h)[a]	26.868357	25.819342	24.833248	24.065890	23.934470	23.098477	12.905374	12.871758	12.658348	12.626004	12.420601	12.191620	12.016449	12.000000	11.967235
[B] Affected tidal constituent																
Q1	Maximum		38.738	15.915	10.809	10.141	7.4738	1.3961	1.3746	1.3540	1.3586	1.3242	1.3000	1.2707	1.2242	1.2742
	Third quartile[b]		24.424	12.349	8.8241	8.4063	6.4215	1.2386	1.2335	1.2019	1.1971	1.1706	1.1456	1.1283	1.0931	1.1234
	Median		22.540	11.411	8.1553	7.7688	5.9364	1.1619	1.1575	1.1294	1.1253	1.0995	1.0714	1.0508	1.0343	1.0446
O1	Maximum	37.490		39.642	18.355	16.494	10.529	1.4717	1.4390	1.4376	1.4175	1.4021	1.3724	1.3369	1.2906	1.3425
	Third quartile[b]	24.422		24.979	13.831	12.833	8.7303	1.3207	1.3153	1.2800	1.2747	1.2418	1.2082	1.1867	1.1618	1.1815
	Median	22.544		23.057	12.782	11.858	8.0683	1.2360	1.2310	1.2000	1.1955	1.1669	1.1355	1.1124	1.0876	1.1058
P1	Maximum	10.462	17.523	48.109		279.30	36.978	1.6464	1.6283	1.5854	1.5830	1.5450	1.5018	1.4666	1.4170	1.4551
	Third quartile[b]	8.8198	13.827	30.985		178.34	23.708	1.4902	1.4833	1.4404	1.4340	1.3941	1.3507	1.3186	1.3010	1.3098
	Median	8.1515	12.777	28.598		160.60	21.892	1.3889	1.3830	1.3447	1.3394	1.3041	1.2662	1.2382	1.1971	1.2302
K1	Maximum	9.8823	16.157	40.527	278.24		42.940	1.6569	1.6484	1.6041	1.5985	1.5627	1.5194	1.4887	1.4301	1.4881
	Third quartile[b]	8.4003	12.827	26.387	178.32		27.357	1.5049	1.4979	1.4542	1.4478	1.4072	1.3633	1.3306	1.3130	1.3215
	Median	7.7639	11.855	24.361	160.62		25.255	1.4020	1.3963	1.3573	1.3518	1.3160	1.2773	1.2492	1.2066	1.2408
N2	Maximum	1.2341	1.3147	1.4030	1.4914	1.4967	1.6056	78.510	90.019		597.62	80.027	33.285	23.501	22.039	20.813
	Third quartile[b]	1.1864	1.2640	1.3486	1.4235	1.4371	1.5323	50.334	58.169		382.53	51.316	25.915	18.712	17.949	17.344
	Median	1.0960	1.1678	1.2469	1.3151	1.3278	1.4158	46.462	53.704		342.65	47.389	23.944	17.296	16.762	16.034
M2	Maximum	1.1976	1.2755	1.3455	1.4376	1.4478	1.5480	32.349	35.338	79.848	92.294		82.243	40.463	36.259	33.644
	Third quartile[b]	1.1523	1.2263	1.3068	1.3776	1.3906	1.4804	25.406	27.265	51.303	59.301		52.291	29.418	27.829	26.182
	Median	1.0646	1.1330	1.2085	1.2726	1.2848	1.3679	23.480	25.196	47.360	54.734		48.271	27.156	25.953	24.192
S2	Maximum	1.1391	1.2082	1.2839	1.3449	1.3555	1.4512	15.201	16.184	21.269	23.290	35.011	90.518	1,050.3		560.55
	Third quartile[b]	1.0942	1.1622	1.2357	1.3002	1.3120	1.3934	13.357	13.855	18.190	19.106	28.191	61.204	719.09		362.31
	Median	1.0111	1.0738	1.1425	1.2012	1.2124	1.2875	12.345	12.805	16.811	17.649	26.047	56.451	610.27		325.96
K2	Maximum	1.1322	1.2039	1.2794	1.3449	1.3450	1.4437	14.573	15.584	20.253	21.684	32.618	79.263	372.58	558.10	
	Third quartile[b]	1.0899	1.1575	1.2305	1.2944	1.3063	1.3870	12.884	13.345	17.326	18.152	26.165	52.359	237.19	352.39	
	Median	1.0069	1.0694	1.1376	1.1958	1.2073	1.2817	11.908	12.338	16.013	16.777	24.176	48.337	216.26	316.74	

* The ratios are calculated by [(output amplitude at [B] frequency site)·(true [A] amplitude)$^{-1}$ / ([A] period approximation deviation)].

[a] Calculated from the frequency lilsted in Kudryavtsev (2004)

[b] Used as “spectral leakage factor” in this study (see text)

the two relevant constituents have close frequencies. The reason is that for two such constituents, the ratio between the leakage and the PAD reaches its maximum at a time series length shorter than approximately 12,000 h and will never approach the maximum again for longer time series (as exemplified in Figs. 8a-c). Such a situation is common for the leakages of diurnal constituent to diurnal and the leakages of semidiurnal constituent to semidiurnal. Using medians or averages (not shown in Table 2) is the simplest way; however, these choices would lead to criteria a little too relaxed for short time series lengths in the case of two constituents with close frequencies. In this study, we make a compromised or moderate choice; we use the third quartiles given in Table 2 as "spectral leakage factors," that is, coefficients for rough quantitative representations of the slopes of the proportionality of the amplitude leakage to the PAD. In the following, the empirically determined spectral leakage factor is used to calculate the representative magnitude of the leakage from the frequency of a tidal constituent to the Fourier analysis output frequency of another constituent. For example, if the O1 constituent has a PAD of 0.050% for an analyzed time series length, we calculate the representative amplitude of its leakage component at Q1 frequency site as 1.22% (= 0.050% × 24.424) of the true O1 amplitude.

To organize practical PAD criteria for the second-category errors, the difference between the amplitudes of the two relevant constituents should be considered. In the case of the last example, the possible maximum error of the Q1 output caused by the calculated representative leakage of O1 is 1.22% in the amplitude and 0.0122 rad in the initial phase, under the condition that the true amplitude of the Q1 constituent in the analyzed time series is equivalent to that of the O1 constituent. Thus, the possible maximum second-category error in the isolation of a tidal constituent affected by another nearby constituent is empirically roughly predicted by:

$$E_2 = A_N \cdot A_D^{-1} \cdot \beta \cdot D, \tag{11}$$

where E_2 is the second-category error; A_N and A_D are the (true) amplitudes of the affecting nearby constituent and the affected desired constituent, respectively; β is the spectral leakage factor we adopted above for the combination of the two constituents; and D is the PAD of the affecting nearby tidal constituent. If the allowable value of the predicted error is set to E_{2A}, the PAD criterion for the affecting tidal constituent (D_{C2}) would be:

$$D_{C2} = A_D \cdot A_N^{-1} \cdot \beta^{-1} \cdot E_{2A}. \tag{12}$$

3. Recommended PAD criteria for the accurate isolation of tidal constituents

In real observation data, the magnitude relationship of the amplitudes of the tidal constituents will differ depending on the observation item, time, and place. In the setting of the criteria below, we use the absolute values of the tidal potentials given by Cartwright & Edden (1973) as representative relative amplitudes of tidal constituents. The potentials are called "cartwright potentials" in oceanographic literature and are used to provide a first guess at the expected relative amplitudes of tidal constituents (Parker 2007). The PAD criterion for the second-category error becomes:

$$D_{C2} = C_D \cdot C_N^{-1} \cdot \beta^{-1} \cdot E_{2A}, \tag{13}$$

where C_D and C_N are the absolute values of the cartwright potentials of the affected desired constituent and the affecting nearby constituent, respectively. The value of β, the spectral leakage factor, is listed in Table 2 as the third quartile. The value of E_{2A}, the allowable predicted error, is left to be determined.

To organize the recommended criteria of the PADs in this study, we set the values of α in Eq. (8) and E_{2A} in Eq. (13) in an inverse way. They are set so that the derived criteria allow a convenient time series length that was conventionally used for tidal analysis (Schureman 1940, Cartwright & Catton 1963, Zetler et al. 1979) as well as already examined and used for the isolation of tidal constituents from groundwater observation data in previous studies (Shirahata et al. 2017, 2018). That length is 8,856 h, which is used for the isolation of the M2, K1, S2, O1, P1, and K2 constituents (in order of the tidal potential). The criteria are also set so that the expected or predicted first- and second-category errors are not extremely different from each other (e.g., with a difference of less than one order of magnitude). Therefore, the value of α is set to 0.05 here. The PAD criterion for the first-category error is given as follows:

$$D_{C1} = T \cdot L^{-1} \times 0.05, \tag{14}$$

where T is the period of the desired isolated constituent and L is the analyzed time series length. With this criterion, the first-category errors that appear in the form of the reduction of the output amplitude will approximately be < 0.4% and, in almost all cases, will be < 0.5%. The allowable value of the predicted second-category error is set to 1.5%, and the PAD criterion for the second-category error is given by:

$$D_{C2} = C_D \cdot C_N^{-1} \cdot \beta^{-1} \times 0.015. \quad (15)$$

The obtained array of the criteria is shown in Table 3. The values in parentheses are the criteria that are actually not used because the PADs of the 15 constituents never exceed the criteria within the time series lengths of a multiple of 1 h and not shorter than 600 h.

Table 4 lists examples of the time series lengths for the accurate isolation of the eight major tidal constituents selected by the criteria. For each of the

Table 3. Recommended criteria (upper limits) of the period approximation deviation of tidal constituents (A) for selecting time series lengths for the accurate isolation of major tidal constituents (B) by Fourier analysis

	[A] Tidal constituents to the period approximation deviations of which criteria are applied														
	Q1	O1	M1	P1	K1	J1	2N2	μ2	N2	ν2	M2	L2	T2	S2	K2
Period (h)[a]	26.868357	25.819342	24.833248	24.065890	23.934470	23.098477	12.905374	12.871758	12.658348	12.626004	12.420601	12.191620	12.016449	12.000000	11.967235
Cartwright potential[b]	0.07217	0.37694	0.02964	0.17543	0.53011	0.02964	0.02301	0.02776	0.17386	0.03302	0.90809	0.02567	0.02476	0.42248	0.11498
[B] Desired isolated tidal constituent															
Q1	**	0.012%	0.296%	0.070%	0.024%	0.569%	(3.798%)	(3.161%)	0.518%	(2.739%)	0.102%	(3.681%)	(3.875%)	0.234%	0.838%
O1	0.321%	**	0.764%	0.233%	0.083%	(2.185%)	(18.605%)	(15.485%)	(2.541%)	(13.433%)	0.501%	(18.230%)	(19.244%)	(1.152%)	(4.162%)
P1	0.413%	0.050%	0.287%	**	0.003%	0.374%	(7.674%)	(6.391%)	(1.051%)	(5.557%)	0.208%	(7.589%)	(8.060%)	0.479%	(1.747%)
K1	1.312%	0.164%	1.017%	0.025%	**	0.981%	(22.964%)	(19.123%)	(3.145%)	(16.633%)	0.622%	(22.722%)	(24.135%)	(1.433%)	(5.233%)
N2	(3.046%)	0.547%	(6.524%)	1.044%	0.342%	(5.742%)	0.225%	0.162%	**	0.021%	0.006%	0.392%	0.563%	0.034%	0.131%
M2	(16.379%)	(2.947%)	(35.167%)	(5.636%)	1.848%	(31.042%)	(2.330%)	(1.800%)	0.153%	0.696%	**	(1.015%)	(1.870%)	0.116%	0.452%
S2	(8.025%)	1.447%	(17.303%)	(2.778%)	0.911%	(15.344%)	(2.062%)	(1.648%)	0.200%	(1.005%)	0.025%	0.403%	0.036%	**	0.015%
K2	(2.193%)	0.395%	(4.729%)	0.760%	0.249%	(4.195%)	0.582%	0.466%	0.057%	0.288%	0.007%	0.128%	0.029%	0.001%	**

** Calculated by [(tidal constituent period)·(time series length)$^{-1}$ × 0.05]

[a] Calculated from the frequency lilsted in Kudryavtsev (2004)

[b] Absolute value of the potential given by Cartwright and Edden (1973)

Table 4. Example list of time series lengths recommended for the accurate isolation of tidal constituents by Fourier analysis

Time series length	Isolated tides	Time series length	Isolated tides	Time series length	Isolated tides	Time series length	Isolated tides	Time series length	Isolated tides
696 h	M	4,212	K	8,532	S	12,445	MKO	19,909	q
708	M	4,235	MO	8,544	KS	13,044	Kp	20,421	q
1,379	M	4,236	MK	8,545	MKO	13,068	Kp	20,580	q
1,391	M	4,260	MKO	8,556	S	13,092	Kp	21,225	q
1,403	M	4,284	K	8,808	KSp	13,116	MKOp	21,226	q
1,404	M	4,285	MO	8,831	MKO	13,140	Kp	21,276	MSO
1,416	M	4,308	K	8,832	KSp	13,164	Kp	21,612	MKSO
1,441	M	4,331	K	8,844	MSk	13,188	Kp	21,636	KSO
1,838	M	4,332	K	8,855	KO	14,483	Mq	21,871	q
2,050	M	4,357	K	8,856	MKSOpk	16,644	MS	21,897	Mq
3,253	O	4,524	K	8,868	MS	16,680	SO	21,948	KSpk
3,254	MO	4,596	MKO	8,880	KS	16,704	SO	21,972	MKSOpk
3,278	O	4,608	Mn	8,904	KS	17,016	MSO	21,996	KSpk
3,279	MO	4,620	MK	9,167	MKO	17,040	SO	22,308	MKSO
3,280	O	4,633	Mn	9,191	MKO	17,340	Sk	25,872	MKSO
3,304	O	5,291	Mn	9,192	KSO	17,352	MKSpk	26,184	KSpk
3,564	O	5,316	Mn	9,204	MS	17,364	MSk	26,208	MKSpk
3,589	O	8,448	KS	9,216	MS	17,376	MKSOpk	26,232	MKSOpk
3,590	MO	8,460	S	9,228	MSn	17,388	Sk	26,256	KSpk
3,899	O	8,472	KS	9,241	Mn	17,688	KSpk	26,544	KSpn
3,900	MO	8,484	S	9,253	Mn	17,700	Sk	26,568	MKSOp
3,925	MO	8,496	MKS	9,266	Mn	17,712	MKSOpk	29,151	MKOq
3,950	MO	8,508	MS	9,291	Mn	17,736	MKSk		
3,975	MO	8,520	MKSO	9,862	MOq	18,048	SO		
4,188	K	8,521	MKO	9,874	Mn	18,433	q		

M: M2; *K*: K1; *S*: S2; *O*: O1; *p*: P1; *n*: N2; *k*: K2; *q*: Q1 eight major constituents, the six combinations of two out of the four major constituents MKSO, and the four combinations of three out of MKSO, the shortest 10 lengths that enable accurate isolations are shown in Table 4. All time series lengths between 600 h and 30,000 h that enable the accurate isolations of four or more tidal constituents are also shown. The two lengths examined and recommended by Shirahata et al. (2017) other than 8,856 h, namely, 708 h for the isolation of M2 and 3,279 h for the isolations of M2 and O1, consistently passed the current criteria. Shirahata et al. (2017) inferred that the time series lengths of the multiples of their recommended lengths 708 h and 3,279 h are also recommendable. The present results showed that it holds to some extent, but for other short time series lengths (e.g., 696 h, 1,379 h, and 3,254 h), the same is not always true. The length of 8,520 h, which has been preferentially used in analyzing tides (Cartwright & Catton 1963, Miyazaki 1967, Ishitobi et al. 1993, Thomson & Emery 2014), is confirmed by the current criteria to enable the accurate isolation of the four major tidal constituents and is the shortest length to do so.

The Fourier analysis of time series longer than a few years may be rare, but the applicability of the criteria should not change for additional longer time series because the approximate stability of the formulas and the values the criteria rest on would apply almost equally well to longer time series. To give some examples, time series lengths 30,468 h and 39,684 h should enable the accurate isolations of the seven constituents MKSOpkq and MKSOpnq, respectively. Time series lengths 70,152 h and 83,292 h should enable the accurate isolations of all the eight major tidal constituents. Because these four time series lengths are multiples of 12 h, the time series data prepared for Fourier analysis can be two-, three-, or four-hourly sampled data, although sparsely sampled time series data will be subject to Fourier analysis output errors caused by random noise (Shirahata et al. 2017). Note that six-hourly sampled data cannot be used for the isolation of the S2 or K2 constituent, because these two tidal constituents have frequencies equal to or higher than the Nyquist frequency, 1/12 cph.

4. Comparison with existing formula for the prediction of spectral leakage

Blair (1979) analytically derived a formula for the approximate prediction of the total error in the Fourier output amplitude of a tidal constituent jointly caused by the amplitude leakages of five nearby constituents. By extracting terms from this formula and modifying the form and variable notations, the following formula that predicts the second-category error owing to one nearby constituent is obtained:

$$E_2 = A_N \cdot A_D^{-1} \cdot \{|[\sin \pi(L / T_D + L / T_N)] / [\pi(L / T_D + L / T_N)]| + |[\sin \pi(L / T_D - L / T_N)] / [\pi(L / T_D - L / T_N)]|\}, \quad (16)$$

where E_2, A_N, and A_D are the same as in Eq. (11); L is the analyzed time series length; and T_D and T_N are the periods of the affected desired tidal constituent and the affecting nearby constituent, respectively. For simple comparison with the results of the present study, assuming $A_N = A_D$, the amplitude leakages of an affecting nearby constituent were calculated as follows:

$$|[\sin \pi(L / T_D + L / T_N)] / [\pi(L / T_D + L / T_N)]| + |[\sin \pi(L / T_D - L / T_N)] / [\pi(L / T_D - L / T_N)]|. \quad (17)$$

These values can be compared with values such as those plotted in Figure 7. The ratios between the amplitude leakage and the PAD of the affecting nearby constituent were calculated by:

$$\{|[\sin \pi(L / T_D + L / T_N)] / [\pi(L / T_D + L / T_N)]| + |[\sin \pi(L / T_D - L / T_N)] / [\pi(L / T_D - L / T_N)]|\}/D, \quad (18)$$

where D is the PAD of the nearby constituent. These values are to be compared with values such as those exemplified in Figure 8.

The curved line in Figure 9a shows an example of amplitude leakage calculated by Eq. (17) based on Blair (1979), the leakage of the N2 constituent to the M2 frequency, for time series lengths from 300 to 3,000 h. The values are plotted against the time series length as in Blair (1979). The dots represent the Fourier analysis results obtained in the present study. They are distributed around the prediction line derived from Blair (1979) but considerably deviate from it. Calculations based on the Blair (1979) formula rarely predict the present outputs accurately. It can be said that, at least for time series longer than approximately 900 h, the fluctuation range of the predictions matches the range of the present outputs. As mentioned in the introduction, the error prediction formula from Blair (1979) was originally derived for cases where the frequency of the desired constituent is exactly represented by an output Fourier frequency. Therefore, the prediction will be applicable if one of the available discrete Fourier frequencies closely approximates the frequency of the desired constituent, M2 in this case, with the PAD of the desired constituent close to zero. This is supported by the distribution of the present outputs for a restricted range of M2 PAD as shown in Figure 9a, which is limited to the vicinity of the prediction line from Blair (1979).

Figure 9b compares the ratio between the N2-to-M2 amplitude leakage and the PAD of the N2 constituent calculated using Eq. (18) based on Blair (1979) and the same ratio obtained via the actual Fourier analyses in the present study. The calculations based on the prediction formula from Blair (1979) yield a wide range of values compared with the actual Fourier analyses. The existing formula should not be of much use for predicting the second-category error owing to a nearby tidal constituent for a time series length that is a multiple of 1 h and generally not a multiple of the periods of the desired tidal constituents.

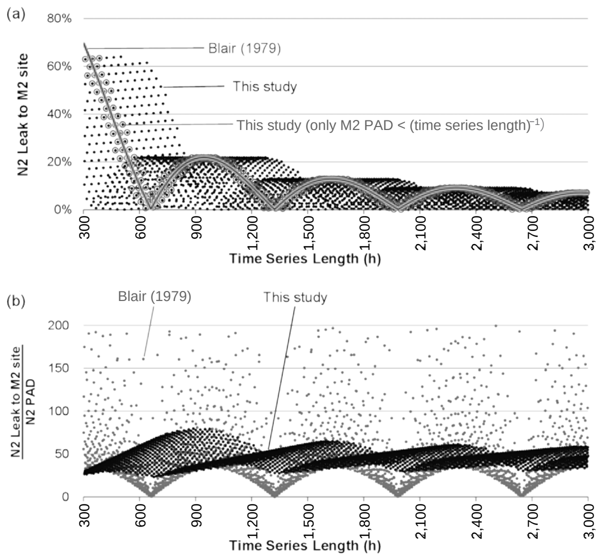

Fig. 9. Comparison between the results of Blair's (1979) error prediction formula and the Fourier analysis of this study

(a) Amplitude leakage of the N2 constituent to the M2 frequency and (b) the ratio of the same amplitude leakage over the PAD of the N2 constituent plotted against the time series length of every hour from 300 to 3,000 h. The predicted values of Blair (1979) in (a) and (b) were calculated using Eqs. (17) and (18), respectively. Note that the Blair (1979) formula was originally intended to be used to predict the leakage at the exact M2 frequency, whereas the present outputs are given at approximate M2 frequencies.

5. Brief comparative demonstration using real observation data

Shirahata et al. (2019b) used hourly sampled groundwater-level time series data to estimate the hydraulic properties of an island aquifer. The observation data that were collected from a site very close to an ocean shore clearly contained tidal components. By the Fourier analyses of the same data with analyzed lengths of 708 h, 3,279 h, and 8,856 h, the amplitudes of the contained tidal components were measured to be approximately 0.44 m-0.47 m for M2, 0.19 m-0.20 m for K1, and 0.17 m-0.18 m for O1 (Shirahata et al. 2014, 2018, 2019a). As a demonstration, Fourier analyses were applied with the 10 lengths from 3,900 to 4,284 h listed in Table 4 to the same observation data starting from 0:00 on August 1, 2008, and the M2, K1, and O1 constituents were isolated following Table 4. For comparison, the Fourier analyses of the same data were also performed with 10 additional lengths that were mechanically set and did not meet the above criteria.

The Fourier analyses with time series lengths selected by the criteria provided consistent amplitudes (Figure 10a), with values in the ranges of 0.459 m-0.464 m for M2, 0.192 m-0.195 m for K1, and 0.173 m-0.176 m for O1. The analyses of the unselected lengths produced inconsistent amplitudes (Fig. 10b), which are unlikely to be accurate.

Concluding remarks

This study investigated errors in the isolation of the eight major semidiurnal and diurnal tidal constituents by the Fourier analysis of finite-length time series.

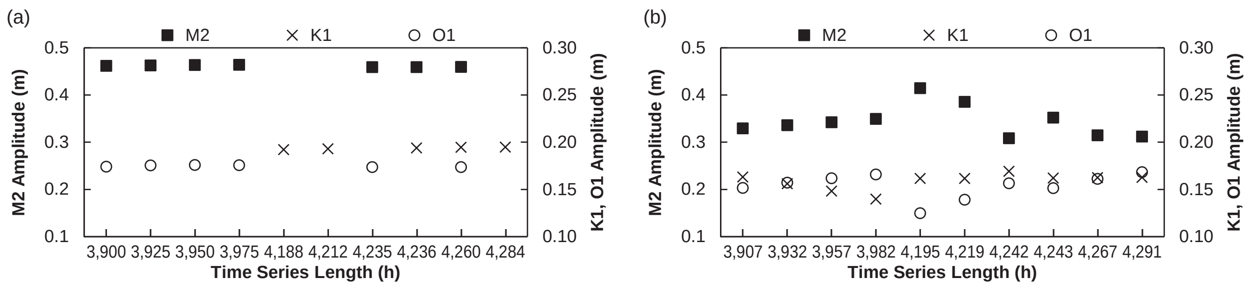

Fig. 10. Amplitudes output from real groundwater-level observation time series data by Fourier analysis (a) with time series lengths selected by the current derived criteria and (b) with unselected lengths

The analyzed observation data were reported by Shirahata et al. (2014, 2018, 2019b).

The investigation included concrete calculations of the amplitudes and initial phases of the eight constituents by the Fourier analyses of 600- to 30,000-h-long time series data each containing one of the 15 major and relatively major tidal constituents. The PAD of a tidal constituent is inevitable in the Fourier analysis of a time series with a length not equal to a multiple of the period of the constituent. The investigation demonstrated that two categories of error are both positively related to the PADs of the tidal constituents, but in different ways. The first-category error due to the PAD of the desired isolated tidal constituent, which results in the reduction of the output amplitude of the desired constituent, can be limited by the appropriate selection of the time series length using a PAD criterion that is a function of the period of the desired tidal constituent and the time series length. The second-category error due to the PAD of a nearby tidal constituent other than the desired constituent, which leads to inaccurate output amplitude and phase, is dependent on the closeness of the frequencies of the two constituents, the magnitude relationship of the two constituents, and the PAD. This error can also be limited if the PAD of the nearby tidal constituent is restricted by a criterion based on the factors of the error empirically quantified in the present investigation. The organized set of PAD criteria can help systematically select time series lengths for the accurate isolation of the eight major semidiurnal and diurnal tidal constituents by Fourier analysis.

The criteria for the PADs of tidal constituents presented in this paper are based on representative relative amplitudes of tidal constituents. If, for a specified observation target, time, and place, the magnitude relationship between the amplitudes of tidal constituents is expected to be significantly different from what we used, another set of criteria may be more appropriate. In that case, the results of the investigations into the factors relevant to spectral leakage will be of practical help for organizing the criteria.

The results of this study enhance the applicability of the simple but accurate Fourier analysis of tidally induced fluctuations in observation data and will contribute to the studies of groundwater resources and aquifers that provide irreplaceable water resources on many remote islands.

Acknowledgements

This study was motivated during the preparation of the first author’s dissertation by a comment made by the author’s main academic advisor, Dr. Tang Changyuan, Professor at Chiba University, Japan. This research was supported by grants from the Project of the NARO Bio-oriented Technology Research Advancement Institution (Research Program on the Development of Innovative Technology).

References

Acworth, R. I. et al. (2015) Understanding connected surface-water/groundwater systems using Fourier analysis of daily and sub-daily head fluctuations. *Hydrogeol. J.*, **23**, 143-159.

Aichi, M. et al. (2011) A new analytical solution of water table response to tidal fluctuations and its application to estimate aquifer properties at the Niijima Island, Japan. *Chikasui gakkaishi* (*J. Groundw. Hydrol.*), **53**, 249-265 [In Japanese with English summary].

Alcolea, A. et al. (2009) Reducing the impact of a desalination plant using stochastic modeling and optimization techniques. *J. Hydrol.*, **365**, 275-288.

Blair, D. (1979) Fourier contamination of a finite length of data in tidal analysis. *Geophys. J. R. Astron. Soc.*, **59**, 523-528.

Burgess, W. G. et al. (2017) Terrestrial water load and groundwater fluctuation in the Bengal Basin. *Sci. Rep.*, **7**, 3872.

Cartwright, D. E. & Catton, D. B. (1963) On the Fourier analysis of tidal observations. *Revue Hydrographique Internationale*, **XL**, 115-128 [In French].

Cartwright, D. E. & Edden, A. C. (1973) Corrected tables of tidal harmonics. *Geophys. J. R. Astron. Soc.*, **33**, 253-264.

Chattopadhyay, P. B. & Singh, V. S. (2013) Hydrochemical evidences: Vulnerability of atoll aquifers in Western Indian Ocean to climate change. *Glob. Planet. Change*, **106**, 123-140.

Dahan, S. (2018) *Rapid review of water knowledge for Pacific Small Islands Developing States.* World Bank, Washington, DC, US.

de Levie, R. (2004) Tidal analysis on a spreadsheet. *Am. J. Phys.*, **72**, 644-651.

Dong, L. et al. (2015) Barometric and tidal-induced aquifer water level fluctuation near the Ariake Sea. *Environ. Monit. Assess.*, **187**, 4187.

Dushaw, B. D. et al. (1995) Barotropic and baroclinic tides in the central North Pacific Ocean determined from long-range reciprocal acoustic transmissions. *J. Phys. Oceanogr.*, **25**, 631-647.

Evans, S. A. (2004) Tidally induced groundwater fluctuations in aquifers located near the Kenai River, Alaska. *Dev. Water Sci.*, **55**, **Part 2**, 1643-1654.

Foreman, M. G. G. et al. (2009) Versatile harmonic tidal analysis: Improvements and applications. *J. Atmos. Ocean. Technol.*, **26**, 806-817.

Fuentes-Arreazola, M. A. et al. (2018) Hydrogeological properties estimation from groundwater level natural fluctuations analysis as a low-cost tool for the Mexicali Valley aquifer. *Water*, **10**, 586.

Ishida, S. et al. (2015) Salt water intrusion into groundwater and problem on Vava’u Island and Lifuka Island, Kingdom of Tonga. *Noson Kogaku Kenkyujo Giho* (*Technical Report of the National Institute for Rural Engineering*), **217**, 1-12 [In Japanese with English summary].

Ishitobi, Y. et al. (1993) Tidal, meteorological and hydrological effects on the water level variation in a lagoon, Lake Shinji.

Jpn. J. Limnol., **54**, 69-79.

Karang, I. W. G. A. et al. (2010) Estimation of tidal energy dissipation and diapycnal diffusivity in the Indonesian Seas. *Int.l J. Remote Sens. Earth Sci.*, **7**, 53-72.

Koizumi, N. (1991) Analyses of tidal fluctuations of groundwater discharge. *J. Phys. Earth*, **39**, 631-647.

Kowalik, Z. & Polyakov, I. (1998) Tides in the Sea of Okhotsk. *J. Phys. Oceanogr.*, **28**, 1389-1409.

Kudryavtsev, S. M. (2004) Improved harmonic development of the Earth tide-generating potential. *J. Geod.*, **77,** 829-838.

Lanyon, J. A. et al. (1982) Observations of shelf waves and bay seiches from tidal and beach groundwater-level records. *Mar. Geol.*, **49**, 23-42.

Miyazaki, M. (1967) A method of Fourier analysis of tides based on the hourly data of 355 days. *Oceanogr. Mag.*, **19**, 7-12.

Möller, O. O. et al. (2007) Tidal frequency dynamics of a southern Brazil coastal lagoon: Choking and short period forced oscillations. *Estuaries Coast.*, **30**, 311-320.

Nishida, H. (1980) Improved tidal charts for the western part of the North Pacific Ocean. *Report of Hydrographic Researches*, **15**, 55-70.

Parker, B. B. (2007) *Tidal analysis and prediction*, US Department Of Commerce, Maryland, US.

Rahi, K. A. & Halihan, T. (2013) Identifying aquifer type in fractured rock aquifers using harmonic analysis. *Ground Water*, **51**, 76-82.

Schureman, P. (1940) *Manual of harmonic analysis and prediction of tides, Department of Commerce Coast and Geodetic Survey Special Publication No. 98, Revised (1940) edition*. US Government Printing Office, Washington, DC, US.

Shirahata, K. et al. (2014) New simple method for estimating hydraulic properties of a freshwater-lens aquifer by analysis of tidal groundwater fluctuations. *Noson Kogaku Kenkyujo Giho* (*Technical Report of the National Institute for Rural Engineering*), **215**, 141-154 [In Japanese with English summary].

Shirahata, K. et al. (2017) Improvements in a simple harmonic analysis of groundwater time series based on error analysis on simulated data of specified lengths. *Paddy Water Environ.*, **15**, 19-36.

Shirahata, K. et al. (2018) Heterogeneous hydraulic properties of an insular aquifer clarified by a tidal response method with simple decomposition techniques. *Geol. Croat.*, **71**, 83-90.

Shirahata, K. et al. (2019a) Estimation of hydraulic parameters of coastal aquifers using tidal response method with frequency decomposition techniques optimized for predominant tides—Application to 147-day groundwater-level data from Tarama Island—. *Nougyo noson kogakukai ronbunshu* (*Irrigation, Drainage and Rural Engineering Journal*), **87**, I_51-I_60 [In Japanese with English summary].

Shirahata, K. et al. (2019b) Study on aquifer hydraulic properties using tidal response method for future groundwater development. *Paddy and Water Environment*, **17**, 531-537.

Siddig, N. A. et al. (2019) Tide and mean sea level trend in the west coast of the Arabian Gulf from tide gauges and multi-missions satellite altimeter. *Oceanologia*, **61**, 401-411.

Tanaka, T. et al. (1967) Tidal analysis by the Fourier transform method. *Sokuchi gakkaishi* (*J. Geod. Soc. Jpn.*), **12**, 77-84 [In Japanese with English summary].

Thomson, R. E. & Emery, W. J. (2014) *Data analysis methods in physical oceanography, Third Edition*. Elsevier, Amsterdam, The Netherlands.

Trefry, M. G. & Bekele, E. (2004) Structural characterization of an island aquifer via tidal methods. *Water Resour. Res.*, **40**, W01505.

Werner, A. D. et al. (2017) Hydrogeology and management of freshwater lenses on atoll islands: Review of current knowledge and research needs. *J. Hydrol.*, **551**, 819-844.

White, I. & Falkland, T. (2010) Management of freshwater lenses on small Pacific islands. *Hydrogeol. J.*, **18**, 227-246.

Whitford, D. J. et al. (2001) Teaching time-series analysis. I. Finite Fourier analysis of ocean waves. *Am. J. Phys.*, **69**, 490-496.

Yoshimoto, S. et al. (2020) Using hydrogeochemical indicators to interpret groundwater flow and geochemical evolution of a freshwater lens on Majuro Atoll, Republic of the Marshall Islands. *Hydrogeol. J.*, **28**, 1053-1075.

Zetler, B. et al. (1979) Some comparisons of response and harmonic tide predictions. *Int. Hydrogr. Rev.*, **LVI**, 105-115.