A Methodology for Fitting the Time Series of Snow Depth on the Arctic Sea Ice

YUZHE WANG AND HUIDONG PAN
Physical Oceanography Laboratory, Qingdao Collaborative Innovation Center of Marine Science and Technology, Ocean University of China, and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

DAOSHENG WANG
College of Marine Science and Technology, China University of Geosciences, Wuhan, and Physical Oceanography Laboratory, Qingdao Collaborative Innovation Center of Marine Science and Technology, Ocean University of China, and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

XIANQING LV
Physical Oceanography Laboratory, Qingdao Collaborative Innovation Center of Marine Science and Technology, Ocean University of China, and Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

(Manuscript received 12 June 2018, in final form 8 May 2019)

ABSTRACT
Snow depth is an important geophysical variable for investigating sea ice and climate change, which can be obtained from satellite data. However, there is a large number of missing data in satellite observations of snow depth. In this study, a methodology, the periodic functions fitting with varying parameter (PFF-VP), is presented to fit the time series of snow depth on Arctic sea ice obtained from the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E). The time-varying parameters are obtained by the independent point (IP) scheme and cubic spline interpolation. The PPF-VP is validated by experiments in which part of the observations are artificially removed and used to compare with the fitting results. Results indicate that the PPF-VP performs better than three traditional fitting methods, with its fitting results closer to observations and with smaller errors. In the practical experiments, the optimal number of IPs can be determined by only considering the fraction of missing data, particularly the length of the longest gaps in the snow-depth time series. All the experimental results indicate that the PPF-VP is a feasible and effective method to fit the time series of snow depth and can provide continuous data of snow depth for further study.

1. Introduction
Once the Arctic sea ice is formed, snow begins to accumulate on the ice surface and thickens throughout the winter (Guerreiro et al. 2017). The ice is mostly free of snow during August. The snow accumulates rapidly in September and October, moderately in November, very slowly in December and January, and then moderately again from February to May (Warren et al. 1999). During the warmer season, the rising temperature and the deformation and drift of sea ice will lead to the melting of snow. In most of the area, snow melts completely before the end of the summer (Lindsay 1998).

Snow depth on sea ice is an important geophysical variable, which has an important influence on the sea ice research. Quantitative estimates of snow depth on the sea ice are needed to provide the necessary parameters for the sea ice model (Häkkinen 1995). Moreover, the state, nature, and distribution of snow have a very important influence on the heat exchange between the ice layer and the atmosphere (Blazey et al. 2013; Pedersen and Winther 2005; Sturm et al. 2006). During winter, snow acts as an insulator that limits the heat transfer from the ocean through the ice to the overlying atmosphere and reduces the rate of sea ice growth (Maykut 1978). During summer, snow reflects up to 80% or more
of the shortwave solar radiation and delays the melting of sea ice (Barry 1996; König et al. 2001). Further, snow can turn to bare ice and melt ponds that decrease the albedo by a factor of up to 5, leading to acceleration of sea ice melting (Perovich et al. 2002; Perovich and Polashenski 2012). Thereafter, the ponds drain into seawater and produce freshwater lenses that affect the ocean circulation (Eicken 1994). Moreover, snow depth on sea ice is also a key parameter affecting the sea ice thickness and volume retrieved by satellite altimetry (Webster et al. 2014).

The snow is also related to the climate of the sea ice region. Snow feedback is expected to amplify global warming caused by increasing concentrations of atmospheric greenhouse gases (Cess et al. 1991). The large air content of snow causes it to be a poor conductor of heat and also causes solar photons to undergo numerous scattering events so that penetration of light is impeded, resulting in a high albedo (Grenfell and Maykut 1977), thereby favoring lower temperatures. Snow thermally isolates the atmosphere from the ground, thus affecting cold extremes of surface air temperature (Räisänen and Eklund 2012).

However, the fact that snow-depth data on sea ice are typically insufficient may cause bias when investigating the sea ice and climate. Therefore, researchers have made efforts to obtain snow depth from limited observations. Warren et al. (1999) carried out a fairly comprehensive study of the Arctic and spatially interpolated snow-depth data from Soviet drifting stations and aircraft landings collected in 1937 and from 1954 to 1991, using a two-dimensional quadratic function for a particular month, irrespective of the year. Webster et al. (2014) spatially interpolated the spring snow-depth data from 2009 to 2013, following the same gridding approach of Warren. The limitations of this spatial interpolation are lack of interannual variability and that snow density and densities have unrealistic north-to-south gradients (Kern et al. 2015; Kwok et al. 2011). At present, there are few methodologies for interpolating or fitting the time series of snow depth. In this paper, a methodology, the periodic functions fitting with varying parameters (PFF-VP), is presented to fit the time series of snow depth on Arctic sea ice, and applied to the grid cells where missing data are less than 40%. The methodology originates from enhanced harmonic analysis to obtain the time-varying amplitudes and phases of tidal components with fixed frequencies. Jin et al. (2018) applied this algorithm to the internal tidal currents in the South China Sea and Pan et al. (2018) applied this algorithm to the river tide and presented a MATLAB toolbox, S_TIDE. In PFF-VP, the period is fixed to 1 year, making the algorithm applicable to Arctic snow-depth data.

This article is organized as follows. The data are introduced in the section 2. In section 3, the fitting methods are introduced. In section 4, the PPF-VP is verified by the experiments where only a subset of observations is used. In section 5, the optimal number of independent points (IPs) is studied and the results of the practical experiments are displayed. Section 6 provides the summary and main conclusions.

2. Data description

The Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E) level-3 daily snow-depth product AE_SI12 (Cavaliere et al. 2003) from 22 July 2002 to 21 July 2011 is obtained from the National Snow and Ice Data Center (NSIDC; https://nsidc.org/node/29078). The snow-depth data in this level-3 gridded product are 5-day running averages, excluding Arctic perennial ice regions. The total number of the observations is 3287 at each grid cell. The spatial resolution of the data is 12.5 km. However, because of the limited penetration depth of microwaves, the effective range of snow depth for AMSR-E data inversion is 0–50 cm. The data availability is shown in Fig. 1a. It can be seen that there is a large area with missing data, and the observations can cover 50% at about 70°N. The grid cells with missing rate smaller than 40% (a total of 7582 grid cells with at least 1973 observations, which is shown in Fig. 1b) are selected for experiments, and the average missing rate is 36.12%.

3. Method

a. PFF-VP

Given that there are many gaps in the time series of snow depth, continuous data can only be obtained by using a fitting method. As discussed in the introduction, the periodicity of snow depth on the Arctic is obvious, and the period is about 1 year. Therefore, we decided to use the periodic function to fit the time series:

\[ h(t) = h_0 + H \cos(\alpha t - \theta), \]

where \( h(t) \) is the measurement of snow depth at time \( t \), \( h_0 \) is a constant denoting the background value, \( H \) is the amplitude, \( \theta \) is the phase, \( \alpha = 2\pi/T \) is the angular velocity, and \( T = 365 \) days is the period.

However, although the periodicity of snow depth is obvious, the snow depth varies from year to year. Therefore, the difference in annual snow depth cannot be well reflected by constant amplitude \( H \) and phase \( \theta \). Therefore, the proposed method assumes that \( H \) and \( \theta \) change with time. Equation (1) can be rewritten into the following form:
interpolation; that is, those at the other points are obtained by cubic spline interpolation. The cubic spline interpolation coefficients are shown in Appendix A. (Lu and Zhang 2006; Guo et al. 2017; Pan et al. 2017; Zhang et al. 2018). Assume that \( N \) is the length of the entire time series and select \( K \) time points that are uniformly distributed in the time domain as IPs. The location of the \( i \)th IP is the same as the \( p(i) \)th point in the time series; that is, \( p(1) = 1 \), \( p(K) = N \), and \( p(i) = (i - 1)\Delta t + 1 \), \( i = 2, 3, \ldots, K - 1 \), where \( \Delta t = (N - 1)/(K - 1) \) is the interval between two adjacent IPs. As in Jin et al. (2018), the cubic spline interpolation is applied to the IP scheme. The parameters \( A[p(i)] \) and \( B[p(i)] \) at IPs are selected as independent parameters and those at the other points are obtained by cubic spline interpolation; that is,

\[
A(t) = \sum_{i=1}^{K} f_{i,t} A[p(i)], \quad 1 \leq t \leq N, \tag{4}
\]

\[
B(t) = \sum_{i=1}^{K} f_{i,t} B[p(i)], \quad 1 \leq t \leq N, \tag{5}
\]

where \( f_{i,t} \) is the interpolation coefficient for the \( i \)th IP at time \( t \). For brevity, the processes for calculating the cubic spline interpolation coefficients are shown in Appendix A.

The reason why the cubic spline interpolation is applied in the method is that it retains the advantages of the piecewise low-order interpolation polynomial and improves the smoothness of the interpolation function. At present, because of stability and smoothness, the cubic spline interpolation is a useful technique, which has become an extremely important branch of numerical approximation and has been increasingly used in many fields (Jiang et al. 2018; Zong et al. 2018).

Note that the values of the independent parameters \( A[p(i)] \) and \( B[p(i)] \), \( i = 1, 2, \ldots, K \), are unknown, but the link between them has been established by cubic spline interpolation. With Eqs. (4) and (5), Eq. (3) can be converted into

\[
h(t) = h_0 + \sum_{i=1}^{K} C_{1,i} A[p(i)] + \sum_{i=1}^{K} C_{2,i} B[p(i)], \tag{6}
\]

where \( C_{1,i} = f_{i,t} \cos(\sigma t) \) and \( C_{2,i} = f_{i,t} \sin(\sigma t) \).

Equation (6) has \( 2K + 1 \) unknown parameters, \( A[p(i)] \) and \( B[p(i)] \), \( i = 1, 2, \ldots, K \), and \( h_0 \). With sufficient observations \( h(t), t = 1, \ldots, N \) and \( N \gg 2K + 1 \), Eq. (6) can be solved by the least squares method.

b. Other fitting methods

1) PERIODIC FUNCTION FITTING WITH CONSTANT PARAMETER

Equation (2), with constant amplitude and phase, can also be used to fit with the least squares method, which is marked as periodic function fitting with constant parameter (PFF-CP).

2) QUADRATIC FUNCTION FITTING

Because the periodicity of snow depth in the Arctic is obvious, we can take the piecewise fitting by least squares method for each cycle. Because of the insulation...
effect of the atmosphere, the temperature continues to rise gradually after the sun’s zenith point reaches the Tropic of Cancer. At the end of July and early August of each year, when the temperature reached the maximum, there was almost no snow on the sea ice. Therefore, during this period, the moment when the snow depth is the lowest was chosen as a partition point, and the data are divided into nine segments (one for each winter). In each segment, the data are fitted with a quadratic function:

\[ h_i(t) = a_{i1}t^2 + a_{i2}t + a_{i3}, \quad i = 1, 2, \ldots, 9, \]  

(7)

where \( a_{i1}, a_{i2}, \) and \( a_{i3} \) are constant coefficients in the \( i \)th segment.

Let the above formula meet

\[ h(t) = 0, \quad t = 1 \quad \text{or} \quad t = N_i, \]

where \( N_i \) is the length of the period in the \( i \)th segment. Thereafter, Eq. (7) can be rewritten as

\[ h(t) = a_{i1}(t-1)(t-N_i). \]  

(8)

The coefficient \( a_{i1} \) can be estimated via least squares fitting. The fitting results of each segment are combined to obtain the fitted snow depth that changes with time.

3) GAUSSIAN FUNCTION FITTING

Considering that the snow always melts faster when temperature rises, with the same segmentation method as in quadratic function fitting (QFF), the data are fitted with the Gaussian function in each segment; that is,

\[ h_i(t) = b_{1i} \exp \left[ -\frac{(t-N_i/2)^2}{2b_{2i}^2} \right], \quad i = 1, 2, \ldots, 9, \]  

(9)

where \( b_{1i} \) and \( b_{2i} \) are constant coefficients in the \( i \)th segment. The coefficients can be solved by nonlinear least squares regression. Because parameter \( b_{1i} \) is related to the largest fitting value in each year, the initial value of \( b_{1i} \) is set as the largest value of snow depth in each year. Parameter \( b_{2i} \) is related to the length of the time domain, and the initial value of \( b_{2i} \) is uniformly set to 800. The fitting results of each segment are combined to obtain the fitted snow depth that changes with time.

4. Experiments

a. Experimental settings

Twelve time series that have the least fraction of missing data among all grid cells were selected for the experiments. Their missing rates are up to 4%–9%, but this has little influence on the experiments. The locations of the 12 grid cells are shown in Fig. 1b. The 12 grid cells are marked as grid cells 1–12 according to the missing rate. The experiments are carried out according to the following steps: 1) For the time series at each grid cell, we introduce artificial gaps to them, trying to mimic the real gaps. Fortunately, other grid cells provide a variety of patterns of missing data. Therefore, we can introduce these kinds of gaps to the time series at the 12 grid cells. In other words, if the data at other grid cells are missing at time \( t \), the data at the 12 grid cells will be removed at time \( t \). 2) These “thinned” time series are fitted by using the four methods introduced in this article, respectively. 3) By calculating the mean absolute errors (MAEs) between the fitting results and the data removed in the artificial gaps, the accuracy of the four methods can be evaluated. 4) Analyze the fitting results with representative time series and representative missing data patterns. 5) Analyze the fitting results of all the missing data patterns statistically at the 12 grid cells.

b. Results analysis

Figures 2 and 3 show the results and MAEs by using the four fitting methods (at grid cells 1 and 8) with two different missing data patterns, respectively. These two time series are extremely representative. The snow at grid cell 1 (Fig. 2) accumulates and melts at a fast speed, which results in snow depth remaining close to its maximum and minimum values for a long time in each year. There is a short process of growth and decline of snow depth. Therefore, the shape of the snow-depth curve in each year looks “fat.” The reason is that there is little snowfall in winter, which makes the snow depth stay invariant for quite a long time. The curve shape is closely related to the location of grid cells. At the locations near the North Pole, water vapor becomes scarce in winter, resulting in the slow accumulation of snow and the “fat” shape. Because of the influence of wind, the time series of snow depth is more complex. Nevertheless, in many areas, the periods of snow accumulation and melting are longer, such as the results shown in Fig. 3. This time series has obvious processes of accumulation and melting of snow. Soon after the snow depth reaches its peak value, it begins to melt quickly. There is a long process of growth and decline of snow depth in each year, making the shape of the snow-depth curve “thin.” (The fitting results for all the 12 time series with artificial gaps are shown in Figs. B1 and B2 in appendix B.)

In the PFF-VP, the selection of the IP number \( K \) can vary. In reality, snow-depth data for almost all grid cells are incomplete, and the incompleteness of data may
have an impact on IP number. The missing parts of real data at each grid cell are different. To find the optimal $K$ value, $K$ is varied from 3 to 50. In Figs. 2a, 2b, 3a, and 3b, the yellow line represents the MAEs corresponding to different IP numbers. As a comparison, the MAEs fitted by the other three methods are given. It can be seen that before the IP number reaches about 20, the MAEs remain almost invariant as the IP number increases. When the IP number is more than 40, the MAEs will become very large and grow rapidly. When the IP number is between 20 and 40, the MAEs are smaller. In this case, we choose the IP number leading to the minimum MAE as the optimal IP number for the corresponding time series.

![Fig. 2. (a) MAEs of (c) fitting results at gridcell 1 using the four methods. (b),(d) As in (a) and (c), but for another missing data pattern.](image)

![Fig. 3. As in Fig. 2, but at gridcell 8.](image)
It is reasonable that larger IP number can lead to worse fitting results. The reason is that the interpolation result at non-IPs always depends more on the value of the parameters (amplitude and phase) at the IP that is nearest to those non-IPs, regardless of the interpolation method. This is the nature of interpolation. If data coverage is severely incomplete in a certain interval, it may lead to poor fitting results in this interval. Since the IPs are selected uniformly in the time domain, it is obvious that the increase of the IP number will reduce the length of the interval and increase the probability of poor fitting.

With the optimal IP number, Figs. 2c, 2d, 3c, and 3d show the results of the four fitting methods. Green dots represent the artificial gaps. We can see that continuous periods of missing data are a common phenomenon in satellite observation. Similarly, the missing data patterns shown in Figs. 2 and 3 are also representative. In Fig. 2c, there is a continuous period of missing data in spring and winter from January 2003 to April 2003, and in autumn and winter from November 2006 to January 2007. In Fig. 2d, there are missing data in summer and some discrete periods of missing data. In Fig. 3c, there is a wide gap near the highest snow depth from November 2004 to May 2005 and from November 2006 to February 2007. In Fig. 3d, there are mainly some short continuous periods of missing data in winter and wide gaps in summer. The percentages of missing data shown in Figs. 2c, 2d, 3c, and 3d are 25.12%, 33.99%, 23.42%, and 31.51%.

We see from Fig. 2 that PFF-CP has a poor fitting performance, which is greatly influenced by the sharp interannual variation of amplitude. The fitted amplitude can only reflect the mean snow depth, which cannot show the specific circumstance in each year. The fitting results have an MAE of about 6 cm at gridcell 8 shown in Table 1. However, PFF-VP causes a smaller MAE at gridcell 1 where little interannual variation exists.

QFF takes the adjacent lowest point as the start and end of each subdomain, and assumes that the lowest snow depth is 0 cm, which can ensure that the shape of the fitting results is consistent with that of the prescribed data. QFF obtains different fitting coefficients according to different circumstances of the year. However, the two fixed ends make the shape of the quadratic function inflexible; therefore, it cannot be used to describe the accumulation and melting of snow with different speeds. As we can see in the Fig. 2 from August 2010 to August 2011, because of the slow accumulation and the fast melting of snow, QFF cannot fit the asymmetrical pattern. In addition, the maximum point of the fitting curve is obviously lower than the highest point of the prescribed data from Fig. 2. Compared to PFF-CP, the MAE of QFF is not reduced in Table 1.

### Table 1. MAEs (cm) of missing data using the four methods at two grid cells in two missing data patterns.

<table>
<thead>
<tr>
<th></th>
<th>Gridcell 1</th>
<th>Gridcell 8</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pattern 1</td>
<td>Pattern 2</td>
</tr>
<tr>
<td>PFF-CP</td>
<td>3.41</td>
<td>3.56</td>
</tr>
<tr>
<td>QFF</td>
<td>3.57</td>
<td>3.33</td>
</tr>
<tr>
<td>GFF</td>
<td>3.46</td>
<td>3.23</td>
</tr>
<tr>
<td>PFF-VP</td>
<td>2.60</td>
<td>2.25</td>
</tr>
</tbody>
</table>

Gaussian function fitting (GFF), like QFF, fits the data in each year, respectively, and then combines fitting curve to a whole. In Figs. 2 and 3, the fitting results by GFF look better than those by QFF. No matter how fast or slowly the snow depth rises (or falls), the rate of increase (or decrease) can be perfectly fitted, but GFF has almost the same two shortcomings as QFF. First, because nine years are divided into nine segments, the continuity and smoothness between segments are poor. Second, the two methods are heavily influenced by data integrity. As can be seen from Fig. 3c, at the end of 2004, the fitted curve by the QFF is much lower than that of prescribed data, and the curve by the GFF is much higher than the prescribed data because of missing data during the maximum value of snow depth. In Fig. 2d, at the beginning of 2011, because of data missing during the whole snowmelt period, the shape of snow depth is far from the Gaussian function, resulting in obvious errors in fitting results. In Table 1, the MAE by GFF is very large in gridcell 8 under pattern 1. Similarly, the MAEs of other methods also become slightly larger, mainly because measurements are lacking during a critical period, (i.e., winter).

After getting the optimal IP numbers, PFF-VP can effectively solve the problems arising from the other methods, because the time-varying amplitude and phase can be adjusted according to the actual situation, obtaining a fitted curve closest to reality. As we can see from Fig. 2c, at the end of 2002, PFF-VP almost perfectly fitted the continuous missing parts in winter, and the fitting results are satisfactory in summer with more missing data but smaller snow depth. However, PFF-VP is also affected by integrity. At the end of 2004 and the end of 2006 in Fig. 3c, with missing data in a critical period, the shape of PFF-VP fitting results is somewhat deviated, but it can roughly reach the highest value of the snow depth, which is acceptable. The MAE by PFF-VP is nearly a half of those by the other methods (Table 1).

In this paper, we only give the fitting results at two grid cells in two missing data patterns. Because there are a large number of missing data patterns, the fitting results in each pattern cannot be displayed one by one,
so a statistical result is important. There are more than 7000 other grid cells. That is, there are more than 7000 missing data patterns. For gridcell 1, 7000 “thinned” time series are constructed by introducing the gaps to them. After fitting these time series, more than 7000 fitting curves are obtained. The fitting curves are averaged and the standard deviations (STDs) for each time are calculated. The gridcell 8 is treated in the same way. Figure 4 shows the statistical results. As we can see, mean fitting results are in good agreement with observations. The STD represents the fluctuation range of these fitting results, which basically covers observation. In Fig. 4b, mean fitting result is a little larger than the observation in the summer of 2004, probably because of many missing data in this period. The STD in summer of 2006 and 2011 is very large because of missing data in these periods. Generally speaking, the STD in summer and winter is larger than that in other periods, because the snow depth is larger and irregular in winter, and more data are absent in summer. Even though the value of snow depth in summer is small, the STD in this period is large, which shows that the fraction of missing data has an important impact on the stability of fitting. In spring and autumn with accumulation and melting of snow, the STD is smaller, which shows better fitting results in these parts. The statistical MAEs at the 12 grid cells with all missing data patterns by the four methods are given in Table 2. The average MAE of PFF-VP (i.e., average over the 12 grid cells) is 55% smaller than that of PFF-CP, 61% smaller than that of QFF and 53% smaller than that of GFF, indicating that PFF-VP is an effective method for various shapes of data and different gap structures.

5. Discussion

For seriously incomplete time series, we are not able to evaluate the accuracy of these time series at those points of time that were missing and lose the way to get the optimal IP number, but there is still a way to get the approximate optimal IP number. The selection of the optimal IP number is related to two factors. One is the location of the gaps and the fraction of missing data and the other is the changes of values of snow-depth data. Figure 3b shows that MAEs are nearly constant near the optimal IP number, while the MAEs near the optimal IP number have

### Table 2. MAEs (cm) of missing data using the four methods at 12 grid cells of all missing data patterns. PFF-VP1 indicates that the optimal IP number is used and PFF-VP2 indicates that Eq. (10) is used.

<table>
<thead>
<tr>
<th>Grid cell</th>
<th>PFF-CP</th>
<th>QFF</th>
<th>GFF</th>
<th>PFF-VP1</th>
<th>PFF-VP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.99</td>
<td>3.15</td>
<td>2.51</td>
<td>1.37</td>
<td>2.12</td>
</tr>
<tr>
<td>2</td>
<td>2.37</td>
<td>2.31</td>
<td>2.06</td>
<td>1.14</td>
<td>1.55</td>
</tr>
<tr>
<td>3</td>
<td>3.52</td>
<td>3.34</td>
<td>3.05</td>
<td>1.75</td>
<td>2.45</td>
</tr>
<tr>
<td>4</td>
<td>1.21</td>
<td>2.05</td>
<td>1.51</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>5</td>
<td>3.56</td>
<td>2.82</td>
<td>3.26</td>
<td>1.61</td>
<td>2.88</td>
</tr>
<tr>
<td>6</td>
<td>1.36</td>
<td>2.38</td>
<td>1.18</td>
<td>0.63</td>
<td>0.83</td>
</tr>
<tr>
<td>7</td>
<td>2.99</td>
<td>2.76</td>
<td>2.57</td>
<td>1.24</td>
<td>1.94</td>
</tr>
<tr>
<td>8</td>
<td>4.17</td>
<td>5.12</td>
<td>4.57</td>
<td>1.42</td>
<td>3.45</td>
</tr>
<tr>
<td>9</td>
<td>4.51</td>
<td>5.63</td>
<td>4.98</td>
<td>1.39</td>
<td>2.97</td>
</tr>
<tr>
<td>10</td>
<td>2.25</td>
<td>3.53</td>
<td>2.82</td>
<td>1.36</td>
<td>1.78</td>
</tr>
<tr>
<td>11</td>
<td>3.04</td>
<td>2.80</td>
<td>1.95</td>
<td>1.02</td>
<td>1.30</td>
</tr>
<tr>
<td>12</td>
<td>1.53</td>
<td>2.50</td>
<td>1.79</td>
<td>1.18</td>
<td>1.22</td>
</tr>
</tbody>
</table>
obvious fluctuation in Fig. 3a. The reason is that a large amount of continuous data are missing in critical periods (especially in winter 2002) in the time series in Fig. 3c.

Because of the obvious periodicity of snow depth at each grid cell, the impact of the changes of values of snow-depth data is far less than the fraction of missing data on the selection of optimal IP number. Here we discuss the selection of the optimal IP number by using a method to eliminate the impact of the changes of values and only consider the characteristics of the data gaps. Figure 5 shows two time series at gridcell A and B (marked in Fig. 1b) with missing degree of 30.24% and 40%, respectively, both of which have continuous and discrete gaps, covering every season. More importantly, gridcell B has the largest missing rate among all the grid cells, which is expected to deteriorate the fitting results. Next, we discuss the selection of the optimal IP number at the two grid cells.

Twelve relatively complete time series were selected in the experiments in section 5. Now we introduce the gaps according to Figs. 5a and 5b to the 12 time series to get 12 different time series with almost the same gap structure, respectively. Then each time series is fitted by PFF-VP with IP number varying from 3 to 35 and the fitting results are evaluated by calculating the MAEs in the gaps. Therefore, we get 12 curves of MAEs varying with IP number. Each curve is normalized by dividing by the average snow depth corresponding to the grid cell; thereafter, the relative MAEs and error bar can be obtained by calculating their average and STD, which are shown in Fig. 6. In this way, the impact of the changes of the values of the snow depth is neutralized, the impact of lack of measurements plays the major role, and the optimal IP number can be obtained by only considering missing data patterns.

As shown in Fig. 6, the curves at the two grid cells share the same characteristics. The relative MAEs show a slowly increasing trend when the IP number is smaller than 15 and decrease rapidly when the IP number is between 15 and 18 (for gridcell B) or 19 (for gridcell A). When the IP number is greater than 19, the relative MAEs increase again. It is concluded that the optimal IP number corresponding to the two grid cells are 19 and 18, respectively. In addition, the STD is smaller when the IP number is smaller than 25, which indicates that the relative MAE at the same IP number is only weakly dependent on the grid cell, no matter where the values of snow depth change dramatically or smoothly. The changes of the values of snow depth have little influence on the selection of the optimal IP number, and the fraction of missing data plays a vital role.

After determining the optimal IP numbers for the two grid cells, Fig. 5 shows the fitting results of the four methods for these two grid cells. There is no way to calculate the MAEs for missing parts, but the above method ensures that the fitting results will still be good in the missing parts. Furthermore, we can evaluate the quality of the fitting by comparison with the observations. Table 3 shows that the MAEs by PFF-VP are half of those by other methods at gridcells A and B.

As shown in Fig. 5a, QFF and GFF show their inflexibility. The fitted curves using QFF and GFF deviate in shape from the actual snow-depth data. It will snow only when there is a lot of water vapor and rising motion over the Arctic, meaning that the real snow-depth accumulation proceeds irregularly in time. The
drawback is evident in Fig. 5b. In addition, in Fig. 5a, the actual maximum of snow depth between August 2003 and August 2004 occurs much closer to August 2004. PFF-VP can solve these problems effectively and can better capture the corresponding changes according to the actual situation. It can be seen from Table 3 that the average MAE using PFF-VP is smaller than those using the other three methods for all grid cells where missing data are less than 40%. However, at the end of the time series in Fig. 5b, the fitting value using PFF-VP is abnormally high, because the fitting value of the boundary is affected by fewer data, which easily leads to anomalous fitting results. The results of middle parts are still acceptable. After all, the missing rate of the time series is 40%.

Following the above method, we can determine the optimal IP numbers of seriously incomplete time series, but the process is complex. We count the optimal IP numbers of more than 7000 grid cells to find their rules. As mentioned above, the optimal IP number may be related to two factors. The obvious periodicity of snow depth makes the impact of snow-depth variation smaller than that of the fraction of missing data. We find that the maximum length of gaps is the main factor in the selection of the optimal IP number as shown in Fig. 7. We can see that there is a strong negative correlation between the maximum length of gaps and the optimal IP number, and the correlation coefficient is −0.7523. When the maximum length of gap is less than 100 days, the optimal IP number is between 29 and 35. When it is between 100 and 200 days, which is also the case at most grid cells, the optimal IP number is mostly between 16 and 24. In this case, “two IPs per year” is a reasonable scheme for practical applications when large fractions of grid cells are considered. The time series start in summer and end in winter. So the main characteristics of the variation of snow depth can be grasped by choosing IPs in summer and in winter. When the maximum length of gap is more than 200 days, the optimal IP number is close to 3, which means that the fitting results are basically the same every year. Too much continuous missing data lead to the loss of too much important information in some regions. If more IPs are selected, the data between adjacent IPs in these regions will become scarce and the fitting results will be deteriorated. In other words, this method is not applicable to this situation. However, even when the maximum length of gaps is between 100 and 200 days, the optimal IP number of some grid cells is close to 3. This is because the data itself will have a certain impact. The simpler the variation of the snow depth is, the fewer IPs are needed. It is not clear whether there are other factors affecting the selection of IP number, but their impact is not major. Besides, the total fraction of missing data does not have a straightforward effect on the selection of the optimal IP number. The optimal IP number for the grid cells with low fraction of missing data may be close to 3, while 30 for those with high fraction of missing data. The maximum length of gaps plays a major role. Once the missing rate is more than 40%, the probability of

<table>
<thead>
<tr>
<th>Grid cell</th>
<th>PFF-CP</th>
<th>QFF</th>
<th>GFF</th>
<th>PFF-VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>5.64</td>
<td>5.32</td>
<td>2.29</td>
</tr>
<tr>
<td>B</td>
<td>3.94</td>
<td>5.64</td>
<td>5.57</td>
<td>2.09</td>
</tr>
<tr>
<td>All</td>
<td>3.69</td>
<td>3.42</td>
<td>3.25</td>
<td>2.47</td>
</tr>
</tbody>
</table>
the maximum length of gaps exceeding 200 days will increase, so the grid cells with missing rate smaller than 40% are selected to fit.

A quadratic function is used to fit the optimal IP number of more than 7000 grid cells. The result is expressed by

\[ y = 8.12 \times 10^{-5}x^2 - 0.12x + 35.16, \]  

where \( y \) is the optimal IP number and \( x \) is the maximum length of gaps, which is shown in red line in Fig. 7. If a lot of data need to be filled by this method, the scheme of selecting the optimal IP number is reliable. According to Table 2, the MAE using PFF-VP with the optimal IP number is approximately 50% smaller than that by other methods, while the MAE using PFF-VP with the suggested scheme is approximately 30% smaller than that by other methods. Although the suggested scheme is not as good as the optimal IP number, this method can still show its advantages.

### 6. Conclusions

Because of the significant temporal gaps in Arctic snow-depth data, continuous time series of snow depth is hard to acquire. The PFF-VP methodology proposed in this paper provides a method for the acquisition of continuous time series of snow depth.

The trigonometric function is selected as the fitting function with a period of 1 year, which is consistent with the actual snow depth. This is because there is no snow on the ice surface in August, and the snow gradually accumulates from September to May, and then melts rapidly from May to August (Warren et al. 1999). Considering the difference in annual snow depth, we set the time-varying amplitude and phase. The change of amplitude can reflect the change of snow accumulation and melting rate, and the change of phase can reflect the nuanced difference of each period. To get time-varying amplitude and phase, some time points are selected uniformly in the time domain as IPs. The amplitude and phase at the IPs are used as independent parameters. The values of the amplitude and phase at other points are obtained by cubic spline interpolation with these independent parameters. Thus, the function is expressed as a linear expression of the independent parameters and \( h_0 \). The least squares method is used to fit the data and obtain values of independent parameters and \( h_0 \). Thus, the time-varying amplitude and phase are obtained as well as the complete time series of snow depth.

In the experiments in section 4, 12 time series that have the smallest fraction of missing data among all grid cells were selected. For the time series, we introduce artificial gaps to them, trying to mimic the real gap structure. The gap structures are provided by other grid cells. The accuracy of the four methods can be evaluated by calculating the MAEs between the fitting results and the data removed in the artificial gaps. Compared with other fitting methods, the curves fitted by PFF-VP are in good agreement with the prescribed curves, having the smallest MAEs. The statistical analysis of the fitting results of all the missing data patterns shows that mean fitting results are close to prescribed data.

In section 5, to study the factors that affect the optimal IP number, almost the same missing periods are introduced to 12 relatively complete time series. Based on statistics of these 12 fitting results, the optimal IP number is set as the one corresponding to the minimum relative MAE. After obtaining the optimal independent points for more than 7000 grid cells, it is found that the optimal IP number is negatively correlated with the maximum length of gaps. By fitting the optimal IP number with a quadratic function, an empirical formula for selecting the optimal IP number is obtained. The other three fitting methods show their advantages in some cases, but they still have some shortcomings. PFF-VP can combine their advantages and effectively fit different snow-depth data at different grid cells. The MAEs of PFF-VP are smaller than those of PFF-CP, QFF, and GFF. In addition, PFF-VP can provide a complete and continuous time series of snow depth, which can be used for future research.

### Acknowledgments

The insightful comments from the two anonymous reviewers are gratefully acknowledged, which help improving the quality of this paper. We thank Anzhou Cao and Zheng Guo for their help in the process of writing the manuscript. This work is supported by the
National Key Research and Development Plan of China (Grants 2016YFC1402705 and 2016YFC1402304) and the National Natural Science Foundation of China (Grant U1806214). The snow-depth data are obtained for years 2002–11 from the National Snow and Ice Data Center (NSIDC; https://nsidc.org/node/29078). The main codes of this article are included in S_TIDE package, and the new version of S_TIDE package can be downloaded (https://www.researchgate.net/project/Adaptation-of-tidal-harmonic-analysis-to-nonstationary-tides).

APPENDIX A

The Solution of Cubic Spline Interpolation Coefficients

For convenience, let

\[ A(x) = \left( \frac{x-x_{i+1}}{h_i} \right)^2 \left( 1 + 2 \frac{x-x_{i+1}}{h_i} \right) y_i + \frac{x-x_{i+1}}{h_i} \left( 1 - 2 \frac{x-x_{i+1}}{h_i} \right) y_{i+1} \]

\[ + \left( \frac{x-x_{i+1}}{h_i} \right)^2 (x-x_{i+1}) m_i + \left( \frac{x-x_{i+1}}{h_i} \right) (x-x_{i+1}) m_{i+1}, \]  

where \( h_i = x_{i+1} - x_i \) and \( x_i \leq x \leq x_{i+1} \).

The second derivative of Eq. (A1) is

\[ A''(x) = \left[ \frac{6}{h_i^2} - \frac{12}{h_i} (x_{i+1} - x) \right] y_i + \left[ \frac{6}{h_i^2} - \frac{12}{h_i} (x-x_i) \right] y_{i+1} \]

\[ + \left[ \frac{2}{h_i} \frac{6}{h_i} (x_{i+1} - x) \right] m_i + \left[ \frac{2}{h_i} \frac{6}{h_i} (x-x_i) \right] m_{i+1}. \]

The second derivative of \( A(x) \) at \( x_i \) point is

\[ A''(x_i^+) = -\frac{6(y_i - y_{i+1})}{h_i^2} - \frac{4m_i + 2m_{i+1}}{h_i}, \]

\[ A''(x_i^-) = -\frac{6(y_i - y_{i-1})}{h_i^2} + \frac{2m_{i-1} + 4m_i}{h_i}. \]

Continuity of the second derivative of \( A(x) \) at point \( x_i \) requires that

\[ A''(x_i^+) = A''(x_i^-), \quad i = 2, 3, \ldots, K - 1. \]  

Equation (A2) can be simplified as

\[ (1 - \alpha_i)m_{i-1} + 2m_i + \alpha_i m_{i+1} = \beta_i, \]  

where

\[ x_i = p(i), \quad i = 1, \ldots, K, \]

\[ y_i = A[p(i)], \quad i = 1, \ldots, K, \]

\[ x = t. \]

Assume that a cubic spline interpolation function \( A(x) \) exists on the interval \([1, N]\). The terms \( m_i \) \((i = 1, 2, \ldots, K)\) represent the first derivatives of \( A(x) \) at \( x_i \), and the curve satisfies the following condition on each interval \([x_i, x_{i+1}]\):

\[ A(x_i) = y_i, \quad A(x_{i+1}) = y_{i+1}, \]

\[ A'(x_i) = m_i, \quad A'(x_{i+1}) = m_{i+1}. \]

Therefore, we can use Hermite interpolation polynomial to write the expression of \( A(x) \) on interval \([x_i, x_{i+1}]\):

\[ \alpha_i = \frac{h_{i-1}}{h_i} + h_i, \]

\[ \beta_i = \frac{1 - \alpha_i}{h_i} (y_i - y_{i-1}) + \frac{\alpha_i}{h_{i-1}} (y_{i+1} - y_i). \]

In this paper, IPs are uniformly distributed; namely, \( h_i = h_{i-1} = h \). Thus, \( \alpha_i = 1/2, \quad \beta_i = (3/2h)(y_{i+1} - y_{i-1}). \)

Additional boundary conditions are required to determine cubic spline functions. We adopt common natural boundary condition: \( A''(x_1) = 0, \ A''(x_K) = 0; \) that is,

\[ 2m_1 + m_2 = \frac{3}{h} (y_2 - y_1), \]

\[ m_{K-1} + 2m_K = \frac{3}{h} (y_K - y_{K-1}). \]

We can get from Eqs. (A3)–(A5) that

\[ DM = Y, \]  

where

\[ M = (m_1, m_2, m_3, \ldots, m_{K-2}, m_{K-1}, m_K)^T, \]

\[ Y = \frac{3}{2h} [2(y_2 - y_1), y_3 - y_1, y_4 - y_2, \ldots, y_{K-1} - y_{K-3}, y_K - y_{K-2}, 2(y_K - y_{K-1})]^T. \]
Combining Eqs. (A1) and (A6) yields expression of $y_i$. Note that this expression is a linear function of $y_i$, and the values before $y_i$ are the interpolation coefficients. For convenience, Eq. (A1) is rewritten into the following form:

$$A(x) = g_1(x)y_1 + g_2(x)y_{i+1} + g_3(x)m_i + g_4(x)m_{i+1}, \quad x_i \leq x \leq x_{i+1}.$$ 

Letting $E = (3/2h)D^{-1}$, $E_{ij}$ represents the element at the $i$th line and $j$th column of $E$:

$$m_i = 2E_{i,1}(y_2 - y_1) + \sum_{j=2}^{K-1} E_{ij}(y_{j+1} - y_{j-1}) + 2E_{i,K}(y_K - y_{K-1})$$

$$= (-2E_{i,1} - E_{i,2})y_1 + (2E_{i,1} - E_{i,3})y_2 + \sum_{j=3}^{K-2} (E_{ij-1} - E_{ij+1})y_j + (-2E_{i,K} + E_{i,K-2})y_{K-1}$$

$$+ (2E_{i,K} + E_{i,K-1})y_K = \sum_{j=1}^{K} t_{ij}y_j,$$

(A7)
Similarly,\[ m_{i+1} = \sum_{j=1}^{K} l_{i+1,j} y_j, \] (A8)

Here, \( f \) at the \( x \)th point in the interval \([x_i, x_{i+1}]\) of the \( j \)th IP can be expressed as
\[
f_{x,j} = \begin{cases} 
  l_i g_3(x) + l_{i+1,j} g_4(x), & j \neq i, i + 1, \\
  l_i g_3(x) + l_{i+1,j} g_4(x) + g_1(x), & j = i, \\
  l_i g_3(x) + l_{i+1,j} g_4(x) + g_2(x), & j = i + 1.
\end{cases}
\]

APPENDIX B

Fitting Results For Gridcells 1–12

The fitting results for all 12 time series with artificial gaps are shown in Figs. B1 and B2.

REFERENCES


