Understanding sea ice melting via functional data analysis

Purba Das, Ananya Lahiri and Sourish Das*

Chennai Mathematical Institute, Chennai 603 103, India

We have addressed the problem of sea ice extent (SIE) melting for Arctic and Southern Ocean. The 'satellite passive microwave remote sensing data' for daily SIE over a year has been considered as a smooth continuous function, called functional data. The analysis of the mean function of SIE over the different year-blocks and 95% bootstrap confidence bands for Arctic Ocean shows a statistically significant drop in SIE. Additional evidence of SIE melting in the Arctic Ocean is provided through phase plane analysis. During the summer, the SIE is observed about 30% less in the current year-block than that of the first year-block.

Keywords: Arctic ocean, bootstrap and confidence band, mean function, sea ice extent, southern ocean.

RECENTLY functional data analysis (FDA)[1] is attracting a lot of attention in the scientific community. With the new digital explosion of data, FDA is a highly effective tool to understand many complex systems. FDA is broadly used in biostatistics, growth curve model, longitudinal data analysis, meteorology, etc.[2,3].

We have addressed the problem of sea ice extent (SIE) melting for Arctic and Southern Oceans over years. The measurement of the extent and trend analysis is available in the literature for the Arctic Ocean[4]. The satellite evidence of the transformation of the Arctic SIE was presented by Johannessen *et al.*[5]. Sharp changes in SIE for the Arctic Ocean over time were reported[6–8]. The effect of Arctic SIE decline on weather and climate has been reviewed[9]. However, the analyses of SIE for the Arctic as well as the Southern Ocean together have been rarely considered. Recently, Josefino[10] provided evidence for asymmetry in the rate of change of SIE melting in the Arctic and the Southern Oceans. To the best of our knowledge, no work has been found considering SIE data as a functional data.

We have treated the 'satellite passive microwave remote sensing' daily data for a year as a functional data for both the Arctic and the Southern Oceans. We have used the FDA package 'FDA'[11] in *R* for the Fourier smoothing of raw data and analysis. In this article, we provide statistically significant evidence of SIE melting in the Arctic Ocean. We also provide analysis of SIE for the Southern Ocean. We present phase plane (PP) analysis and a functional measure of severity of the change in SIE over year-blocks.

*For correspondence. (e-mail: sourish@cmi.ac.in)

Functional data

Data source and description

We have used two different data sources. The daily SIE data for the Arctic and the Southern Oceans are available from National Snow and Ice Data Center (NSIDC), website ftp://sidads.colorado.edu/DATASETS/NOAA/G02135/. These data are available from November 1978 to December 2015. As we wanted to study the yearly effect, we have considered data from 1 January 1979 to 31 December 2015 (37 years data) for both the Arctic and the Southern Oceans. NASA website http://neptune.gsfc.nasa.gov/csb/index.php?section=59 is another source of data of SIE for Arctic and Southern Oceans. These data are available from November 1978 to December 2013.

In both sources, from 1 January 1979 to 20 August 1987 the SIE data are available only on alternate days and from 21 August 1987 onward we have daily data. The data contain the total SIE of the Arctic Ocean along with the break-up of the nine different seas of the Arctic Ocean. In Figure 1 snapshots of SIE of both the Arctic and the Southern Oceans have been provided.

Motivation for FDA and modelling

Figure 2 gives the 37 years of raw SIE data. It shows that that the daily SIE over a year can be considered to be a smooth continuous function.

We have considered the model as

$$y_j^i = x^i(t_j) + e_j^i, \quad j = 1,...,n_i, \quad i = 1,...,37, \tag{1}$$

where i represents year and j represents days. The error term e_j^i assumes to be distribution free but with zero mean and finite variance, $y^i = (y_1^i, y_2^i, ..., y_{n_i}^i)$ is the observed data vector for ith year with length n_i. The $x^i(t_j)$

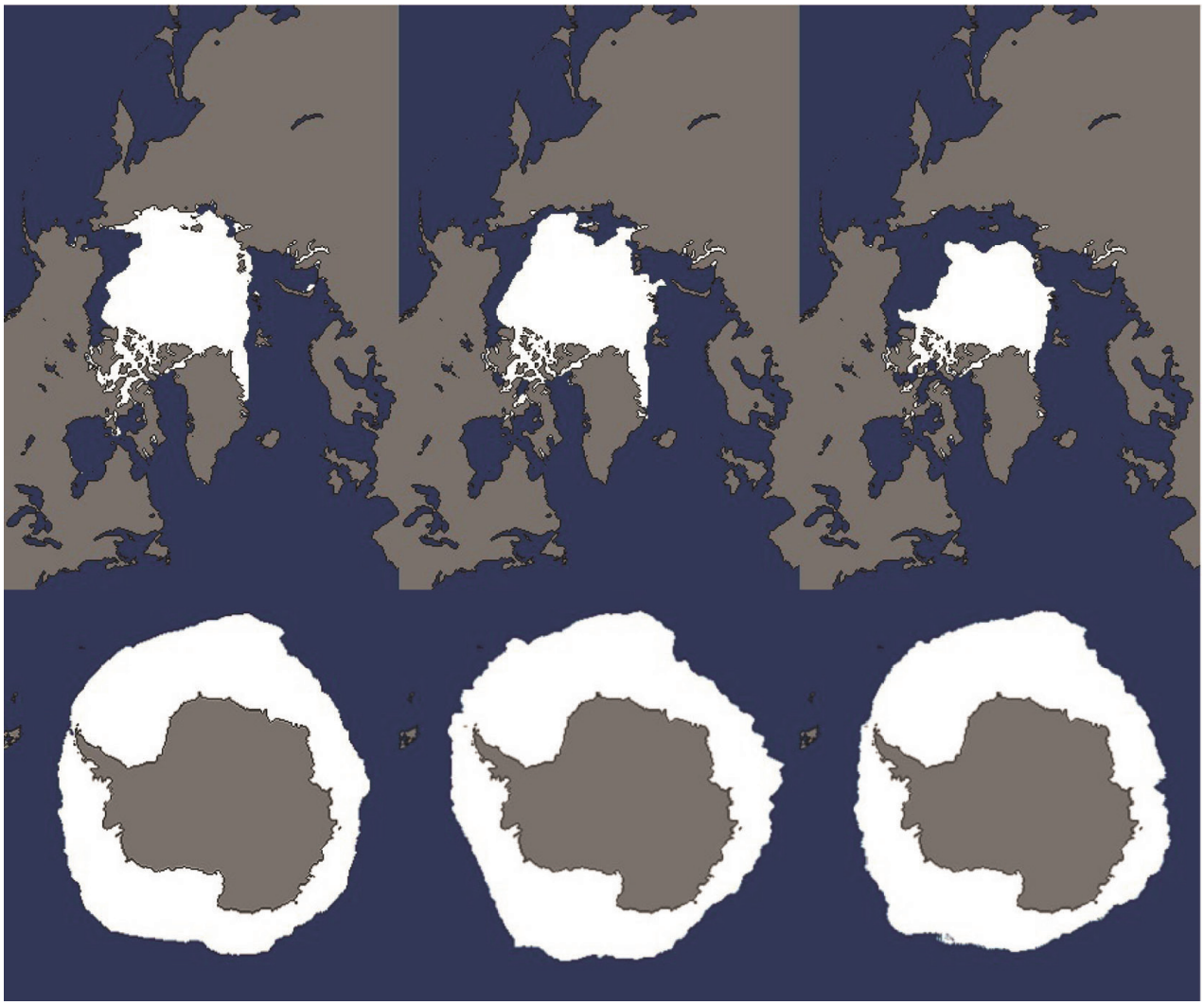

Figure 1. Snapshots of SIE of the Arctic and the Southern Oceans for September in the years 1979, 1997, 2017 respectively. Source: NASA.

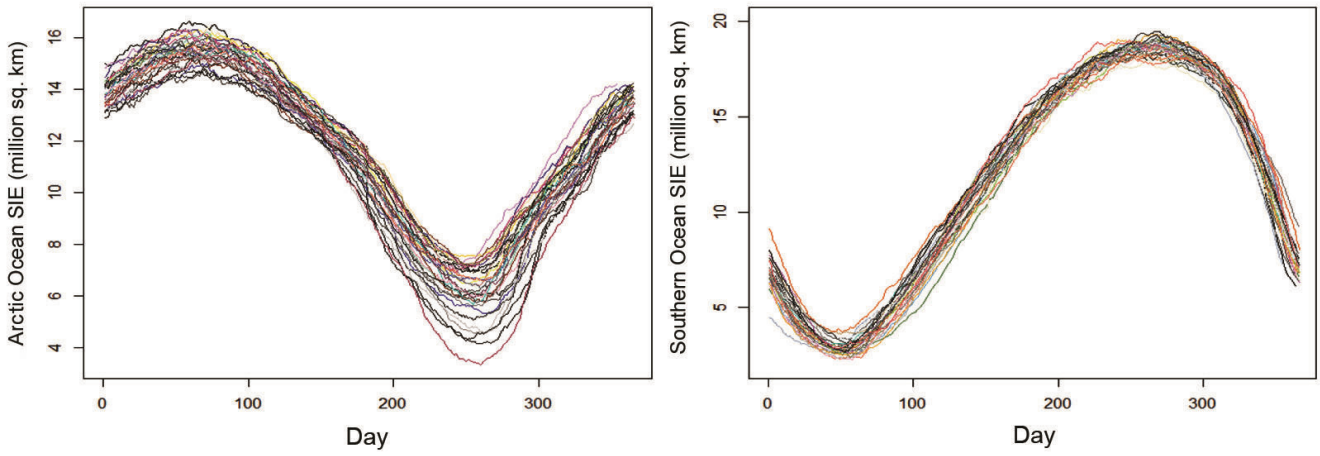

Figure 2. Arctic and Southern Ocean SIE (in a million sq. km) over days for 37 years (raw data).

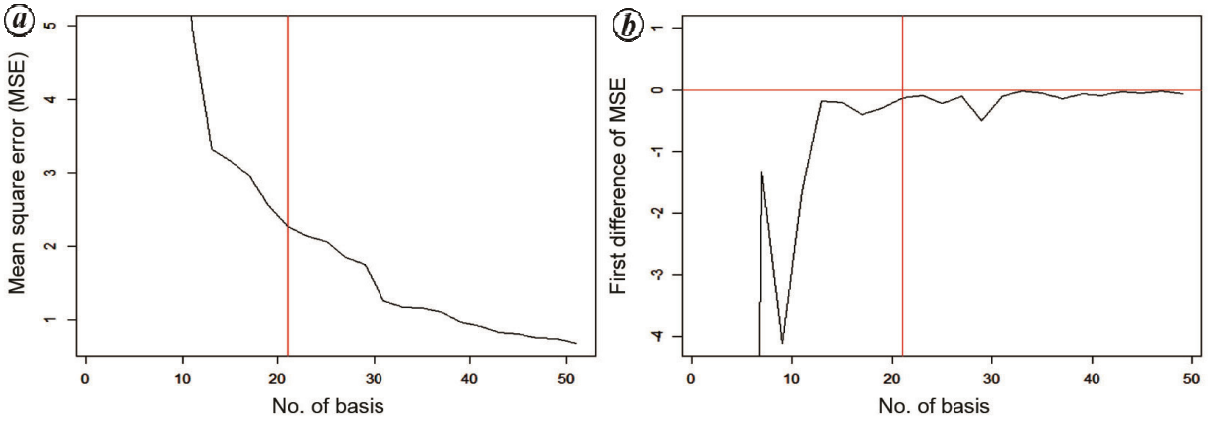

Figure 3. ***a***, $\widehat{\text{MSE}}$ versus number of Fourier basis (grey line represents number of basis = 21). ***b***, First difference of $\widehat{\text{MSE}}$ (grey lines correspond to number of basis = 21 and $\widehat{\text{MSE}} = 0$).

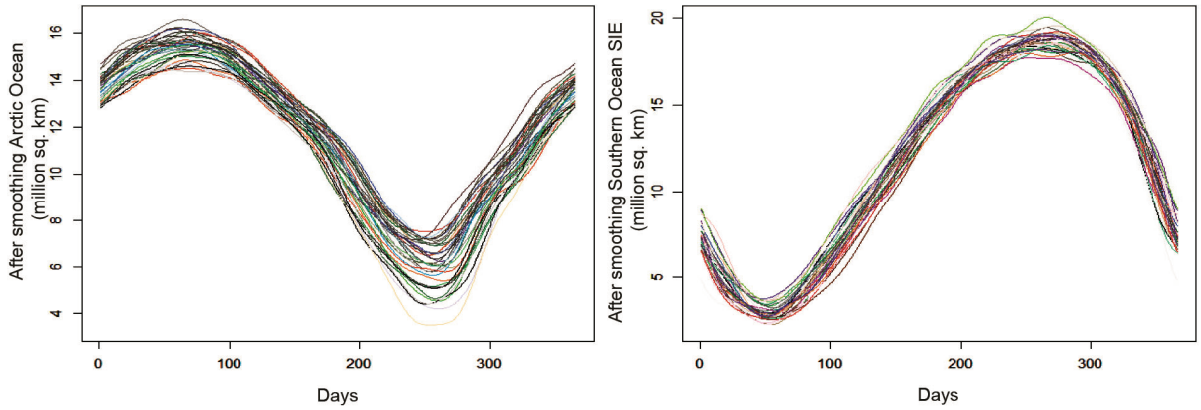

Figure 4. Arctic and Southern Ocean SIE (in a million sq. km) after smoothing over days for 37 years.

is the ith function evaluated at the time point t_j. The functional form of the model is as follows

$$x(t) = \sum_{k=1}^{p} c_k \phi_k(t), \tag{2}$$

where p is assumed to be a fixed constant for all 37 functions, t represents time, c_k's are real constants, ϕ_k's are Fourier basis as follows

$$\phi_k(t) = \begin{cases} 1, & \text{if } k = 1, \\ \sin(\frac{k}{2}\omega t), & \text{if } k \text{ is even}, \\ \cos(\frac{k-1}{2}\omega t) & \text{otherwise.} \end{cases} \tag{3}$$

Fourier smoothing and number of basis selection

We want to fit FDA model (1, 2, 3) for SIE data. To find the optimal number of basis we have calculated the mean squared error (MSE) for ith year as

$$\text{MSE}^i(p) := \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{e}_j^{i\,2},$$

where $i = 1, 2, \ldots, 37$, $\hat{e}_j^i$ is the estimated error after fitting the FDA model (1) using Fourier basis. We have computed the coefficients c_k's by the weighted least squared method using FDA package in R[11]. We have defined, for each fixed odd p, as

$$\widehat{\text{MSE}}(p) := \frac{1}{37} \sum_{j=1}^{37} \text{MSE}^i(p).$$

We have plotted $\widehat{\text{MSE}}(p)$ for all odd p between 1 and 51 in Figure 3 *a*. Figure 3 *a* shows that the slope is changing at $p = 21$, but for further justification we have studied the plot of first difference of $\widehat{\text{MSE}}(p)$ in Figure 3 *b*. We conclude $p = 21$, as the fluctuation is nearly zero thereafter. Figure 4 shows daily SIE over a year using Fourier basis smoothing. Plots in Figure 3 correspond to the Arctic Ocean SIE. For the Southern Ocean SIE also, the value of p turns out to be 21.

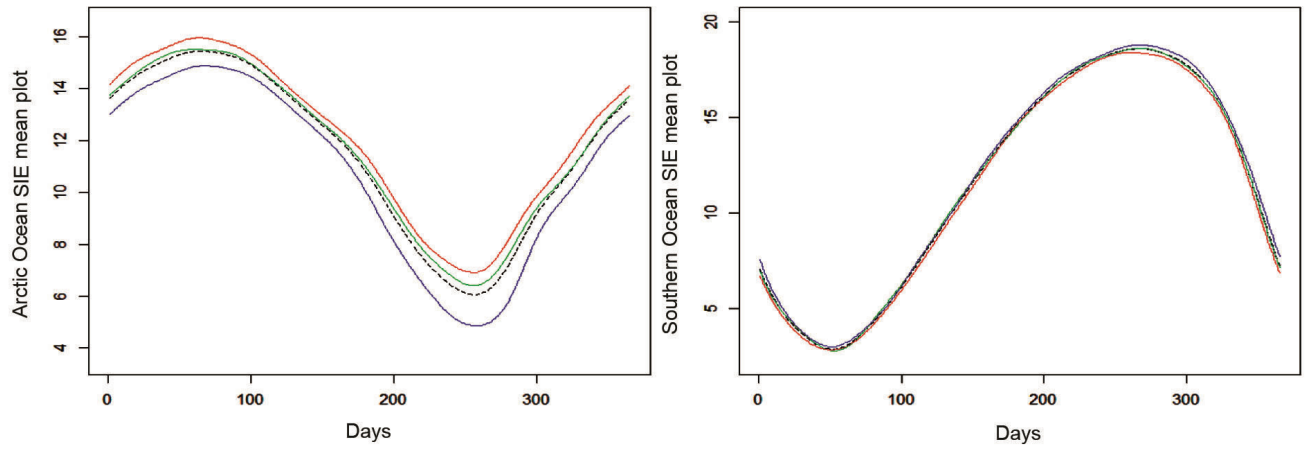

Figure 5. Mean curves (year-block-wise) of the Arctic and the Southern Ocean SIE (in a million sq. km) after smoothing. For 37 years (dotted line); years 1979–1990; years 1991–2003; years 2004–2015. (y-axis unit: million sq. km).

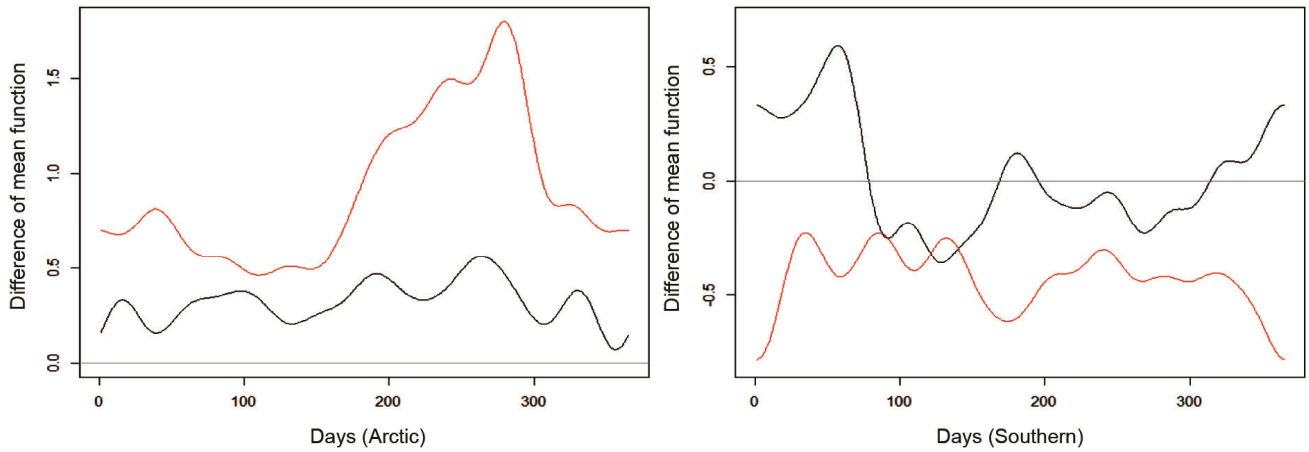

Figure 6. Consecutive difference of year-block-wise mean curves for both the Arctic and the Southern Oceans.

Summary statistics

For the Arctic Ocean, the SIE reaches the maximum between 5 and 9 March and minimum between 10 and 14 September. This highest and lowest SIE occur about 30 days later than the occurrence of minimum and maximum sea-surface temperatures (SST) (source: mean temperature for the north of the 80° north can be found here: http://ocean.dmi.dk/arctic/meant80n.uk.php) respectively.

For the Southern Ocean, the SIE reaches the maximum between 22 and 26 September and minimum between 19 and 23 February. This highest and lowest SIE occur about 40 days later than the occurrence of minimum and maximum SST (source: mean temperature for three different stations of Southern Ocean can be found here: http://www.coolantarctica.com/Antarctica%20fact%20file/antarctica%20environment/vostok_south_pole_mcmurdo.php) respectively. The average SIE of Arctic and Southern Oceans is 11.50, 11.76 million sq. km respectively.

Mean function analysis

The mean function[1] of FDA is defined as

$$\bar{x}(t) := \frac{1}{N}\sum_{i=1}^{N} x_i(t),$$

where $\bar{x}(t)$ is the point-wise mean of N functions. Figure 5 shows the mean functions of SIE for both the Arctic and the Southern Oceans. The dashed line represents the mean function of smoothed SIE computed over all 37 years.

We have divided 37 years into t blocks which we call year-blocks d_1, d_2,, …, d_t. Here we take $t = 3$. The red, green, blue curves represent mean functions of SIE for the year-blocks, d_1 = (1979–1990), d_2 = (1991–2003) and d_3 = (2004–2015) respectively.

For the Arctic Ocean, the minimum SIEs (occur between 265th and 269th days of the year) are 6.96, 6.43, 4.95 million sq. km and the maximum SIEs (occur

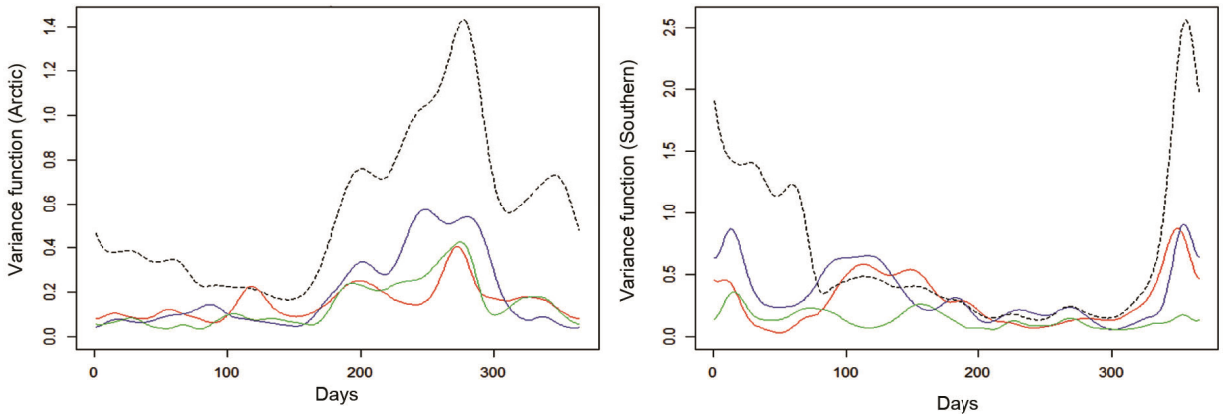

Figure 7. Variance curves (year-block-wise) of Arctic and Southern Oceans SIE (in million sq. km) after smoothing. For 37 years (dotted line): years 1979–1990, years 1991–2003, years 2004–2015.

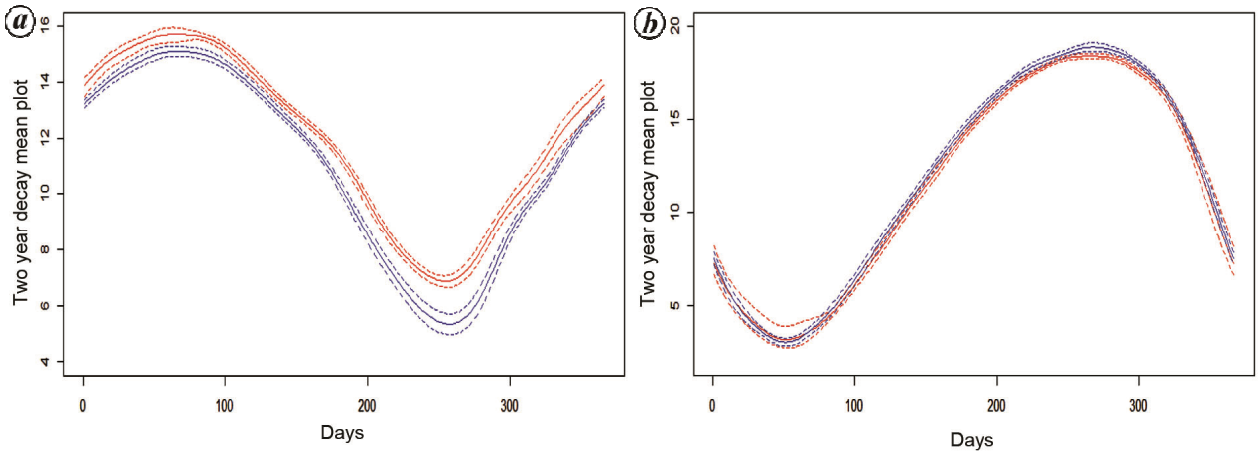

Figure 8. Confidence band for Arctic (***a***) and Southern (***b***): years 1979–1996, years 1997–2015. (*y*-axis unit: million sq. km).

between 50th and 54th days of the year) are 15.83, 15.51, 14.92 million sq. km for the three consecutive year-blocks respectively. For the Southern Ocean the minimum SIEs (occur between 64th and 68th days of the year) are 3.34, 2.78, 3.16 million sq. km and the maximum SIEs (occur between 253th and 257th days of the year) are 18.35, 18.57, 19.01 million sq. km for the three consecutive year-blocks respectively. From the year-block-wise mean function plot, a downward trend in the Arctic Ocean SIE is observed. However, from the corresponding plot for the Southern Ocean no such trend is found.

In Figure 6, the difference between the mean functions of the first two year-blocks (the difference between green and red curves in Figure 5) and the difference between the mean functions of the last two year-blocks (the difference between blue and green curve in Figure 5) are represented by the black and red curves respectively. For the Arctic Ocean, both the curves fall in the positive region but for the Southern Ocean, both the curves are hovering around zero line. During the 170th to 300th day of the year, the change in the Arctic Ocean SIE is more prominent than the rest of the year.

Variance function analysis

In FDA the variance function[1] is defined as

$$\mathrm{Var}(x(t)) := \frac{1}{N-1}\sum_{i=1}^{N}[x_i(t)-\bar{x}(t)]^2,$$

where $\mathrm{Var}(x(t))$ is the point-wise sample variance of the given N functions corresponding to N years. In Figure 7, the variance functions corresponding to same year-blocks, as in case of mean functions, are presented. For the Arctic Ocean, the variance is much higher in the interval [170, 300] days. However, for the rest of the year it is low. So we have a non-constant variance indicating high volatility during summer. For the Southern Ocean, we have observed non-constant volatility. This nonhomogeneity of

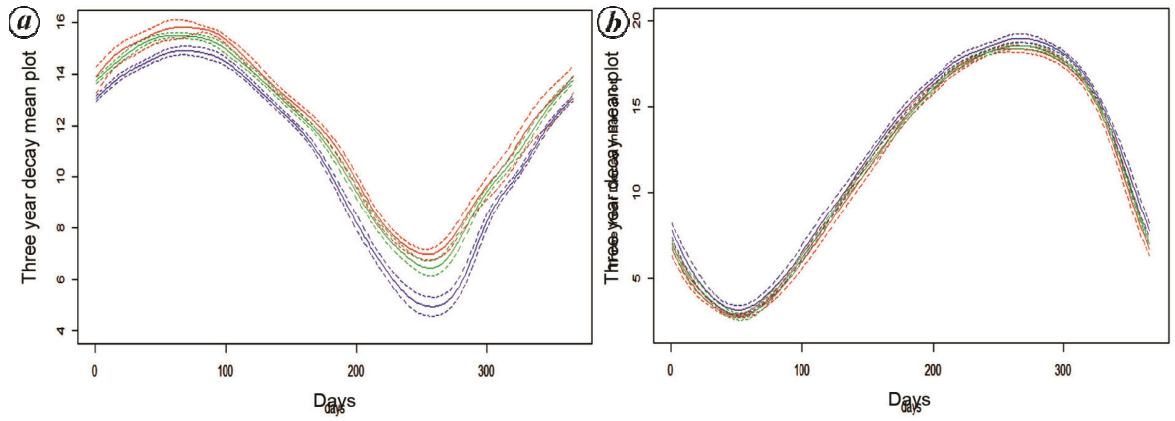

Figure 9. Confidence band for Arctic (***a***) and Southern (***b***): years 1979–1990, years 1991–2003, years 2004–2015. (*y*-axis unit: million sq. km).

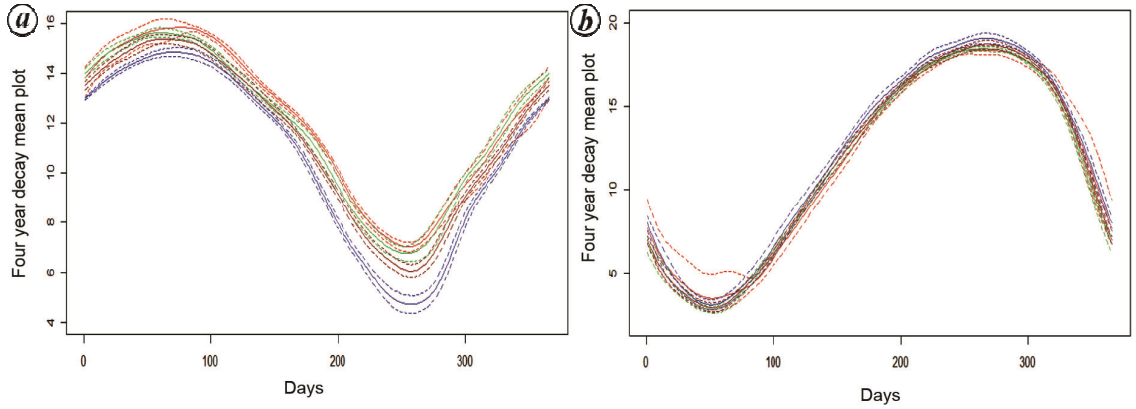

Figure 10. Confidence band for Arctic (***a***) and Southern (***b***): years 1979–1987, years 1988–1996, years 1997–2005, years 2006–2015. (*y*-axis unit: million sq. km).

variances justifies the weighted least squares method of estimation. This non-homogeneity will also reflect in the width of the confidence band.

Bootstrap confidence band for mean functions

In this section, we have taken $t = 2, 3, 4, 5$ in our analysis. Given t, for the ith year-block d_i with $i = 1, 2, \ldots, t$, we have computed the 95% bootstrap confidence band[1] or the year-block d_i along with its bootstrap mean function.

Suppose year-block d_i consists of b_i years. We have drawn b_i resample curves with replacement and have computed the mean function. We have repeated resampling process B times, B being the bootstrap sample size. We have taken the point-wise mean, variance and 95% confidence band for these B curves. Throughout the article, we have considered $B = 5000$.

Analysis for Arctic Ocean

We have presented the confidence band for SIE of the Arctic Ocean over different year-blocks. The dotted lines in Figures 8–11 (panel *a*) show the confidence band.

For $t = 2$, $n_1 = 18$ (years 1979–1996) and $n_2 = 19$ (years 1996–2015) and corresponding confidence band is shown in Figure 8.

For $t = 3$, $n_1 = 12$ (years 1979–1990), $n_2 = 12$ (year 1991–2003) and $n_3 = 13$ (years 2004–2015) and corresponding confidence band is shown in Figure 9.

For $t = 4$, $n_1 = 9$ (years 1979–1987), $n_2 = 9$ (years 1988–1996), $n_3 = 9$ (years 1997–2005) and $n_4 = 10$ (years 2006–2015) and corresponding confidence band is shown in Figure 10.

For $t = 5$, $n_1 = 7$ (years 1979–1985), $n_2 = 7$ (years 1986–1992), $n_3 = 7$ (years 1993–1999), $n_4 = 8$ (years 2000–2007) and $n_5 = 8$ (years 2008–2015) and corresponding confidence band is shown in Figure 11.

From each of these figures, it is clear that 95% bootstrap confidence bands for t year-blocks are mostly non-overlapping and non-uniform in width. Hence we can statistically conclude that SIE for the recent year-block is significantly lower than the previous year-blocks.

Analysis for Southern Ocean

We have adopted the same methodology for the Southern Ocean as for the Arctic Ocean. The same analysis for the

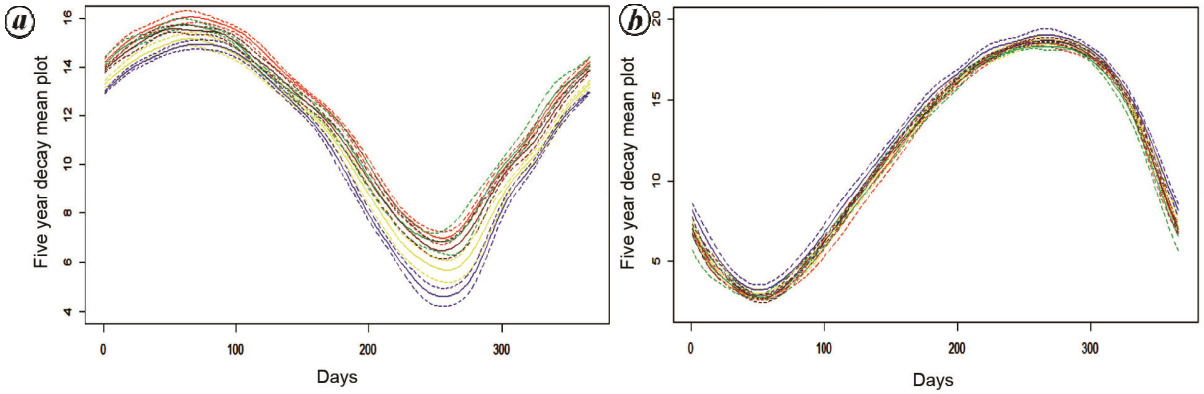

Figure 11. Confidence band for Arctic (***a***) and Southern (***b***): years 1979–1985, years 1986–1992, years 1993–1999, years 2000–2007, years 2008–2015. (*y*-axis unit: million sq. km).

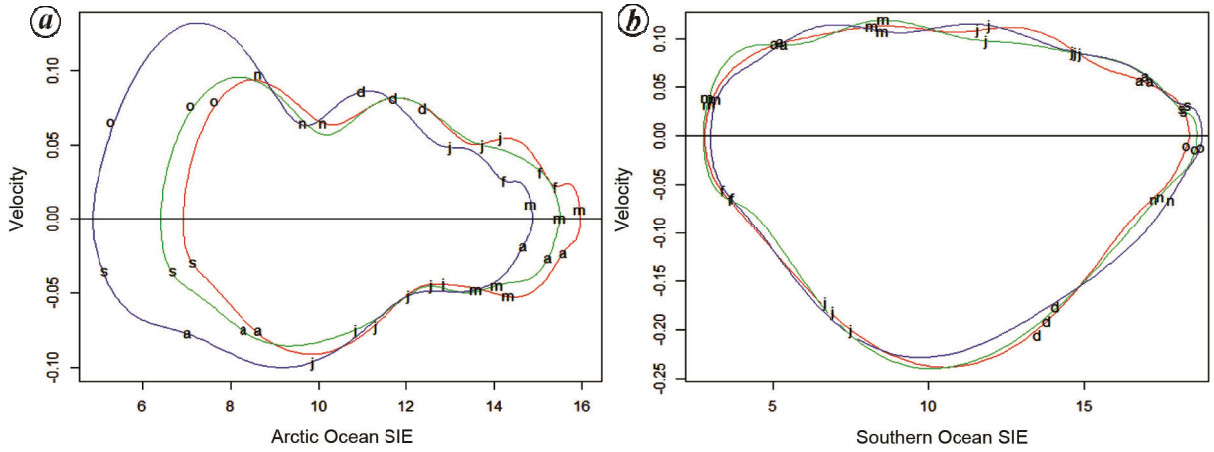

Figure 12. ***a***, Area versus velocity plot of SIE for Arctic Ocean. ***b***, Area versus velocity plot of SIE for Southern Ocean: years 1979–1990, years 1991–2003, years 2004–2015.

Southern Ocean is depicted in Figures 8–11 (panel *b*). In all these figures, the 95% bootstrap confidence bands are overlapping and non-uniform in width. Hence we have concluded that there is no statistical evidence of change in SIE in the Southern Ocean.

Phase plane analysis

In the phase plane (PP) analysis, we have compared the SIE, its first and second differences called area, velocity and acceleration respectively, for both the Arctic and Southern Oceans. The first and second difference on the smoothed Arctic (Southern) Ocean SIE data can be assumed to be a good approximation for the first and second derivatives of the function SIE. These differences have been used for further analysis.

Analysis of area versus velocity

We have considered the plot of area versus velocity for the three year-blocks. The year-blocks are defined in the previous section. One single curve represents 365 days phase plot. For our convenience, we have tagged the first day of each month with the first letter of the respective month in all three year-blocks. The area versus velocity plot is represented in Figure 12 for both the Arctic and the Southern Oceans. The red (first), green (second) and blue (third) curves represent area versus velocity plot for year-blocks d_1 = (1979–1990), d_2 = (1991–2003) and d_3 = (2004–2015) respectively.

The third curve is more leftward than the first curve. When SIE is minimum (September), the leftward shift over year-blocks is higher than that when SIE is maximum (March). This shows SIE is decreasing over year-blocks.

It is observed that between 1 and 15 March SIE reaches the maximum while between 5 and 15 September, it reaches the minimum for all three year-blocks in the Arctic Ocean (Figure 12 *a*).

During the time interval from July to October, when the SIE reaches minimum, the third curve dominates the second curve, and the second curve dominates the first one for the Arctic Ocean. During this time span for

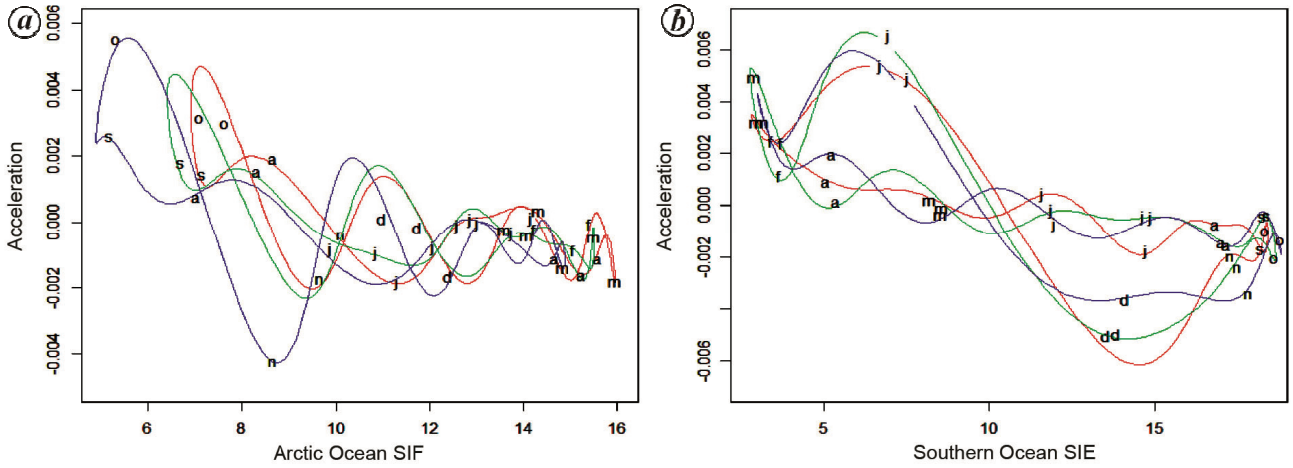

Figure 13. ***a***, Area versus acceleration plot of SIE for Arctic Ocean. ***b***, Area versus acceleration plot of SIE for Southern Ocean: years 1979–1990, years 1991–2003, years 2004–2015.

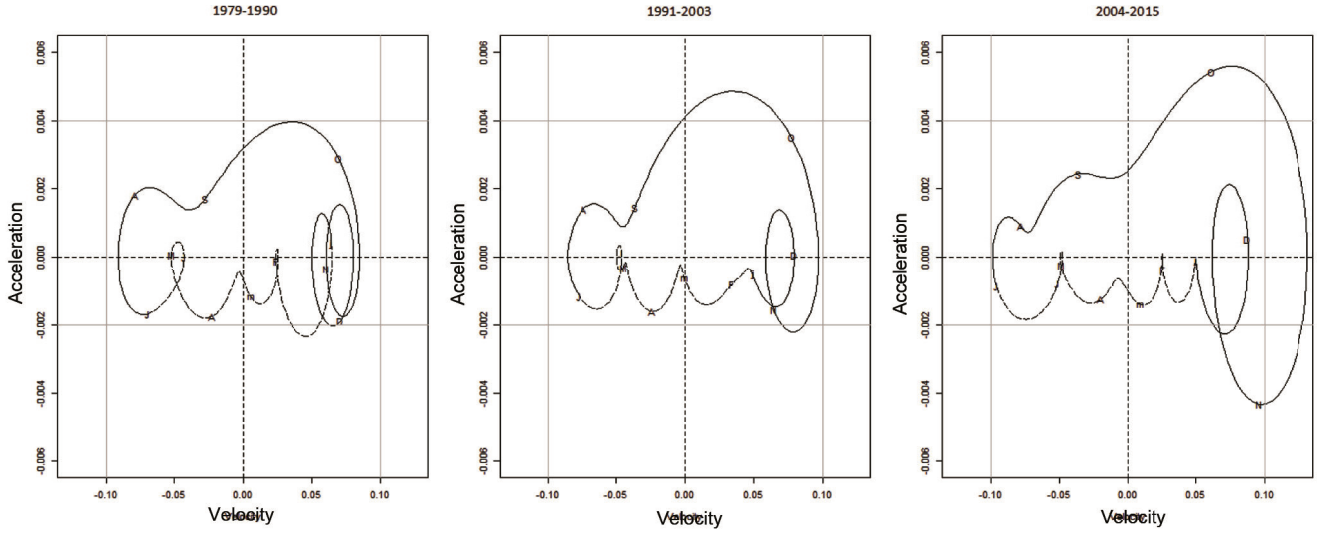

Figure 14. Velocity versus acceleration plot of SIE for Arctic Ocean.

later year-blocks melting of SIE occurs at a faster rate as well as the formation of SIE also takes place at a faster rate.

During January to April, when SIE reaches the maximum, the first curve dominates the second curve and the second curve dominates the third one. SIE decreases more rapidly over the year-blocks. During later year-blocks, while SIE forms at a slower rate and melting of SIE takes place at a faster rate. The asymmetry of SIE formation (melting) over seasons should be noted here.

During May, June, November, December, the velocity variations are mostly the same for all year-blocks. So for these time periods, there are no velocity changes across year-blocks. For the Southern Ocean no such observations can be made (Figure 12 *b*).

Analysis of area versus acceleration

From Figure 13 it is observed that the area versus acceleration curve has mostly similar behaviour over year-blocks with a time lag of 6 months between the Arctic and the Southern Oceans. All notations used for the plots are exactly the same as in Figure 12. For January to April the acceleration is mostly negative whereas for July to October the acceleration is mostly positive (Figure 13 *a*). The maximum amplitude of the curve with positive acceleration is observed to be higher than the maximum amplitude of the curve with negative acceleration. This explains the asymmetry of SIE formation (melting) over seasons. Like the area versus velocity curve for July to October, here the left shift of the curve persists.

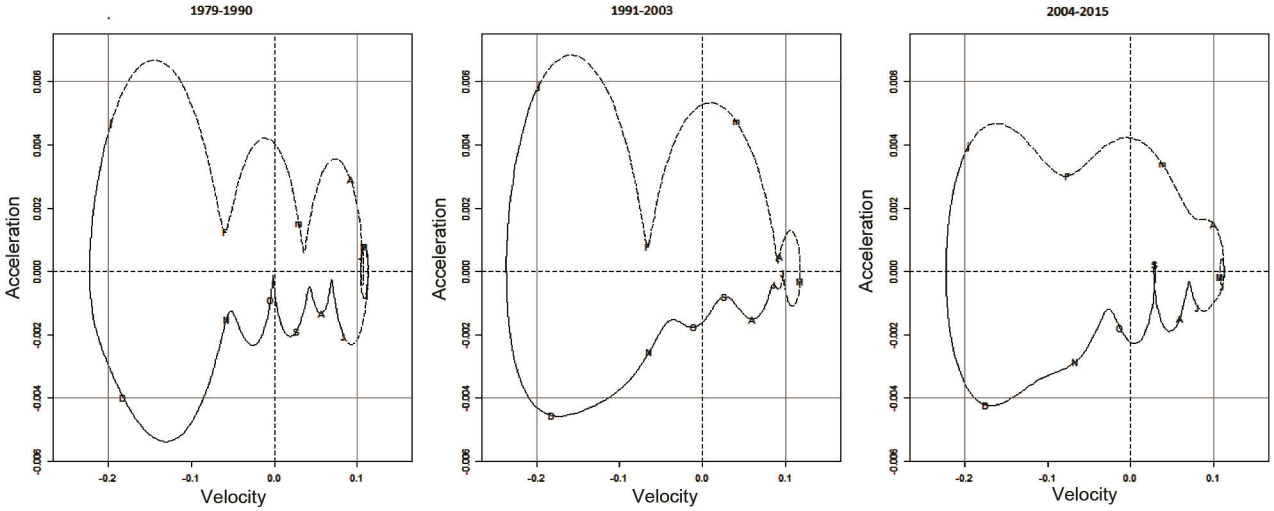

Figure 15. Velocity versus acceleration plot of SIE for Southern Ocean.

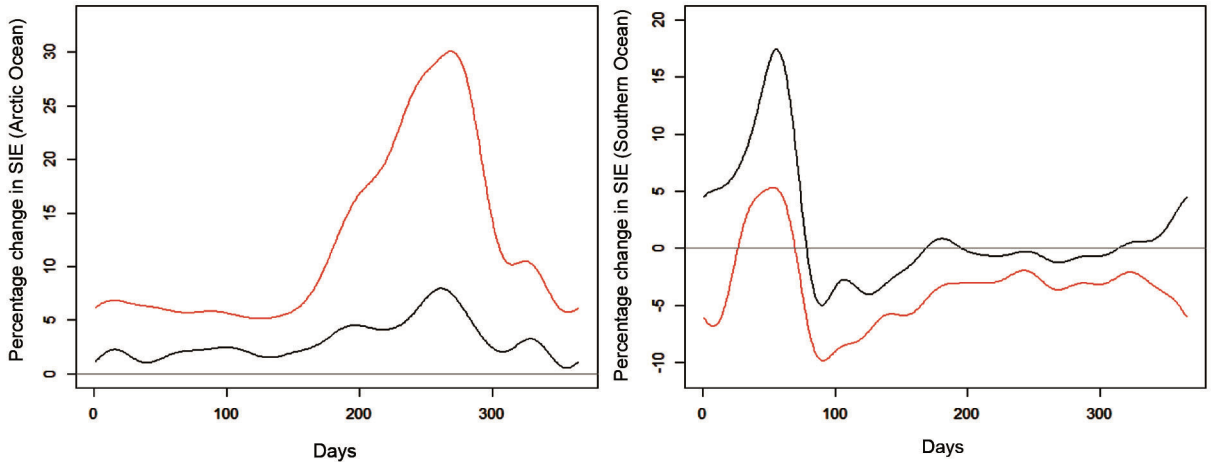

Figure 16. Percentage change in SIE of Arctic and Southern Ocean.

This implies that the rate of change of melting and formation of ice occur at a faster rate.

For the Southern Ocean in the first half of the year the acceleration is positive and for the other half of the year, it is negative (Figures 13 *b*). But no other significant findings are noted.

Analysis of velocity versus acceleration

The velocity versus acceleration plot has a lot of physical interpretations. The velocity axis is considered to be the representative of the kinetic energy and the acceleration axis of the potential energy level of the system. The explanation for the velocity versus acceleration plot can be found in ref. 1 (p. 32). The larger radius implies the increase in energy transfer, which means the system is going away from equilibrium. If the horizontal location of the center is towards the right, there is a net positive velocity in the system and if it is towards left there is a net negative velocity in the system. If the vertical location of the center is above zero there is a net velocity increase in the system and if it is below zero, there is a drop in velocity. The total energy change can be captured through the variations of the shapes of the cycles from year to year. Figures 14 and 15 represent the velocity versus acceleration plots for the Arctic and the Southern Oceans respectively.

In this section, we have used FDA package for our plots. Symbol reference is taken from *R* code of FDA package as follows: letters indicate mid-months, with lowercase letters used for January and March. For clarity, the first half of the year is plotted as a dashed line and the second half as a solid line.

In June to August the sub-cycle decrease in average radius explains the faster rate of melting of SIE. During this period the radius of the curve strictly decreases year-block-wise. Therefore there is a drop in energy in the system over year-blocks.

In the October to December sub-cycle, the increase in average radius explains the faster rate of formation of SIE. In this period the radius of the curve strictly increases year-block-wise. Therefore there is a rise in energy in the system over year-blocks.

For the Southern Ocean the change in shape of the velocity versus acceleration plot represents the internal change in the energy of the system though we do not have the significant change in the area covered by the curves. This indicates a structural change in the system over year-blocks.

Measurement of percentage

In the previous sections the statistical evidence of decrease in SIE over year-blocks in the Arctic Ocean is observed. In this section a measure of percentage change of drop in SIE in the Arctic Ocean, over the year-blocks, is proposed. Let m_1, m_2, m_3 be the mean SIE function of Arctic (Southern) Ocean for the first, second and third year-block d_1, d_2, d_3 respectively. The $m_j(i)$ represents the value of mean SIE evaluated at the ith day of the function m_j for $i = 1, 2, \ldots, 365$ and $j = 1, 2, 3$. We have defined two measures $\mu_{1,j}$ with $j = 2, 3$ as follows

$$\mu_{1,j}(i) := \frac{(m_j - m_1)}{m_1}(i) \times 100\%.$$

These two measures $\mu_{1,2}$ and $\mu_{1,3}$ provide the percentage change of mean SIE in the second and third year-block with respect to that of the first year-block. The measurement of percentage change curve is shown in Figure 16. The red (black) curve presents the percentage change of mean SIE for both the Arctic and Southern Oceans for the third (second) year-block with respect to the first year-block. Figure 16 indicates that for the Arctic Ocean the third year-block observes a size of the mean SIE about 30% less during summer than what it used to be during the first year-block.

Conclusion

In this article, we have shown how to model SIE daily data for both the Arctic and the Southern Oceans as functional data. FDA is useful for missing value and unequal length prediction. For each year, daily SIE data has been considered as a smooth continuous function. We have used Fourier basis to estimate the smooth functions for 37 years (1979–2015). We have also provided an analysis of an optimal number of basis selection. After analysing the functional data we have observed significantly strong statistical evidence of decrease in SIE the of Arctic Ocean over year-blocks whereas in the Southern Ocean no statistical evidence for SIE melting has been found. For our analysis of SIE we have considered the mean curve and the bootstrap confidence band (95%) for the progressive blocks of years, called year-blocks. We have found the statistical evidence of SIE melting in the Arctic Ocean over year-blocks. We have found that the SIE in the third year-block is decreased by 30% during the summer in the Arctic Ocean compared to that of the SIE in first year-block. Similar analyses have been performed for the Southern Ocean also, but no significant statistical evidence of SIE melting has been found over the observed 37 years.

1. Ramsay, J. O. and Silverman, B. W., *Functional Data Analysis*, Springer, 2nd edn, 2005.
2. Wang, J.-L., Chiou, J.-M. and Müller, H.-G., Functional data analysis. *Ann. Rev. Stat. Appl.*, 2016, **3**, 257–295.
3. Ramsay, J. O. and Silverman, B. W., *Applied Functional Data Analysis: Methods and Case Studies*, Springer, 2002, 2nd edn.
4. Parkinson, C. L., Cavalieri, D. J., Gloersen, P., Zwally, H. J. and Comiso, J. C., Arctic sea ice extents, areas, and trends 1978–1996. *J. Geophys. Res. Oceans*, 1999, **104**, 20837–20856.
5. Johannessen, O. M., Shalina, E. V. and Miles, M. W., Satellite evidence for an Arctic sea ice cover in transformation. *Science*, 1999, **286**, 1937–1939.
6. Comiso, J. C., A rapidly declining perennial sea ice cover in the Arctic. *Geophys. Res. Lett.*, 2002, **29**(20), 17-1–17-4.
7. Stroeve, J., Holland, M. M., Meier, W., Scambos, T. and Serreze, M. C., Arctic sea ice decline: faster than forecast. *Geophys. Res. Lett.*, 2007, **34**(9), 1–5.
8. Comiso, J. C., Parkinson, C. L., Gersten, R. and Stock, L., Accelerated decline in the Arctic sea ice cover. *Geophys. Res. Lett.*, 2008, **35**(1), 1–5.
9. Vihma, T., Effects of Arctic Sea ice decline on weather and climate: a review. *Surv. Geophys.*, 2014, **35**(5), 1175–1214.
10. Josefino, C. C., Global changes in the sea ice cover and associated surface temperature changes. ISPRS – international archives of the photogrammetry. *Remote Sensing Spatial Infor. Sci.*, 2016, **XLI-B8**, 469–479.
11. Ramsay, J. O., Wickham, H., Graves, S. and Hooker, G., *Functional Data Analysis*, 2014, R package v 2.4.4.
12. Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap*, CRC Press, 1st edn, 1994.

ACKNOWLEDGEMENTS. This research is partially supported by Infosys research grant and TATA trust grant to CMI. A.L. acknowledges DST, New Delhi for research funding.

Received 22 September 2016; revised accepted 7 May 2018

doi: 10.18520/cs/v115/i5/920-929