# Data-adaptive Harmonic Decomposition and Stochastic Modeling of Arctic Sea Ice

Dmitri Kondrashov, Mickaël D. Chekroun, Xiaojun Yuan, and Michael Ghil

**Abstract** We present and apply a novel method of describing and modeling complex multivariate datasets in the geosciences and elsewhere. Data-adaptive harmonic (DAH) decomposition identifies narrow-banded, spatio-temporal modes (DAHMs) whose frequencies are not necessarily integer multiples of each other. The evolution in time of the DAH coefficients (DAHCs) of these modes can be modeled using a set of coupled Stuart-Landau stochastic differential equations that capture the modes' frequencies and amplitude modulation in time and space. This methodology is applied first to a challenging synthetic dataset and then to Arctic sea ice concentration (SIC) data from the U.S. National Snow and Ice Data Center (NSIDC). The 36-year (1979–2014) dataset is parsimoniously and accurately described by our DAHMs. Preliminary results indicate that simulations using our multilayer Stuart-Landau model (MSLM) of SICs are stable for much longer time intervals, beyond the end of the 21st century, and exhibit interdecadal variability consistent with past historical records. Preliminary results indicate that this MSLM is quite skillful in predicting September sea ice extent.

Dmitri Kondrashov
Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, U.S.A, e-mail: dkondras@atmos.ucla.edu

Mickaël D. Chekroun
Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, U.S.A e-mail: mchekroun@atmos.ucla.edu

Xiaojun Yuan
Lamont-Doherty Earth Observatory of Columbia University, Palisades, U.S.A e-mail: xyuan@ldeo.columbia.edu

Michael Ghil
Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, U.S.A, Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), École Normale Supérieure and and PSL Research University, F-75231 Paris Cedex 05, FRANCE. e-mail: ghil@atmos.ucla.edu

## 1 Data-adaptive harmonic (DAH) decomposition

The DAH decomposition introduced in [CK17] is a signal processing methodology that allows for a data-adaptive decomposition of power and phase spectra by adapting the time embedding approach to the study of time series introduced in [BK86, VG89, ET96] and its multivariate extensions. However, unlike other methodologies that rely on time embedding — such as Multichannel Singular Spectrum Analysis (M-SSA) [GAD+02] or Laplacian spectral analysis [GM12] — DAH uses integral-operator techniques that help decompose the original signal into narrow-banded signals; while data-adaptive, these elementary signals remain narrow-banded for each separate, discrete Fourier frequency.

At a practical level, the key feature of the DAH method is that it relies on the construction of matrices that exploit cross-correlations in a different way than found in standard statistical methods, such as in Principal Component Analysis (PCA) [Pre88]. As explained in [CK17] and discussed below, the eigenmodes associated with the matrices constructed by DAH exhibit a data-adaptive feature that shows up in their phase rather than in their shape. To wit, these modes form an orthogonal set of oscillating functions within the embedding window that is characterized by an interlacing of their zeros, as is the case for the eigenfunctions of Sturm-Liouville boundary-value problems for ordinary differential equations (e.g., [Har86]). While this interlacing property is intrinsic to the modes obtained by the DAH approach, the location of their zeros depends on the dataset at hand.

It is for this reason, that these modes are referred hereafter as *data-adaptive harmonic modes* (DAHMs). As a result, the elementary signals come in pairs, which are composed—as far as permitted by the available information and resolution—by such modes in exact phase quadrature. This property allows one to extract the aforementioned narrow-banded but amplitude-modulated time series, whose sum represents the original signal, as time series of DAH coefficients (DAHCs) obtained by projecting the input dataset onto the DAHMs. These features are at the core of identifying spatio-temporal oscillatory modes in the noisy synthetic dataset introduced in Sec. 2, as well as in the DAH analysis of a dataset of Arctic Sea Ice extent [FSHCC10] performed in Sec. 3; finally they permit the DAH-enabled nonlinear stochastic modeling of Sec. 4. Numerical details appear in Appendices 1 and 2.

## 2 DAH identification of spatio-temporal oscillatory modes

Here we evaluate our DAH methodology by applying it to a synthetic dataset designed as a testbed for the classical Prony problem of identifying "hidden periodicities" in a noisy environment (e.g., [Mar87]). Pisarenko harmonic decomposition [Pis73] is a well-known method of frequency estimation by using time-lagged correlations, and it assumes that a signal $x(n)$ consists of $p$ complex exponentials superimposed on white noise. However, the algorithm is restricted to the univariate case,

and its practical usefulness is somewhat limited due to the white-noise assumption and to the fact that $p$ must be known a priori.

The M-SSA methodology [GAD+02] also relies on time-lagged correlations, and it can be applied for identifying oscillatory modes without the limitations inherent in [Pis73]. A challenge for M-SSA, however, is the degeneracy problem in discriminating between oscillatory modes having similar energy but distinct temporal frequencies and spatial patterns; A. Groth and M. Ghil [GG11] introduced a suitably modified varimax rotation of the M-SSA modes that helps to deal with this shortcoming. To demonstrate the DAH capabilities for mode identification, we will rely on the synthetic dataset provided at `http://www.atmos.ucla.edu/tcd/ssa/guide/mssa/mssarot.html`, as part of the SSA-MTM Toolkit for time series analysis, `https://dept.atmos.ucla.edu/tcd/ssa-mtm-toolkit`; this dataset is used in the freeware Toolkit to illustrate the varimax-rotated M-SSA algorithm introduced in [GG11].

We thus consider a short and noisy spatio-temporal dataset describing the time evolution of a $d$-dimensional vector $\mathbf{y}(t_n) := (y_1(t_n), ..., y_d(t_n))$ over the interval $n = 1, ..., N$; here $d = 6$ and $N = 130$. The full dataset shown in Fig.1f consists of a coherent component $\mathbf{s}(t)$ embedded into *temporally correlated*, albeit spatially uncorrelated noise $\mathbf{r}(t)$:

$$\mathbf{y}(t_n) = (1-\nu)^{1/2}\,\mathbf{s}(t_n) + \nu^{1/2}\,\mathbf{r}(t_n)\,. \tag{1}$$

The coherent component $\mathbf{s}(t)$ in Fig.1e is the sum of the four oscillatory modes $x_k^i(t)$ with varying amplitude and phase across the six spatial channels, as shown in Figs.1(a–d):

$$s_k(t_n) = \sum_{i=1}^{4} x_k^i(t_n), \quad k = 1, \cdots, 6; \tag{2}$$

these modes are given by

$$x_k^i(t) = \left(\frac{\alpha_k^i}{2}\right)^{1/2} \sin(2\pi f_i t + \Phi_k^i), \quad k = 1, \cdots, 6, \;\; i = 1, \cdots, 4, \tag{3}$$

and each phase $\Phi_k^i$ is obtained independently as a random variable uniformly distributed in $[0, 2\pi]$.

The periodicities of the four oscillatory modes are not integer multiples of the sampling time nor of each other, while the respective frequencies $f_1 = 1/7.5, f_2 = 1/6, f_3 = 1/2.8$ and $f_4 = 1/2.3$ (in sampling units) are located in both the low-frequency and high-frequency part of the power spectrum. The amplitudes $\alpha_i^j$ are prescribed across the spatial channels so that 3 distinct modes contribute to each channel, albeit with different amplitudes; see Table 1. The random choice of the phases $\Phi_k^i$ in Eq. (3) results in arbitrary phase shifts across the spatial channels; see Fig. 1. The coefficient $\nu = 0.7$ in Eq. (1) guarantees that the noise component has larger variance than the signal; this fact is obvious from a comparison of the "clean" Fig. 1e with the "noisy" Fig. 1f), and it makes the identification problem that much more challenging.

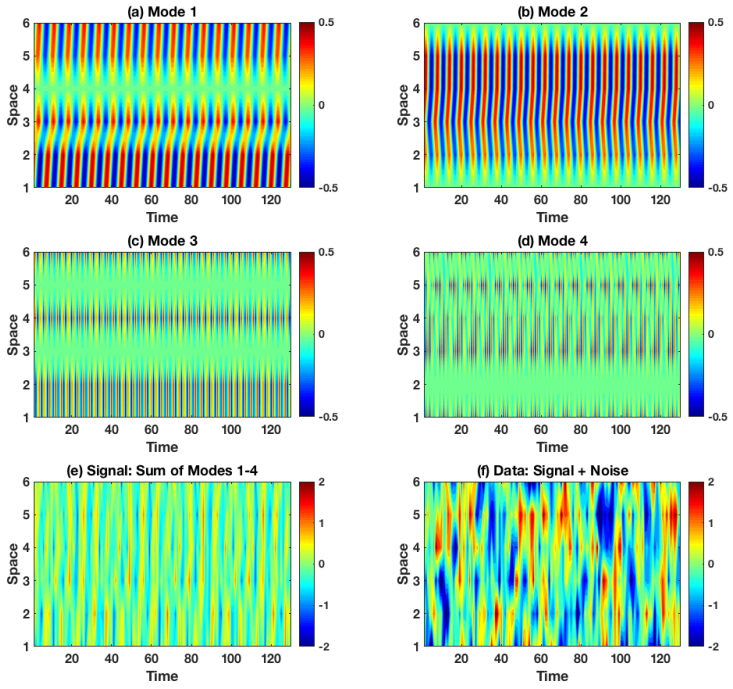


**Fig. 1** Multivariate spatio-temporal dataset representing six channels in space and 130 points in time: (a–d) four harmonic modes $\{\mathbf{x}^i(t) : i = 1,\ldots,4\}$ having fixed temporal frequencies but different amplitudes and phases in each of the six channels; see Eq. (3). Their sum $\mathbf{s}(t)$ defines the coherent component given by Eq. (2) shown in panel (e); (f) total dataset representing the sum of the coherent component $\mathbf{s}(t)$ and of the temporal red noise $r_k(t)$ in each of the $\{k = 1,\ldots,d\}$ channels; see text for details.

.

**Table 1** Amplitude modulation of the four oscillatory modes across six spatial channels; see Eq. (3). The index $k$ is for the channels, while the index $i$ is for the modes.

| $\alpha_k^i$ | $i=1$ | $i=2$ | $i=3$ | $i=4$ |
|---|---|---|---|---|
| $k=1$ | 0.4 | 0.0 | 0.3 | 0.3 |
| $k=2$ | 0.4 | 0.2 | 0.4 | 0.0 |
| $k=3$ | 0.3 | 0.3 | 0.0 | 0.4 |
| $k=4$ | 0.0 | 0.4 | 0.4 | 0.2 |
| $k=5$ | 0.2 | 0.4 | 0.0 | 0.4 |
| $k=6$ | 0.3 | 0.0 | 0.4 | 0.3 |

The block-Hankel matrix $\mathfrak{C}$ of the DAH decomposition (see Appendix 1) has $d = 6$ blocks of dimension $M' \times M'$, where $M'$ is the embedding dimension. The choice of $M'$ is based on two competing goals: (i) to obtain reliable estimates of autocorrelations from noisy and short datasets; and (ii) to resolve the dataset's frequency domain for identification purposes with sufficient accuracy. We chose $M' = 119$, which results in a total number $dM' = 714$ of DAH eigenvalues $\lambda_j$ and eigenvectors $\mathbf{E}_j$, i.e. $1 \le j \le dM'$.

Each of the DAH eigenvectors represents a data-adaptive spatio-temporal pattern associated with a fixed temporal frequency; the latter are equally spaced at intervals of $1/(M'-1)$ in the Nyquist interval $[0, 0.5]$. Moreover, each temporal frequency is associated with $d$ pairs of DAH eigenvalues that are opposite in sign but equal in absolute value, except at $f = 0$, where there is only one eigenvector per eigenvalue.

Figure 2 shows the DAH spectrum composed of the values $|\lambda_j|$ (red full circles), and obtained here for the synthetic dataset in Fig. 1f. The frequencies of the oscillatory modes that make up the coherent component are identified by eigenpairs located above the noisy background, and marked by the black arrows.

The time-embedded structure of these eigenvectors is shown in Fig. 3, with each pair $(E_j, E'_j)$ plotted by red and blue lines, respectively. This structure conveys information about the amplitude modulation across spatial channels, and the figure demonstrates that indeed the eigenvectors for each pair, except at zero frequency, are in phase quadrature, i.e. shifted by one quarter of the associated period.

The latter property is reminiscent of Fourier decomposition, based on sine and cosine pairs with the same periodicity, as well as of the similar property of oscillatory SSA eigenpairs [GAD+02]. The $k$-th spatial channel $\mathbf{E}_k^j$ of a particular multivariate DAHM — i.e., for a DAH with $d \ge 2$ — that is associated with a frequency

$$\omega_\ell = \frac{2\pi(\ell-1)}{M'-1}, \;\; \ell = 1, \cdots, \frac{M'+1}{2}. \tag{4}$$

can be analytically expressed—for each $1 \le j \le dM'$—as an oscillatory function in the embedding time-window variable $\tau$ as follows:

$$\mathbf{E}_k^j(\tau) = B_k^j(\omega_\ell)\sin(\omega_\ell \tau + \phi_k^j(\omega_\ell)), \quad 1 \le k \le d,\ 1 \le \tau \le M'; \tag{5}$$

here both amplitudes $B_k^j(\omega_\ell)$ and phases $\phi_k^j(\omega_\ell)$ are *data-adaptive* [CK17].

Moreover, the theory shows that the phases $\phi_k^j(\omega_l)$ for the modes in each pair are shifted by one fourth of the period, i.e. DAHMs are in exact phase quadrature, as for sine–cosine pairs, but in a data-adaptive fashion, encapsulated into the phase. Indeed, as proved in [CK17], in the case of univariate time series, the DAH modes provide the phase spectrum contained in each frequency $\omega_\ell$ (given in (4)) via the analytical formula:

$$\Phi(\omega_\ell) = \arg(\lambda_j \widehat{\mathbf{E}}^j(\omega_\ell)) - \arg(\overline{\widehat{\mathbf{E}}^j}(\omega_\ell)),\ 1 \le j \le dM', \tag{6}$$

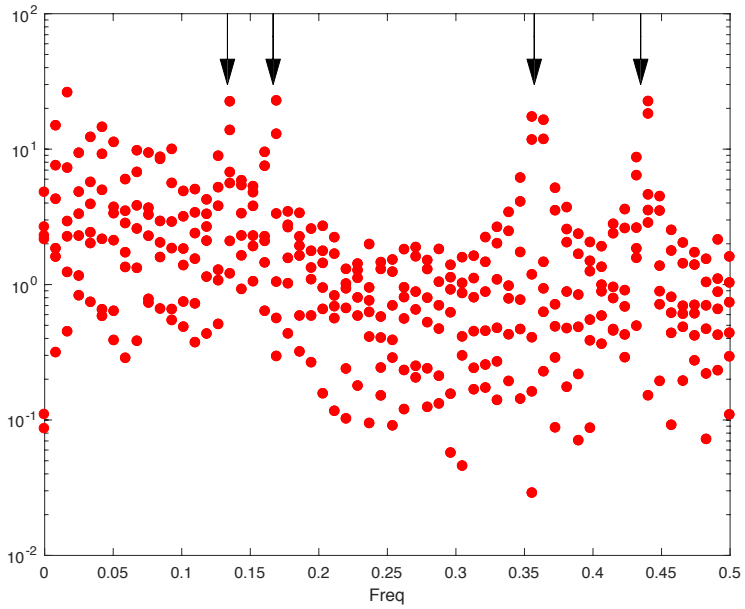


**Fig. 2** DAH spectrum of the noisy dataset in Fig. 1f. Each red full circle corresponds to a pair $\pm|\lambda_j|$ with distinct eigenvectors $(\mathbf{E}_j, \mathbf{E}'_j)$; the latter represent the same temporal frequency $f_j$ but are time-shifted so as to be in phase quadrature, cf. Fig. 3 below. Arrows point to the temporal frequencies of four oscillatory modes that do correspond to those shown in Figs. 1(a–d). The frequencies of the DAH eigenvectors are equally spaced between 0 and 0.5, and the total number of DAH pairs in each frequency bin is equal to the number of channels $d = 6$ in the dataset. The data-adaptive DAH modes describe amplitude and phase modulation between the spatial channels and are shown in Fig. 3; they do permit the faithful reconstruction of the reference modes in Figs. 1(a–d), as shown in Figs. 4(a–d) below.

where $\widehat{\mathbf{E}}^j$ and $\overline{\widehat{\mathbf{E}}^j}$ denote respectively the Fourier transform of $\mathbf{E}^j$ and its complex conjugate.

The precise information about amplitude and phase modulation of the oscillatory modes captured by the DAHMs allows one to perform highly accurate reconstructions in the space-time domain, cf. Eq. (14) in Appendix 1 below.Figure 4 shows the space-time patterns of the harmonic reconstruction components (HRCs) given by Eq. (15); these patterns are obtained using all the DAH pairs in the frequency bins that contain the target periodicities $f_1, f_2, f_3$ and $f_4$. These patterns match quite well in frequency and phase those of the reference coherent components in Figs. 1a–d, although they do underestimate their amplitude as a consequence of the large noise level. In fact, the normalized root-mean-square (rmse) error, averaged over time and space, is roughly 0.5 for all four modes.

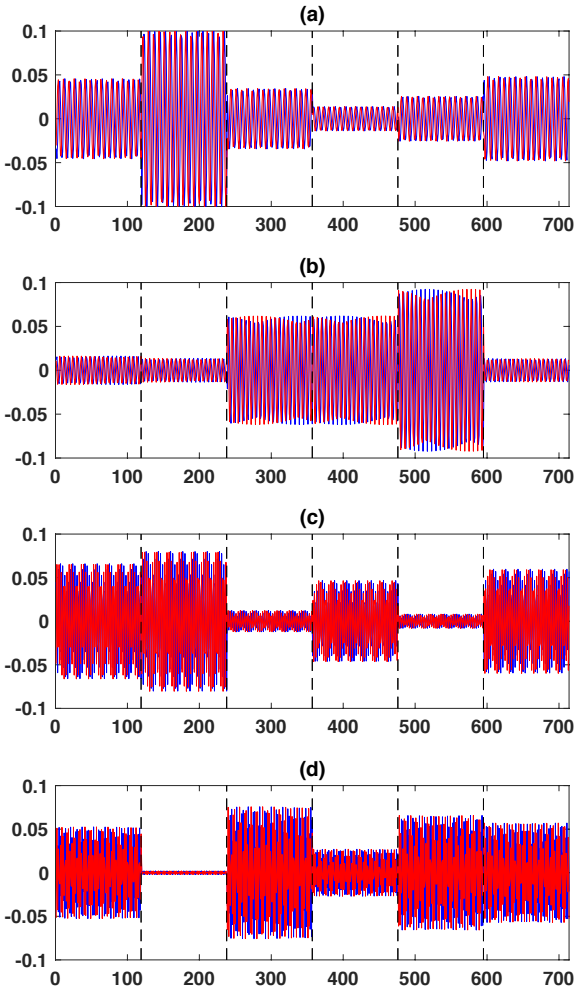


**Fig. 3** Eigenvectors $(\mathbf{E}_j, \mathbf{E}'_j)$ of the leading spectral DAH pairs for the four frequencies that are closest to those of the four spatio-temporal oscillatory modes in Figs. 1(a–d), i.e. $f_1, f_2, f_3$ and $f_4$, respectively; see Eq. (3). The $x$-axis represents the embedding dimension $dM'$, while the vertical dashed lines mark six $M'$-long segments that correspond to $d = 6$ spatial channels. For each spatial channel, the eigenvectors of a given frequency convey different phases and are shifted by a quarter of the associated period, i.e. they are in exact phase quadrature.

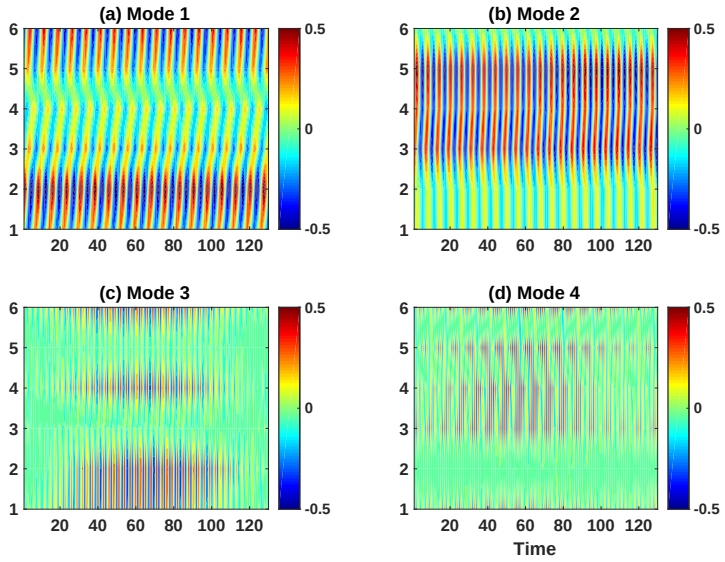


**Fig. 4** DAH reconstruction associated with the frequencies of the four dominant oscillatory DAH pairs, as marked by the arrows in Fig. 2, and obtained by using the DAH pairs in the corresponding frequency bins. The resulting patterns match reasonably well the reference patterns shown in Figs. 1(a–d).

These results show that DAH does correctly detect the temporal frequencies of distinct oscillatory modes in a very noisy multichannel dataset. Moreover, it also captures fairly well their distinct phase and amplitude across the spatial channels.

## 3 DAH decomposition of Arctic sea ice concentrations (SICs)

Decline in Arctic Sea ice extent is an area of active scientific research with profound climatic and socio-economic implications, both negative—on global temperatures—and positive—by facilitating navigation in polar waters [SRF+16]. The key variable of interest to study Arctic Sea ice dynamics is so-called sea ice concentration (SIC), which measures the relative amount of reference area covered by ice at a given location; SIC is given in percentage points (0% – 100%). An important indicator of Arctic sea ice conditions is the so-called Sea Ice Extent (SIE), defined as the total surface area of the Arctic region having SIC greater than 15%.

The widely used Sea Ice Index (SII) from the National Snow and Ice Data Center (NSIDC) relies exclusively on passive microwave measurements, which provide a 35-year–long dataset of daily SICs from 1979 to the present. The satellite observations are automatically processed by the National Aeronautics and Space Adminis-

tration (NASA) Team [CPGZ96] and Bootstrap [Com14] algorithms to create daily SIC maps; both algorithms have their own biases and limitations.

We have used the monthly NSIDC dataset for SIC over the Jan. 1979–Dec. 2014 interval, available on a 25 km $\times$ 25 km polar stereographic grid; this dataset is based on the Bootstrap algorithm [Com14]. The data version used has been coarse-grained onto a $2^\circ \times 0.5^\circ$ grid, representing 7400 spatial degrees of freedom each month and $N = 432$ monthly maps.

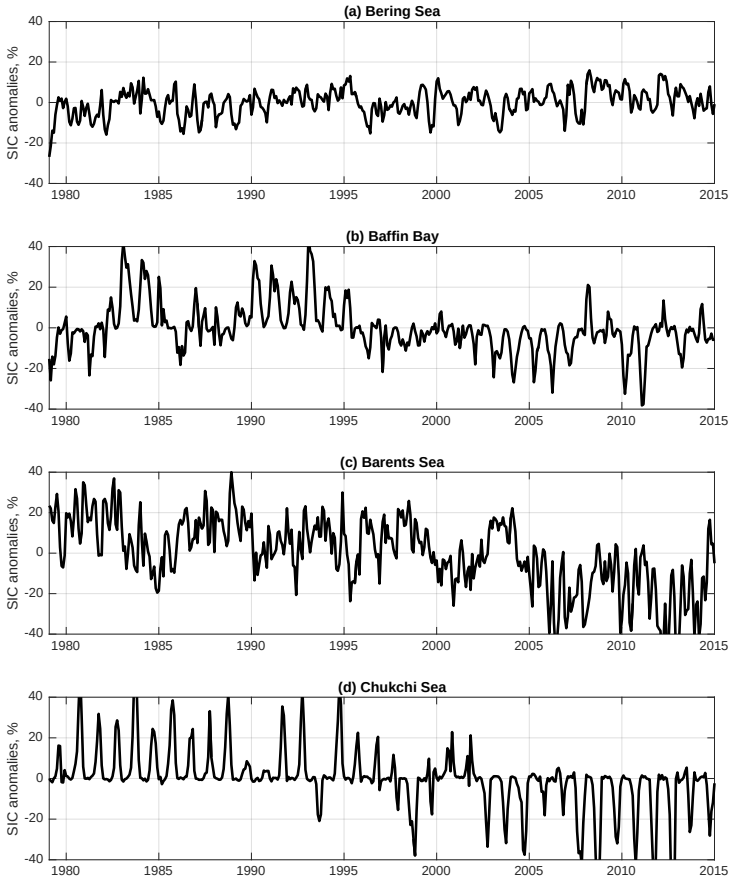


**Fig. 5** Monthly time series for sea ice concentration (SIC) anomalies in key Arctic regions; see text for details. (a–d) Bering Sea (182°E–192°E, 58°N–62°N); Baffin Bay (298°E–304°E, 61°N–66°N); Barents Sea (34°E–54°E, 76°N–80°N); and Chuckhi Sea (190°E–210°E, 72°N–76°N).

First, we removed the seasonal cycle by computing SIC anomalies with respect to each calendar month. Figure 5 shows that the dynamics of SIC anomalies is very different in key Arctic regions, namely the Bering Sea, Baffin Bay, Barents Sea, and Chuckhi Sea. In particular, SIC anomalies in the Baffin Bay and Chuckhi Sea are dominated by the seasonal cycle and a strong downward trend, while internal dynamics is more prominent in the Bering and Barents Seas.

Figure 6a shows that the variability of SIC anomalies is mostly concentrated in the marginal seas of the Arctic Ocean, while it is very small over the North Pole, where the sea remains ice-covered at all times. To extract the dominant modes of SIC variability, empirical orthogonal function (EOF) decomposition [Pre88] was applied to the dataset. The 12 leading EOFs account for 82% of SIC anomaly variance: excluding the Bering Sea, which is only in very limited contact with the Arctic Ocean, these EOFs capture most of the variance in the marginal seas, cf. Fig. 6b.

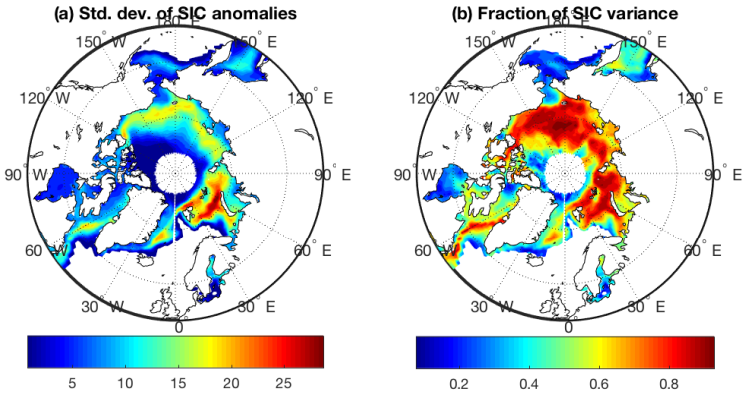


**Fig. 6** Spatial distribution of SIC variability. (a) Standard deviation of SIC anomalies; and (b) fraction of SIC variance captured by the 12 leading EOFs of SIC anomalies. Color bars are in percentage units and nondimensional, in (a) and (b), respectively.

Figure 7 shows the corresponding time series of principal components (PCs). The trend component is most prominent in the leading pair of PCs, although it is present, to a lesser extent, in other PCs as well. Moreover, the trend component strongly depends on the calendar month, being more pronounced in fall than in winter; hence there is also strong annual variability in the 1st and 2nd PC, superimposed on the

trend. To summarize, SIC PCs exhibit a complex mixture of annual cycle, intraseasonal, interannual and long-term time scales; this complexity represents a serious challenge for data-driven analysis and modeling techniques, but will be successfully addressed by DAH decomposition.

Figure 8 shows the multivariate DAH spectrum of $d = 12$ PCs for the SIC dataset, with an embedding dimension of $M' = 59$ months. Each full circle in this figure is associated with a pair of DAHMs, except at zero frequency, where the modes are not paired, cf. Eq. (5). The seasonally dependent trend is clearly isolated by the pairs associated with annual-cycle harmonics and located well above the continuous background.

The spatio-temporal patterns of the DAH modes shown in the left and center panels of Fig. 9 reveal useful dynamical information on the combined evolution and mutual influence of SIC's PCs in particular frequency bands. For example, the dominant variability patterns— i.e. those corresponding to the pair having the largest $|\lambda_j|$ at a particular frequency—convey in-phase, out-of-phase and time-lagged influences between different PCs. The DAHMs associated with the *same* frequency and ranked top-to-bottom by their DAH spectral value behave in a similar fashion, as shown in Fig. 10 for the 12-month periodicity. Note that the DAHMs are always in phase-quadrature, except at zero frequency.

On the other hand, although the DAH coefficients $A_j$ are not formally orthogonal in time — see Eq. (13) and its discussion in Appendix 1 — they also exhibit a certain phase-quadrature relationship that depends on whether the window $M$ is sufficiently large to resolve the decay of temporal correlations of a given dataset. Typically, the larger $M$ (subject to the length of the record), the more apparent is the phase quadrature between a pair of DAHCs associated with the same frequency.

Shown in the right panels of Figs. 9 and 10, the DAHCs constituting a given $\mathbf{A}_j$-pair account for narrow-band temporal information contained at the characteristic frequency associated with the respective $\mathbf{E}_j$-pair. The latter pairs are shown in the left and center panels of these two figures, respectively, and they reflect differences in amplitude and a shift of, approximately, a quarter of a period. As we can see, the phase-quadrature property of the DAHCs is satisfied to a reasonable degree, which bodes well for the success of the stochastic-modeling approach described in the next section.

## 4 Stochastic modeling of Arctic SICs

The recent *Multilayer Stochastic Model* (MSM) framework introduced in [KCG15] emphasizes the key role of nonlinear, stochastic and non-Markovian effects in deriving data-driven closure models. Such models have been shown to posses considerable skill in simulating and predicting the main dynamical features of a targeted spatio-temporal field, given either as the output of a high-end geophysical model or as a set of observations. The MSM approach generalizes various multilevel inverse models, including Empirical Model Reduction (EMR) [KKG05, KKG09]: it allows

for greater flexibility in the choice of the nonlinear predictors, while ensuring stable asymptotic behavior, such as the existence of a global random attractor [CSG11]; see Theorem 3.1 and Corollary 3.2 in [KCG15].

However, if the input dataset is not large enough and exhibits a mixture of several time scales, this approach may propose numerous predictors that require one to estimate too many model coefficients, a situation that makes accurate and stable estimates quite difficult. Alternative algorithms are thus called for, and DAH decomposition provides such an alternative. We show here, in the context of Arctic sea ice modeling, that an appropriate change of the basis—in a data-adaptive manner—reduces the data-driven modeling effort to elemental MSMs stacked by frequency, and requires only estimating a fixed and much smaller number of coefficients.

These elemental models fall into the class of networks of linearly coupled Stuart-Landau oscillators [ZLS+16], which may include memory terms [SLD+12] and are described below. Given a sequence of partial observations of a dynamical-model simulation, the DAHCs allow one to recast these observations so that they can be reproduced by a simple stochastic model. Such a model can be inferred within a universal parametric family, provided, roughly speaking, that the window whether the window $M$ is sufficiently large to resolve the decay of temporal correlations of a given dataset, as discussed in Appendix 1.

Stuart-Landau (SL) models with additive noise form a generic class of models that capture (i) the frequency $f$ and (ii) the amplitude modulations of the $A_j$'s corresponding to a given narrow-band DAHC pair, denoted by $(x(t), y(t))$:

$$\dot{z} = (\mu + i\gamma)z - (1 + i\beta)|z|^2 z + \varepsilon_t, \;\; z \in \mathbb{C}; \tag{7}$$

here $z(t) = x(t) + iy(t)$ ($i^2 = -1$) and the real parameters $\mu, \gamma$ and $\beta$, as well as the properties of the driving noise $\varepsilon_t = (\varepsilon_j^x, \varepsilon_j^y)$, are estimated from the time history of $z(t)$ by the aforementioned MSM approach. To reproduce the global phase coherence of the collective behavior of $d$ DAH pairs $(x_j(t), y_j(t))$, at a given frequency $f \neq 0$, requires an appropriate dynamical coupling between individual SL oscillators, along with taking into account the temporal and spatial cross-pair correlations in the driving noise $\varepsilon_t$; see Appendix 2 and Eq. (*MSLM*) there.

Thus, for each frequency $f$, the 12 associated pairs of temporal DAHCs are modeled by Eq. (*MSLM*). First, the model coefficients can be estimated *in parallel* for each frequency, i.e. by successive pairwise regressions, subject to linear constraints on $\beta_j(f), \alpha_j(f)$ and $\sigma_j(f)$ that impose the necessary model structure in Eq. (*MSLM*) for each $(x_j, y_j)$ pair; these constraints entail antisymmetry for the linear part, without the coupling terms, as well as equal and nonpositive values $\sigma_j(f) \leq 0$ to ensure asymptotic stability. Hence the overall number of independent coefficients to estimate is fixed and relatively small for each $(x_j, y_j)$ pair; e.g., the main layer of Eq. (*MSLM*) involves estimation of $3 + 4(d-1) = 47$ coefficients from the $2N' = 748$ DAH-processed Arctic SIC observations, over the full time interval 1979–2014; see Appendix 1 for the definition of $N' = N - M' + 1$, with the window width $M' = 59$ months. Extra layers are added as needed until the regression residuals for the last layer can be approximated by white noise, according to the stopping

test described in [KCG15, Appendix A]; these layers convey temporal correlations in the stochastic forcing $\varepsilon_t$ on the main layer of the model for $(x_j, y_j)$.

Second, the DAH-MSLMs are run *in parallel* across the frequencies by the same white-noise realization in the last layer of the model, which represents a dynamical mechanism for coupling between different frequencies. Finally, the simulated time series of the temporal DAHCs are converted back to the phase space of the SIC dataset's PCs, by convolution with the spatio-temporal DAHM's.

Despite the limited amount of available data and their nonstationarity, Figs. 11 and 12 show very good modeling skill in reproducing the complex structure of the autocorrelation functions (ACFs) of the SIC dataset's PCs, as simulated by the optimal DAH-MSLM model with $M' = 59$ and having three additional layers in Eq. (*MSLM*) to model the noise $\varepsilon_t$. The model also captures sufficiently well skewness & kurtosis of the probability density functions (PDFs), although it is more challenging to capture the bumps in the PDFs' "tails," due to the record's shortness.

Figures 13 and 14 show the evolution in time of the leading PCs of two stochastic ensemble members, as simulated by our DAH-MSLM model and initialized in January 1979. These extended, 129-year–long simulations demonstrate that our optimized stochastic-dynamic model agrees well with the existing 36-year–long SIC record, is numerically stable for much longer times, and displays interesting dynamical behavior such as multidecadal variability in PC-1. Such variability has been documented by J. Walsh and W. Chapman [WC15] in their reconstruction of sea ice extent anomalies from historical records.

One reason for the success of our model's simulations relies on the ability of the DAH approach to extract modulated time series of DAHCs that are narrow-banded in the frequency domain and exhibit phase quadrature in the time domain. Another important reason is that the class of MSLMs introduced herein is intrinsically well adapted to the modeling of such time series.

It is worth mentioning that the less narrow-banded the DAHCs, the worse their modeling using MSLM. For the Arctic Sea Ice dataset of [Com14], as represented by the SIC PCs, the DAH decomposition provides just the right time series of DAHCs for the MSLM modeling approach to be efficient; see Figs. 9 and 10.

Our DAH-MSLM model is able to produce a remarkable near-synchronization of the simulations with observations during the first four years that start with January 1979. This approximate synchronization holds for almost every noise realization, as shown, for instance, in Fig. 15 for one ensemble member, using a particular noise realization: plotted in the figure are September SIC anomalies for 1979–1982 in gridded physical space, with the maps of the observations in the left column and the simulations in the right one. The match between simulation and observation is visually excellent and only starts deteriorating in September 1982. The potential predictive skill of our DAH-MSLM model suggested by these plots, implies highly promising potential of developed approach for real-time forecasting of September SIE.

Indeed, this potential forecast skill has been tentatively confirmed by the present authors in [KCG17] by using the Multisensor Analyzed Sea Ice Extent (MASIE) dataset [FSHCC10] for the Sea Ice Prediction Network (SIPN, `http://www.`

arcus.org/sipn). Our DAH-MSLM model's real-time SIE forecast for September 2016 [SBWG+15, HS16] outperformed most other statistical models and physics-based models in the SIPN network. In 2016, the multimodel-median September SIPN estimate in August was $4.4 \cdot 10^6$ km$^2$, with a quartile range of $4.2 - 4.7 \cdot 10^6$ km$^2$, vs. the actual observed value of $4.72 \cdot 10^6$ km$^2$. The real-time DAH-MSLM August prediction for SIPN's 2016 September Outlook was $4.79 \cdot 10^6$ km$^2$.

**Acknowledgements** The authors would like to acknowledge Andreas Groth for developing the synthetic dataset in the SSA-MTM Toolkit example of varimax-rotated M-SSA (http://www.atmos.ucla.edu/tcd/ssa/guide/mssa/mssarot.html); it is this dataset that was utilized in Sec. 2.

Preliminary results of this research were reported at "30 Years of Nonlinear Dynamics in Geosciences" conference in Rhodes, Greece, July 2017. This research was supported by ONR's Multidisciplinary Research Initiative (MURI) grants N00014-12-1-0911 and N00014-16-1-2073, and by the National Science Foundation grants OCE-1243175 and DMS-1616981.

## Appendix 1. Details on the DAH decomposition

The DAH modes (DAHMs) are obtained as follows. First, we estimate from a given $d$-channel time series $\mathbf{X}(t_n) = (X_1(t_n), \ldots, X_d(t_n))$, $n = 1, \cdots, N$, the *cross-correlation coefficient* (CCF) $\rho_\tau^{(p,q)}$ at lag $\tau$ between channels $p$ and $q$, where $-M+1 \leq \tau \leq M-1$. In spectral analysis, it is common to refer to $M$ as the window width.

Next, we form the following Hankel matrix:

$$\mathbf{H}^{(p,q)} = \begin{pmatrix} \rho_{-M+1}^{(p,q)} & \rho_{-M+2}^{(p,q)} & \cdots & \rho_0^{(p,q)} & \rho_1^{(p,q)} & \cdots & \rho_{M-1}^{(p,q)} \\ \rho_{-M+2}^{(p,q)} & \iddots & \iddots & \iddots & \iddots & \iddots & \rho_{-M+1}^{(p,q)} \\ \vdots & \iddots & \iddots & \iddots & \iddots & \iddots & \rho_{-M+2}^{(p,q)} \\ \rho_0^{(p,q)} & \iddots & \iddots & \iddots & \rho_{-M+1}^{(p,q)} & \iddots & \vdots \\ \rho_1^{(p,q)} & \iddots & \iddots & \iddots & \rho_{-M+2}^{(p,q)} & \iddots & \rho_0^{(p,q)} \\ \vdots & \rho_{M-1}^{(p,q)} & \rho_{-M+1}^{(p,q)} & \iddots & \iddots & \iddots & \vdots \\ \rho_{M-1}^{(p,q)} & \rho_{-M+1}^{(p,q)} & \rho_{-M+2}^{(p,q)} & \cdots & \rho_0^{(p,q)} & \cdots & \rho_{M-2}^{(p,q)} \end{pmatrix}. \tag{8}$$

Equivalently, this matrix can be viewed as a left circulant matrix formed from the $(2M-1)$-dimensional row $r = (\rho_{-M+1}^{(p,q)}, \cdots, \rho_0^{(p,q)}, \cdots, \rho_{M-1}^{(p,q)})$, i.e.:

$$\mathbf{H}^{(p,q)} = l\text{-circ}(\rho_{-M+1}^{(p,q)}, \cdots, \rho_{-1}^{k,k'}, \rho_0^{(p,q)}, \rho_1^{(p,q)} \cdots, \rho_{M-1}^{(p,q)}); \tag{9}$$

in other words, the rows of $\mathbf{H}^{(p,q)}$ are obtained by successive shifts to the left by one position, starting from $r$ as a first row. Finally, we consider the block-Hankel matrix $\mathfrak{C}$ formed by $d^2$ blocks of size $(2M-1) \times (2M-1)$, each given according to

$$\begin{aligned} \mathfrak{C}^{(p,q)} &= \mathbf{H}^{(p,q)}, \text{ if } 1 \le p \le q \le d, \\ \mathfrak{C}^{(p,q)} &= \left(\mathbf{H}^{(q,p)}\right), \text{ otherwise.} \end{aligned} \tag{10}$$

Note that $\mathfrak{C}$ is symmetric by construction due to symmetry of its building blocks $\mathbf{H}^{(p,q)}$, i.e. $\mathfrak{C}^{(p,q)} = \mathfrak{C}^{(q,p)}$, and hereafter we use $M' = 2M-1$ for concision, reindexing the string $\{-M+1,\ldots,M-1\}$ from 1 to $M'$ as necessary.

The DAH eigenpairs $(\lambda_j, \mathbf{E}^j)$, with $1 \le j \le dM'$, reveal useful information about the variability contained in the multivariate time series. In contrast to other data-adaptive methods built from cross-correlations, each of the DAH eigenvectors $\mathbf{E}^j$ represents a data-adaptive spatio-temporal pattern naturally associated with a Fourier frequency $\omega_l$ given by

$$\omega_\ell = \frac{2\pi(\ell-1)}{M'-1}, \;\; \ell = 1,\cdots,\frac{M'+1}{2}. \tag{11}$$

These frequencies are equally spaced within the Nyquist interval $[0, 0.5]$ with a resolution of $1/(M'-1)$, essentially given by the embedding dimension $M$.

Each temporal frequency $\omega_\ell$ is associated with $d$ pairs of DAH eigenvalues $\pm\lambda_j$ that are opposite in sign but equal in absolute value, except at zero frequency, where there is only one eigenvector per eigenvalue, for a total of $2d(M-1)+d$ eigenvalues. The association between a particular frequency and a given DAHM is obtained by counting zero-crossings $\delta_j$ across the window width $M$ for all channels:

$$\delta_j = \sum_{k=1}^{d}\sum_{\tau=1}^{M'-1}\left(1 - \text{sign}(\mathbf{E}_k^j(\tau)\mathbf{E}_k^j(\tau+1))\right), \;\; 1 \le j \le dM'. \tag{12}$$

One can thus assign a frequency that is in one-to-one correspondence to $\delta_j$. In Eq. (12), $\mathbf{E}_k^j$ denotes the $k$-th spatial component of the DAHM, $\mathbf{E}^j$. One can then rank the DAHMs from the lowest to the highest frequency by simply looking at their number of sign changes. As shown in [CK17], the corresponding fraction of the energy they capture is given by $|\lambda_j|$, up to a scaling factor.

By analogy with M-SSA [GAD+02], the multivariate dataset $\mathbf{X}$ can be projected onto the orthogonal set formed by the $\mathbf{E}^j$'s, to obtain the DAH expansion coefficients (DAHCs):

$$A_j(t) = \sum_{\tau=1}^{M'}\sum_{k=1}^{d} X_k(t+\tau-1)\mathbf{E}_k^j(\tau), \tag{13}$$

where $t$ varies from 1 to $N' = N - M' + 1$.

Although the DAHCs are not formally orthogonal in time, they also exhibit a phase-quadrature relationship that depends on whether the window $M$ is sufficiently large to resolve the decay of temporal correlations of a given dataset. Typically, the larger $M$ (subject to the length of the record), the more apparent is the phase quadrature between a pair of DAHCs associated with the same frequency.

Furthermore, any subset $\mathbf{B} \subset \mathbf{A}$ of DAHCs, as well as the full set $\mathbf{A}$, can be convolved with associated $\mathbf{E}_j$'s, for partial or full reconstruction of the original data,

respectively. The transformation between $\mathbf{X}$ and $\mathbf{A}$ is unitary, i.e., there is no loss of variance. Thus, the $j$th RC at time $t$ for channel $k$ is given by:

$$R_k^j(t) = \frac{1}{M_t} \sum_{\tau=L_t}^{U_t} A_j(t-\tau+1)\mathbf{E}_k^j(\tau). \tag{14}$$

The normalization factor $M_t$ equals $M'$, except near the ends of the time series [GAD+02], and the sum of all the RCs recovers the original time series.

It is also useful to consider harmonic reconstruction components (HRCs), namely a sum of $d$ RC pairs corresponding to a particular frequency $\omega_\ell \neq 0$:

$$R_k^{\omega_\ell}(t) = \sum_{j \in \mathscr{J}_\ell} R_k^j(t), \tag{15}$$

where $\mathscr{J}_\ell$ denotes the set of all the indices $j$ associated with the frequency $\omega_\ell$. By construction, for each nonzero frequency, this set is constituted by $2d$ elements.

## Appendix 2. Details on the MSLM modeling

As discussed in Sec. 4, the DAHMs extract harmonic components of variability that allow for a reduction of the data-driven modeling effort to a simple class of elemental multilayer stochastic models (MSMs: [KCG15]); these MSMs are stacked by frequency and only coupled at different frequencies by the same noise realization.

In the simplest case of one layer for the modeled noise, this construction leads to stochastic models of the form:

$$\begin{aligned}
\dot{x_j} &= \beta_j(f)x_j - \alpha_j(f)y_j + \sigma_j(f)x_j(x_j^2+y_j^2) + \sum_{i\neq j}^{d} b_{ij}^x(f)x_i + \sum_{i\neq j}^{d} a_{ij}^x(f)y_i + \varepsilon_j^x,\\
\dot{y_j} &= \alpha_j(f)x_j + \beta_j(f)y_j + \sigma_j(f)y_j(x_j^2+y_j^2) + \sum_{i\neq j}^{d} a_{ij}^y(f)x_i + \sum_{i\neq j}^{d} b_{ij}^y(f)y_i + \varepsilon_j^y,\\
\dot{\varepsilon}_j^x &= L_{11}^j(f)x_j + L_{12}^j(f)y_j + M_{11}^j(f)\varepsilon_j^x + M_{12}^j(f)\varepsilon_j^y + \\
&\qquad Q_{11}^j(f)\dot{W}_1^j + Q_{12}^j(f)\dot{W}_2^j + \sum_{i\neq j}^{d}\sum_{k=1}^{2} Q_{1k}^i(f)\dot{W}_k^i,\\
\dot{\varepsilon}_j^y &= L_{21}^j(f)x_j + L_{22}^j(f)y_j + M_{21}^j(f)\varepsilon_j^x + M_{22}^j(f)\varepsilon_j^y + \\
&\qquad Q_{21}^j(f)\dot{W}_1^j + Q_{22}^j(f)\dot{W}_2^j + \sum_{i\neq j}^{d}\sum_{k=1}^{2} Q_{2k}^i(f)\dot{W}_k^i.
\end{aligned} \tag{MSLM}$$

In (*MSLM*), the index $j$ varies in the set of indices $\mathscr{J}_f$ associated with a single frequency $f$, determined by the zero-crossings of the corresponding $\mathbf{E}^j$'s. When $f \neq 0$, this set consists of $d$ elements. In practice $f = \omega_\ell/(2\pi)$ is determined by a

Fourier frequency $\omega_\ell$ given in Eq. (11). The $W_k^j$'s with $k$ in $\{1,2\}$ and $j$ in $\{1,\cdots,d\}$ form $2d$ independent Brownian motions.

We call these models *multilayer stochastic Stuart-Landau models* (MSLM). At a given frequency $f$, the $d$ pairs are linearly coupled as indicated by the terms in the sums apparent in the $x_j$- and $y_j$-equations. In (*MSLM*) and for a given pair indexed by $j$, the noise term $(\varepsilon_j^x, \varepsilon_j^y)$ is modeled by means of linear dependencies involving only $(\varepsilon_j^x, \varepsilon_j^y)$, on the one hand, and the $j$-th pair $(x_j, y_j)$, on the other.

Obviously, for a given pair, and following [KCG15], more layers can be added as needed to (*MSLM*), when the noise term $(\varepsilon_j^x, \varepsilon_j^y)$ at the first level is not white. In this case, the extra layers will depend linearly on the $j$-th pair $(x_j, y_j)$, and on the noise residuals from the previous layers. The sums in the $\varepsilon_j^x$- and $\varepsilon_j^y$-equations take into account "spatial" correlations between the pairs, at the level of the noise. Note that for the null frequency, $f \equiv 0$, there are exactly $d$ modes that are not paired, and they are modeled by a linear multilayer stochastic model as in [KCG15].

Note that equations (*MSLM*) can be generalized further by allowing coupling of $(x_j, y_j)$ pairs at neighboring frequencies, which can be useful for certain applications where cross-frequency interactions are important. Equations (*MSLM*) are discretized in time and integrated numerically forward from initial conditions that respect the initialization procedure described in [KCG15, Appendix B].

## References


BK86. D. S. Broomhead and G. P. King, *Extracting qualitative dynamics from experimental data*, Physica D: Nonlinear Phenomena **20** (1986), no. 2, 217–236.

CK17. M. D. Chekroun and D. Kondrashov, *Data-adaptive harmonic spectrum and stochastic-dynamic inverse Stuart-Landau models*, In preparation (2017).

Com14. J. C. Comiso, *Bootstrap Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS. Version 2 [Northern Hemisphere daily data]*, Digital media, NASA National Snow and Ice Data Center, Distributed Active Archive Center, Boulder, Colorado USA, 2014.

CPGZ96. D. Cavalieri, C. Parkinson, P. Gloersen, and H. J. Zwally, *Updated Yearly Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, 1979–2010*, Digital media, National Snow and Ice Data Center, Boulder, Colo., 1996.

CSG11. M. D. Chekroun, E. Simonnet, and M. Ghil, *Stochastic climate dynamics: Random attractors and time-dependent invariant measures*, Physica D **240** (2011), 1685–1700.

ET96. J. B. Elsner and A. A. Tsonis, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, Springer Science & Business Media, 1996.

FSHCC10. F. Fetterer, M. Savoie, S. Helfrich, and P. Clemente-Colón, *Multisensor Analyzed Sea Ice Extent - Northern Hemisphere*, Digital media, National Snow and Ice Data Center, Boulder, Colo., 2010.

GAD+02. M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, *Advanced spectral methods for climatic time series*, Rev. Geophys. **40** (2002).

GG11. A. Groth and M. Ghil, *Multivariate singular spectrum analysis and the road to phase synchronization*, Phys. Rev. E **84** (2011), 036206.

GM12. D. Giannakis and A. J. Majda, *Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability*, Proc. Natl. Acad. Sci. USA **109** (2012), no. 7, 2222–2227.

Har86. P. Hartman, *Ordinary Differential Equations*, 2nd ed., Classics in Applied Mathematics, vol. 38, SIAM, 1986.

HS16. L. C. Hamilton and J. Stroeve, *400 predictions: the SEARCH Sea Ice Outlook 2008–2015*, Polar Geography **39** (2016), no. 4, 274–287.

KCG15. D. Kondrashov, M. D. Chekroun, and M. Ghil, *Data-driven non-Markovian closure models*, Physica D **297** (2015), 33–55.

KCG17. D. Kondrashov, M. D. Chekroun, and M. Ghil, *Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent*, In preparation (2017).

KKG05. S. Kravtsov, D. Kondrashov, and M. Ghil, *Multi-level regression modeling of nonlinear processes: Derivation and applications to climatic variability*, J. Climate **18** (2005), no. 21, 4404–4424.

KKG09. _____, *Empirical model reduction and the modeling hierarchy in climate dynamics and the geosciences*, Stochastic Physics and Climate Modeling (T. N. Palmer and P. Williams, eds.), Cambridge University Press, 2009, pp. 35–72.

Mar87. S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall, 1987.

Pis73. V. F. Pisarenko, *The retrieval of harmonics from a covariance function*, Geophys. J. Intl. **33** (1973), no. 3, 347–366.

Pre88. R. W. Preisendorfer, *Principal component analysis in meteorology and oceanography*, Elsevier, New York, 425 pp., 1988.

SBWG+15. J. Stroeve, E. Blanchard-Wrigglesworth, V. Guemas, S. Howell, F. Massonnet, and S. Tietsche, *Improving predictions of Arctic sea ice extent*, Eos, Trans. AGU **96** (2015).

SLD+12. A. A. Selivanov, J. Lehnert, T. Dahms, P. Hövel, A. L. Fradkov, and E. Schöll, *Adaptive synchronization in delay-coupled networks of Stuart-Landau oscillators*, Phys. Rev. E **85** (2012), 016201.

SRF+16. M. Sigmond, M. C. Reader, G. M. Flato, W. J. Merryfield, and A. Tivy, *Skillful seasonal forecasts of Arctic sea ice retreat and advance dates in a dynamical forecast system*, Geophysical Research Letters **43** (2016), no. 24, 12,457–12,465.

VG89. R. Vautard and M. Ghil, *Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series*, Physica D: Nonlinear Phenomena **35** (1989), no. 3, 395–424.

WC15. J. Walsh and W. Chapman, *Variability of sea ice extent over decadal and longer timescales*, Climate Change: Multidecadal and Beyond (C. P. Chang, M. Ghil, M. Latif, and J. M. Wallace, eds.), World Scientific Publ. Co./Imperial College Press, 2015, pp. 203–217.

ZLS+16. A. Zakharova, S. Loos, J. Siebert, A. Gjurchinovski, J. C. Claussen, and E. Schöll, *Controlling chimera patterns in networks: interplay of structure, noise, and delay in control of self-organizing nonlinear systems*, Control of Self-Organizing Nonlinear Systems (P. Hövel E. Schöll, S. H. L. Klapp, ed.), Springer, Berlin, Heidelberg, 2016, pp. 35–72.

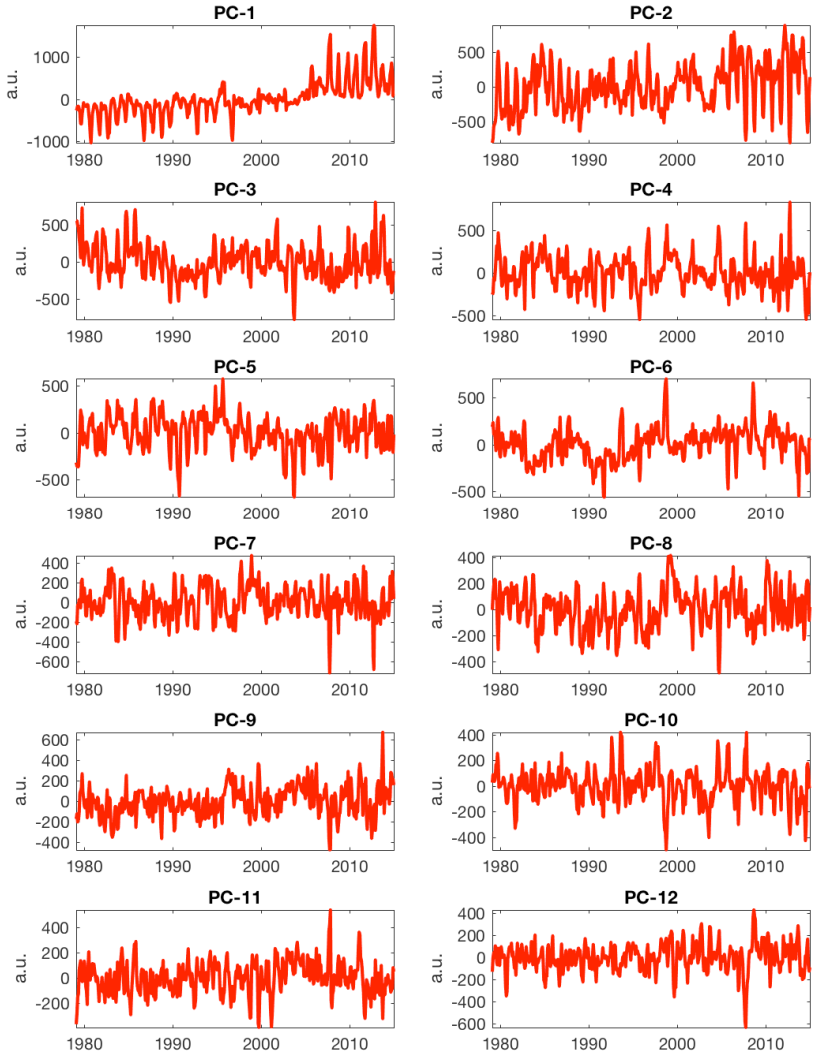


**Fig. 7** Time series of the 12 leading principal components (PCs) of SIC anomalies. The seasonally dependent trend component is very prominent in the 1st and 2nd PC.

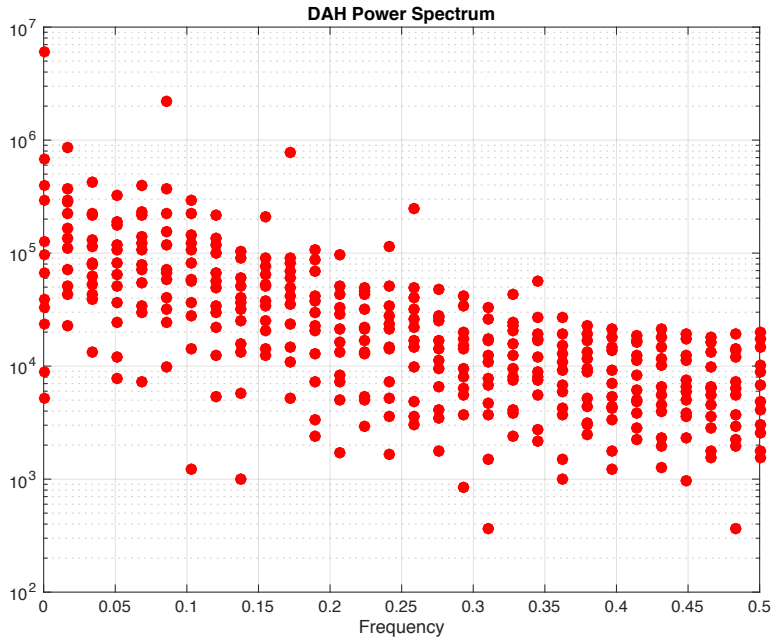


**Fig. 8** DAH spectrum of the 12 leading PCs of the SIC dataset, using an embedding window of $M' = 59$ months.

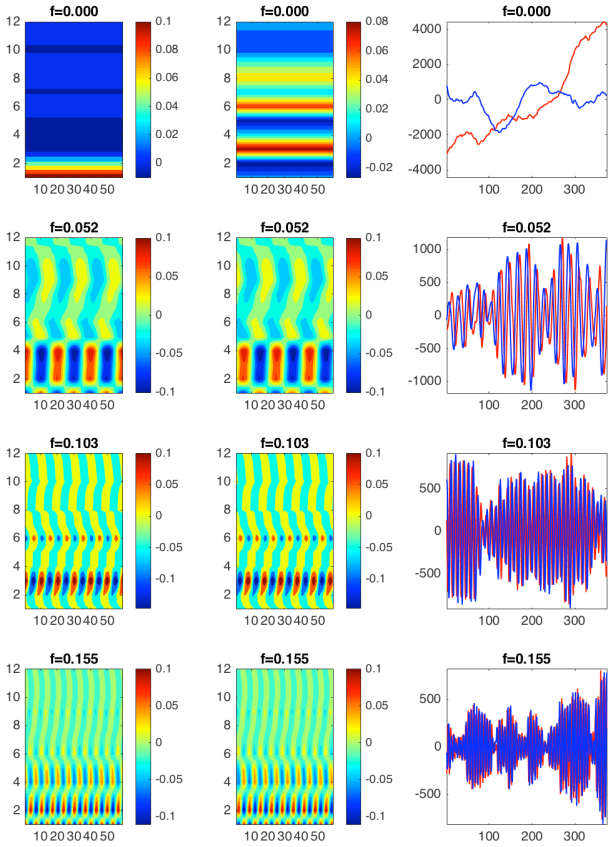


**Fig. 9 Left and center columns:** Spatio-temporal DAH modes (DAHMs) that correspond to the leading DAH pair (1,2) in the SIC dataset's spectrum at selected frequencies: $x$-axis – embedding dimension, $y$-axis – PC index. **Right column:** Corresponding temporal DAH coefficients (DAHCs). The four selected frequencies, $f = 0.0, 0.052, 0.103$ and $f = 0.155$, appear in the caption of each panel.

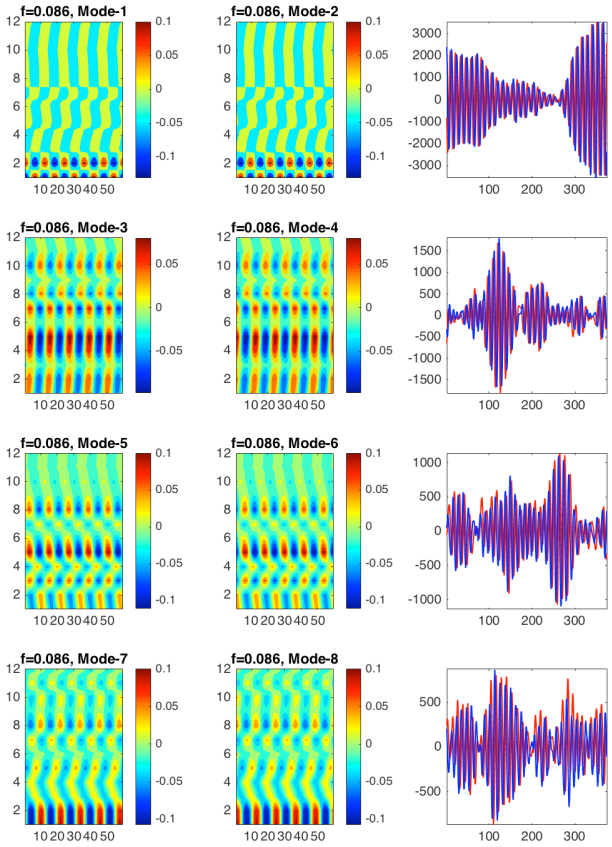


**Fig. 10** Same as Fig. 9, except for showing the four leading pairs at the 12-month periodicity, $f = 0.086$. The DAHMs (1,2), (3,4), (5.5) and (7.8) appear in the caption of each panel.

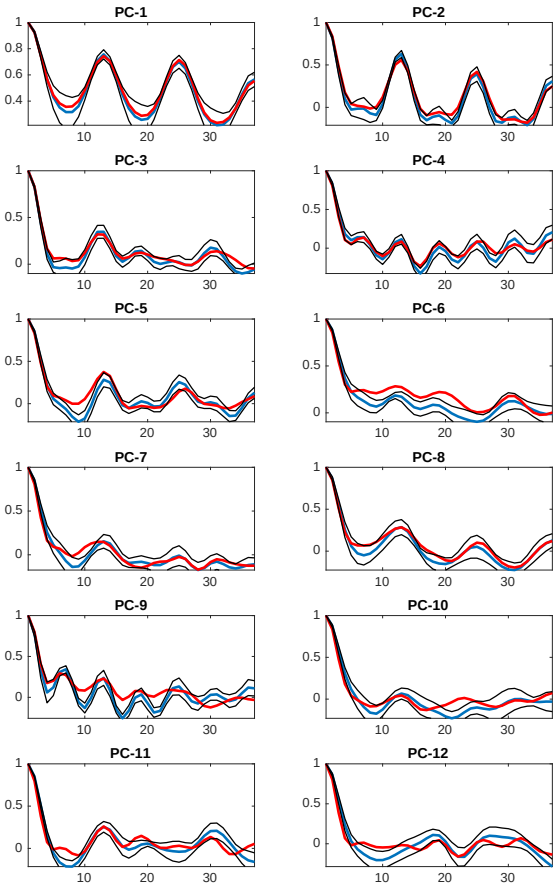


**Fig. 11** The autocorrelation functions (ACFs) of the SIC dataset's PCs: red – observations, black – ensemble mean of stochastic-dynamic simulations by the DAH-MSLM approach; blue – standard deviation of the ensemble.

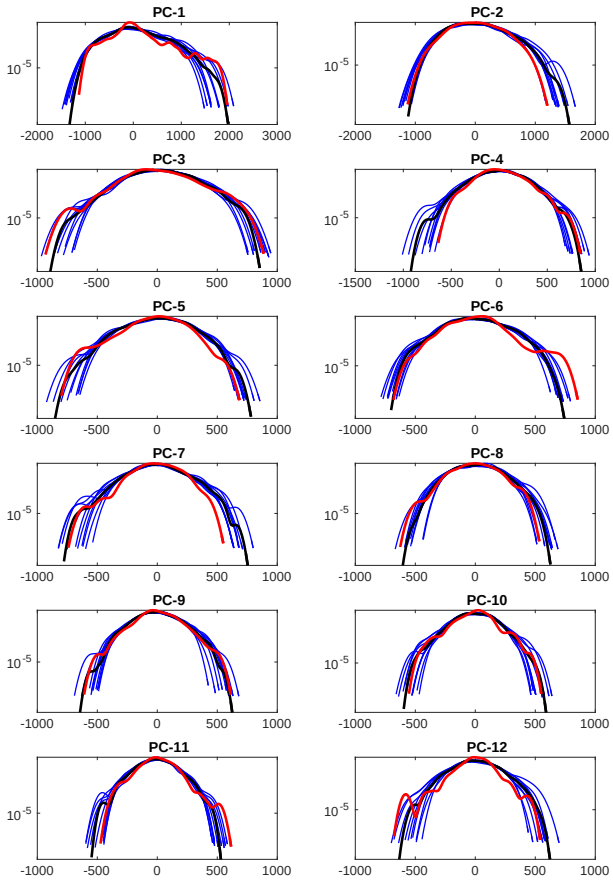


**Fig. 12** Same as Fig. 11, except for the probability density functions (PDFs): the blue lines now represent individual ensemble members.

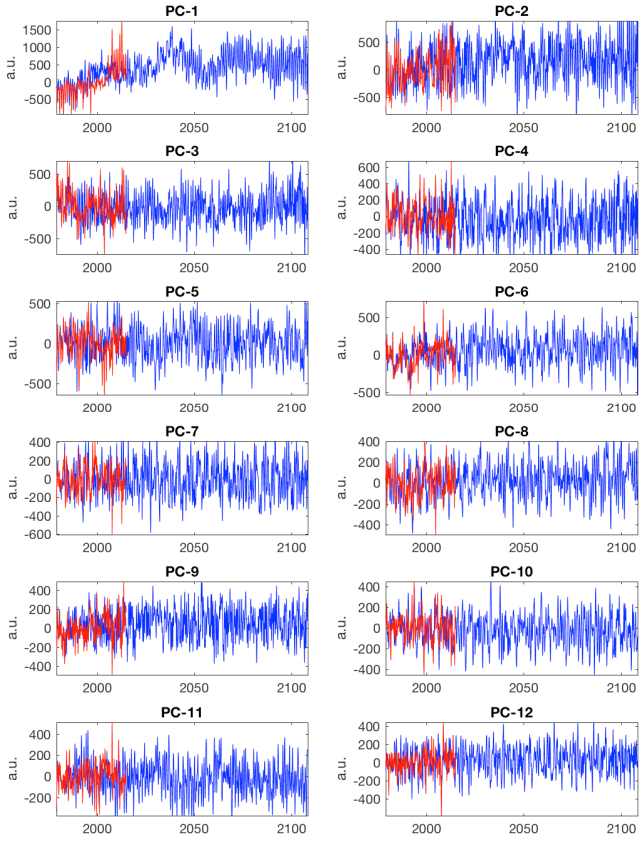


**Fig. 13** Extended simulation of the Arctic SIC conditions. Red – observational dataset of the 12 leading PCs for 1979–2014 (36 years); blue – 129-year–long stochastic simulation by the DAH-MSLM approach.

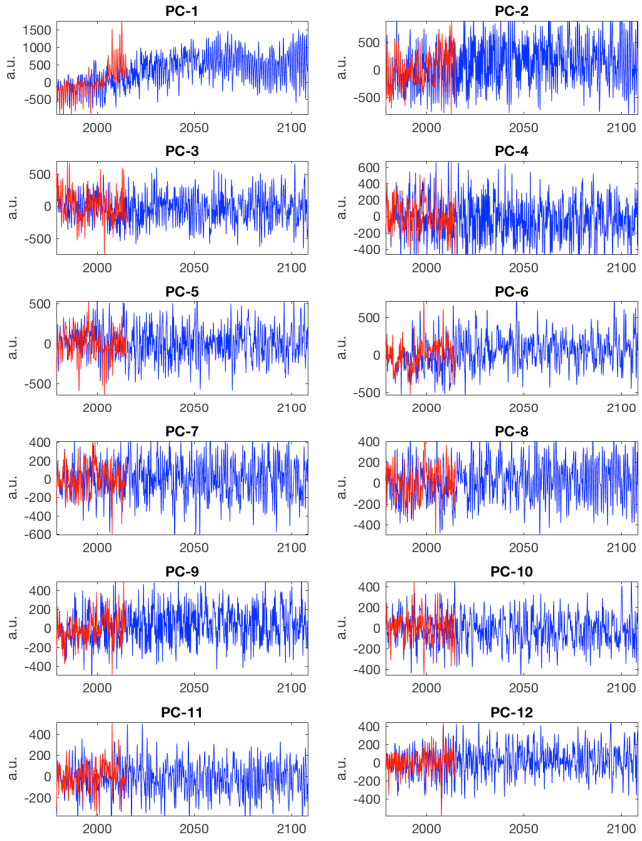


**Fig. 14** Same as in Fig. 13, but for another stochastic realization.

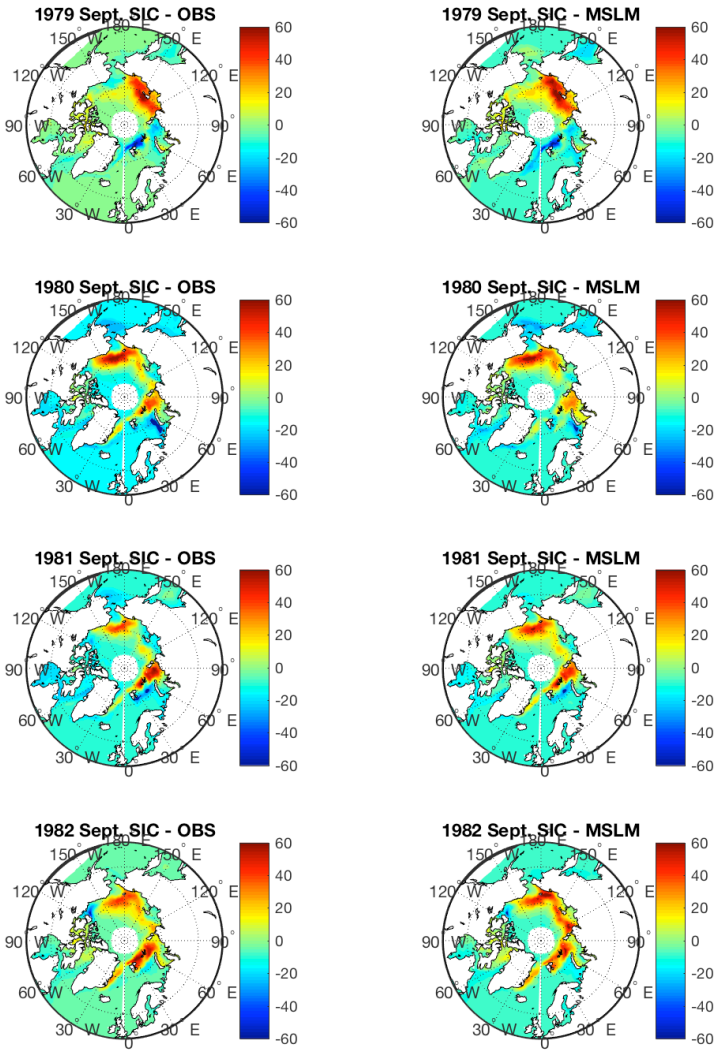


**Fig. 15** Simulations of September SICs by using our DAH-MSLM approach. Left – observed September SIC anomalies; right – hindcast of the DAH-MSLM model, initialized in January 1979. Caption of each panel indicates the particular September being compared, OBS vs. MSLM, for 1979, 1980, 1981, and 1982.