

Article

# Multiscale Stuart-Landau Emulators: Application to Wind-Driven Ocean Gyres

Dmitri Kondrashov <sup>1,\*</sup> , Mickaël D. Chekroun <sup>1</sup>  and Pavel Berloff <sup>2</sup>

<sup>1</sup> Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095, USA; mchekroun@atmos.ucla.edu

<sup>2</sup> Department of Mathematics, Imperial College London, London SW7 2AZ, UK; p.berloff@imperial.ac.uk

\* Correspondence: dkondras@atmos.ucla.edu; Tel.: +1-310-825-1038

Received: 13 February 2018; Accepted: 28 February 2018; Published: 6 March 2018

**Abstract:** The multiscale variability of the ocean circulation due to its nonlinear dynamics remains a big challenge for theoretical understanding and practical ocean modeling. This paper demonstrates how the data-adaptive harmonic (DAH) decomposition and inverse stochastic modeling techniques introduced in (Chekroun and Kondrashov, (2017), *Chaos*, **27**), allow for reproducing with high fidelity the main statistical properties of multiscale variability in a coarse-grained eddy-resolving ocean flow. This fully-data-driven approach relies on extraction of frequency-ranked time-dependent coefficients describing the evolution of spatio-temporal DAH modes (DAHMs) in the oceanic flow data. In turn, the time series of these coefficients are efficiently modeled by a family of low-order stochastic differential equations (SDEs) stacked per frequency, involving a fixed set of predictor functions and a small number of model coefficients. These SDEs take the form of stochastic oscillators, identified as multilayer Stuart–Landau models (MSLMs), and their use is justified by relying on the theory of Ruelle–Pollicott resonances. The good modeling skills shown by the resulting DAH-MSLM emulators demonstrates the feasibility of using a network of stochastic oscillators for the modeling of geophysical turbulence. In a certain sense, the original quasiperiodic Landau view of turbulence, with the amendment of the inclusion of stochasticity, may be well suited to describe turbulence.

**Keywords:** cross-correlations; eddy-resolving; Hankel matrices; inverse modeling; low-frequency variability; Ruelle–Pollicott resonances; stochastic modeling; stochastic oscillators

## 1. Introduction

The turbulent oceanic flows consist of complex motions, jets, vortices and waves that co-exist on very different spatio-temporal scales, but also without clear scale separation. Mesoscale oceanic eddies populate nearly all parts of the global ocean and play important roles in maintaining the oceanic general circulation. The most straightforward, but also the most computationally intensive and, thus, unfeasible, way of accounting for the eddy effects on the large-scale circulation is resolving them dynamically with eddy-resolving ocean general circulation models (GCMs). This brute-force approach requires a computational grid resolution of about 1 km, which makes it feasible only for relatively short time simulations, whereas the Earth system and climate change modeling routinely require much longer simulations over centuries and millennia. One way to afford these time scales, while simulating the ocean in a qualitatively correct way, is to parameterize the important eddy effects with simple and affordable, but still accurate models embedded in the prognostic ocean circulation models.

Thus, the search for suitable eddy parameterizations remains a challenging theoretical topic with clear practical dimension. More often, the embedded prognostic models are dynamical, in the sense that they solve some coarse-grained approximations of the governing primitive equations, but computationally still fairly expensive, because they solve for many degrees of freedom. Here, the

parameterizations of the eddy effects can be deterministic, where eddy diffusion is by far the most common approach (e.g., [1,2]), stochastic (e.g., [3–8]), or a combination of both. The latter approach is attracting increasing attention because it can, not only account for nondiffusive and antidiiffusive eddy effects (e.g., negative eddy viscosity, eddy backscatter, etc.), but also takes into account the inherent “randomness” of complicated turbulent processes.

Taking a broader view, the exact form of such reduced models can be found rigorously from the full model equations only in special cases, by adopting, for instance, the Mori–Zwanzig (MZ) projection approach [9–13] or methods rooted in the approximation theory of local invariant manifolds, deterministic [14–17] or stochastic ([18,19] and the references therein). Alternatively, when a rigorous derivation of the reduced dynamics is mathematically intractable or when the governing equations are not known, a data-driven inverse modeling approach naturally leads to the development of emulators, i.e., purely statistical, severely truncated in terms of degrees of freedom, prognostic models of reduced complexity that can reproduce the whole complexity of turbulent oceanic motions across scales. Such inverse models can be trained on the data from model simulations, or from observations, and they can be viewed as inexpensive statistical oceanic emulators. Their potential applications range from representing the ocean in climate process studies to providing a framework for analyses of multiscale flow interactions.

For successive inverse modeling, one of the key steps is to find relevant data-adaptive basis functions. Numerous techniques are known for decomposition of time-evolving datasets into data-adaptive modes, which include: (i) variance-based decomposition, such as principal component analysis (PCA) [20,21], its probabilistic formulation [22], as well as its nonlinear [23,24] and time-embedded extensions [25]; (ii) eigenfunctions of Markov matrices that reflect the local geometry and density of the data [26–29]; and (iii) approaches that rely on Koopman operator theory [30–35]. While working in a suitable subspace spanned by the leading modes, one needs then to derive evolution equations that characterize the long-term dynamics of the full dynamical system that has produced the dataset.

Many methods addressing this task include nonlinear autoregression, moving averages with exogenous inputs (NARMAX) [36–38], artificial neural networks (ANNs) [39–41], proper orthogonal decomposition (POD) techniques and alike [42–45], stochastic linear inverse models (LIMs) [46,47], multilevel approaches and empirical model reduction (EMR) [48–53] generalized by a multilayer stochastic model (MSM) framework [54]. In a data-driven context, the MZ formalism actually provides theoretical guidance for the determination of the proper nonlinear and memory effects for the optimal reduced model [54]. Nevertheless, whether the approach retained uses an a priori set of predictor functions or allows for dictionaries of such functions, there is often either an explanatory deficit or a selection problem of the appropriate class of predictors.

This study demonstrates how to overcome such hurdles by applying the data-adaptive harmonic decomposition (DAHD) techniques introduced in [55] to emulate the full spectrum of dynamically important scales in the coarse-grained eddy-resolving ocean model solution, including mesoscale eddies. As discussed below, this is made possible by the ability of DAHD to extract spatio-temporal modes whose time-evolution can be efficiently modeled by a universal family of frequency-based low-order stochastic differential equations (SDEs) involving a fixed set of predictor functions, namely coupled stochastic oscillators made of multilayer Stuart–Landau models (MSLMs) ([55], Section VIII).

To make the expository as self-contained as possible, the DAHD technique and the MSLMs are summarized in Sections 4.2 and 4.3, respectively. It is noteworthy that, after prerequisites from the theory of Ruelle–Pollicott resonances [56] recalled in Appendices A and B, the Appendices C and D below communicate, in the context of the oceanic model [8], on the mathematical foundations underlying the modeling by the class of Stuart–Landau models (of type (25) used hereafter), for the time-evolution of the corresponding DAH modes (DAHMs). The resulting DAH-MSLM approach has already shown its flexibility for datasets such as analyzed ([55], Section IX) and a great promise for the stochastic data-driven modeling of challenging datasets in different branches of geophysics

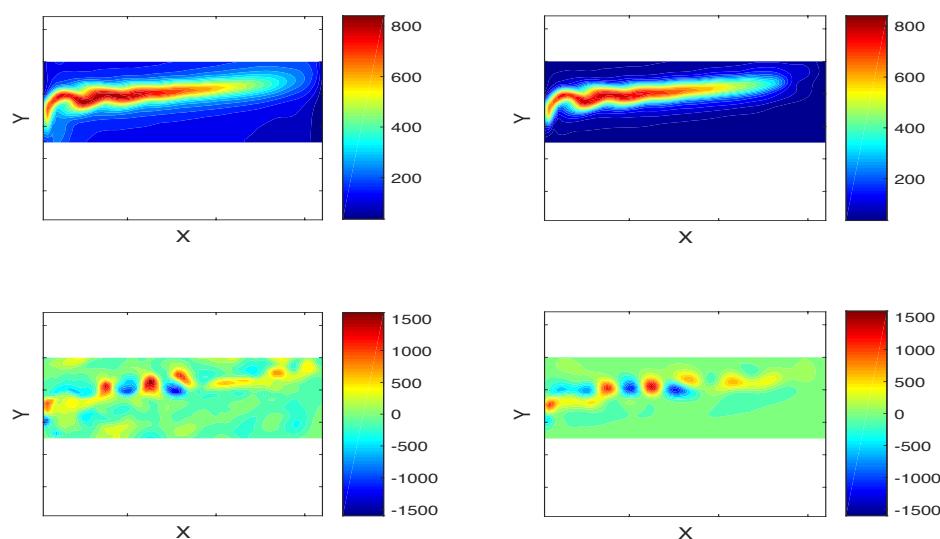
where modeling and prediction is hampered, on the one hand, by the shortness of observational records and, on the other, by the shortcomings of physics-based models: Arctic sea ice extent and concentrations [57,58] and solar wind-magnetosphere interactions [59]. This study demonstrates, in the context of ocean modeling, that the DAH-MSLM approach allows for efficiently simulating the time-evolution of the spatio-temporal DAHMs across the full range of temporal scales covered by the oceanic model of [8], reproducing fairly accurately, in particular, the key statistical properties of the flow's nonlinear dynamics.

## 2. Results

### 2.1. Oceanic Dataset

We consider a model of mid-latitude ocean gyres where flow dynamics is simulated by the three-layer quasi-geostrophic (QG) equations forced by the wind stress in the upper layer, in which merging western boundary currents form powerful eastward jet extension, similar to the Gulf Stream and Kuroshio; see Section 4.1. The ocean circulation of eddy-resolving simulation with  $\approx 10^6$  spatial degrees of freedom (d.o.f.) at reference model parameters [8] is characterized by a robust large-scale decadal low-frequency variability (LFV) at  $\approx 17$  years and involving coherent meridional shifts of the eastward jet extension separating the gyres. This LFV is driven by transient mesoscale eddies, qualitatively similar to the turbulent oscillator studied by [60,61]; see [62–68] for alternative theories. Still, there is no clear scale separation between the eddies and large-scale flow, and in fact, the LFV accounts for a relatively small fraction of the total variability, as demonstrated below in the detailed DAH analysis.

First, we consider the upper-ocean stream function anomalies, given by the dataset of  $6.5 \times 10^5$  days, sampled 10 days apart and coarse-grained in the spatial region centered around the eastward jet (see Figure 1) with 64 points in the  $x$ -direction and 26 points in the  $y$ -direction. Then, the resulting field having  $N = 3.25 \times 10^5$  time snapshots and sampled 20 days apart is compressed by applying a standard PCA [21]. The leading  $d = 30$  empirical orthogonal function (EOF) modes (from a total of 1664 modes), which capture  $\approx 70\%$  of the total variance, are retained for the subsequent analysis. Discarded residual variability accounts for small spatial scales away from the jet; the latter being shown in Figure 1. The corresponding time series of principal components (PCs) are obtained by projecting the upper-ocean stream function anomalies onto the retained EOFs.



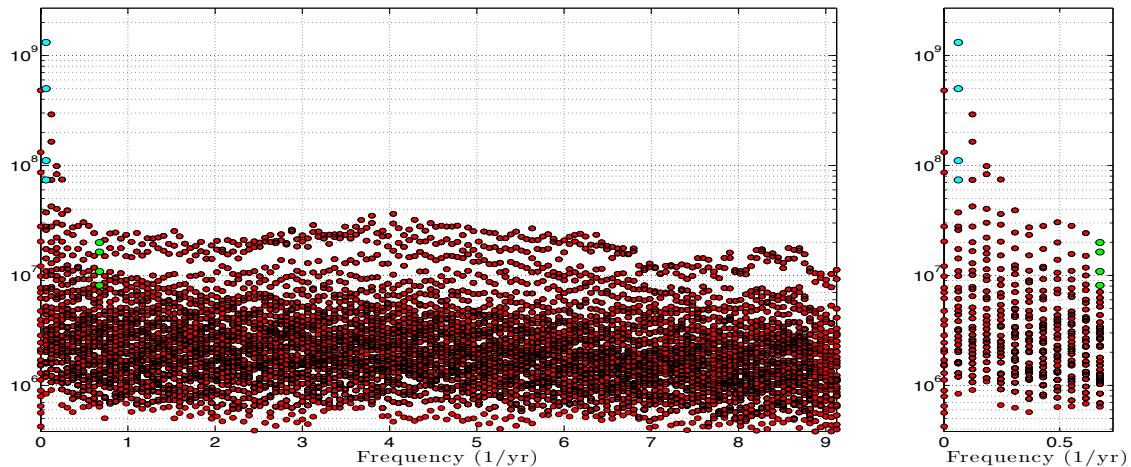
**Figure 1.** Upper-level stream function anomalies. (Left) Top: standard deviation; bottom: snapshot of instantaneous flow. (Right) Same, but in the truncated subspace of 30 leading spatial empirical orthogonal functions (EOFs). The nondimensional units are arbitrary, but are the same for all panels.

## 2.2. DAHD, DAH Power Spectrum and DAHMs

Next, we apply DAHD (Section 4.2) for the spectral analysis of the dataset constituted by the 30 corresponding PCs to diagnose, in particular, the multiscale ocean variability such as simulated by the model studied here; see Section 4.1 below for a description. Section 4.2 below recalls from [55] the main mathematical ingredients needed for the computations of the DAH spectra and modes.

To compute the latter, we first compute all the combinations of two-sided cross-correlations among the PCs time series up to lag  $M = 150$  (in sampling units) and formed the grand matrix  $\mathfrak{C}$  given by (12) below with  $d = 30$  and  $M' = 2M - 1 = 299$  ( $\approx 17$  y), allowing us thus to resolve the decadal variability. Given the temporal sampling of 20 days used here, the Nyquist interval of resolved frequencies is therefore  $f < 9.1$   $y^{-1}$ . After computing the DAH eigenvalues ( $\lambda_j$ ) and associated eigenvectors ( $\mathbf{W}_j$ ) of the grand matrix  $\mathfrak{C}$ , we re-arrange them according to the procedure described Section 4.2.1 in order to form the DAH power spectrum  $\mathcal{P}_\ell$  given in (15), over each Fourier frequency  $f_\ell$ ; see (16) below.

Figure 2 shows the resulting DAH power spectrum with the corresponding  $|\lambda_j|$  plotted as red filled circles. The DAH power spectrum is evenly spaced in frequency and the number of frequency bins  $B_f = (M' + 1)/2 = 150$ . There are exactly  $d = 30$  pairs of  $|\lambda_j|$  at each equidistant frequency  $f$ , except at  $f = 0$ , where there are  $d$  unpaired values [55]. Note that in Figure 2 and the figures hereafter, the frequency values  $f$  that appear therein correspond to  $f_\ell/(2\pi)$ . The DAH power spectrum exhibits a complicated “bumpy” plateau-like shape stretching across most of the frequencies and without a clear distinction between fast and slow timescales; a bumpy plateau that coexists with a pronounced sharp peak at low frequencies corresponding to decadal periodicity  $\approx 17$  y ( $0.061$   $y^{-1}$ ) as shown in Figure 2 (cyan dots in Figure 2). A closer look at the details of the DAH power spectrum indicates that the bumpy part of the plateau is actually made of much broader, but less pronounced peaks spanning over an intermediate range of frequencies,  $3$   $y^{-1} < f < 6$   $y^{-1}$ , as well as over a range of very high frequencies near  $f \approx 8.5$   $y^{-1}$ , thus close to the Nyquist frequency.

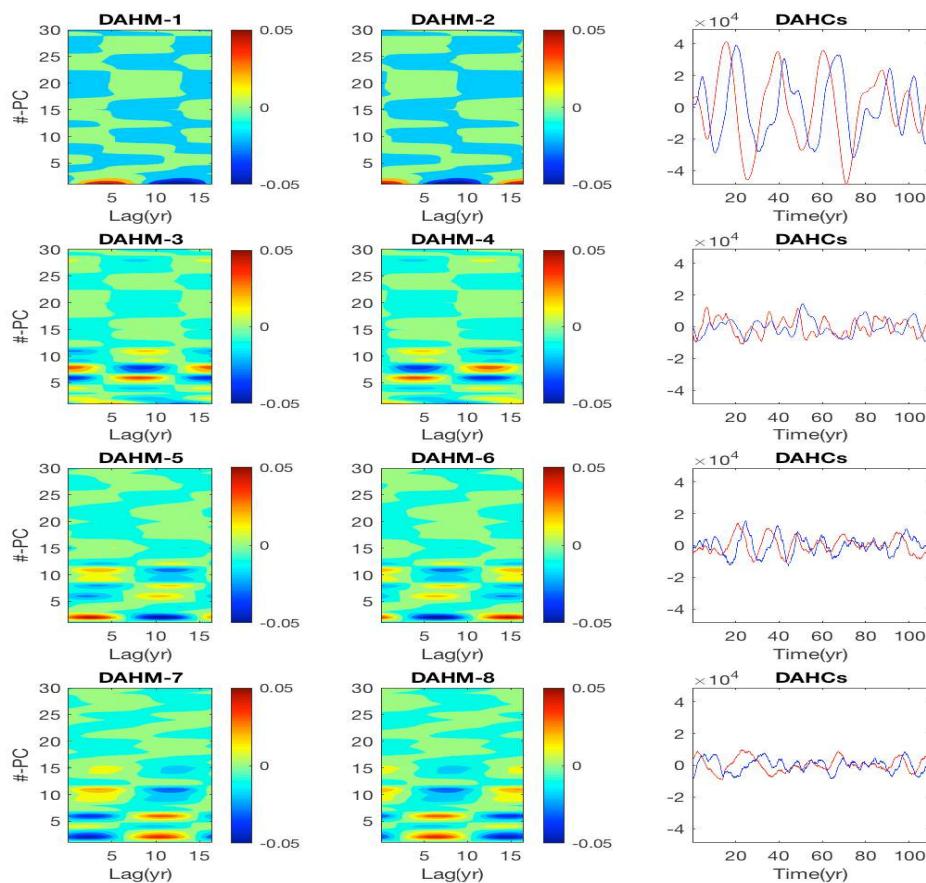


**Figure 2.** DAH power spectrum  $\mathcal{P}_\ell$  of 30 leading PCs of the upper layer stream function anomalies. Left panel: Each discrete set  $\mathcal{P}_\ell$  consists of 30 eigenelements (equal to number of input dataset’s channels), corresponding to the pairs of DAH eigenvalues  $|\lambda_j|$  and eigenvectors  $\mathbf{W}_j$  at a given temporal frequency  $f_\ell$  (see Section 4.2.1 and (15)–(17)). Figures 3 and 4 show DAHMs associated with four largest  $|\lambda_j|$  at two selected frequencies: cyan, decadal LFV peak  $f_D = 0.061$   $y^{-1}$  ( $\approx 17$  y); green, interannual  $f_I = 0.674$   $y^{-1}$  ( $\approx 1.5$  y). The right panel shows a magnification of the low-frequency part of the spectrum.

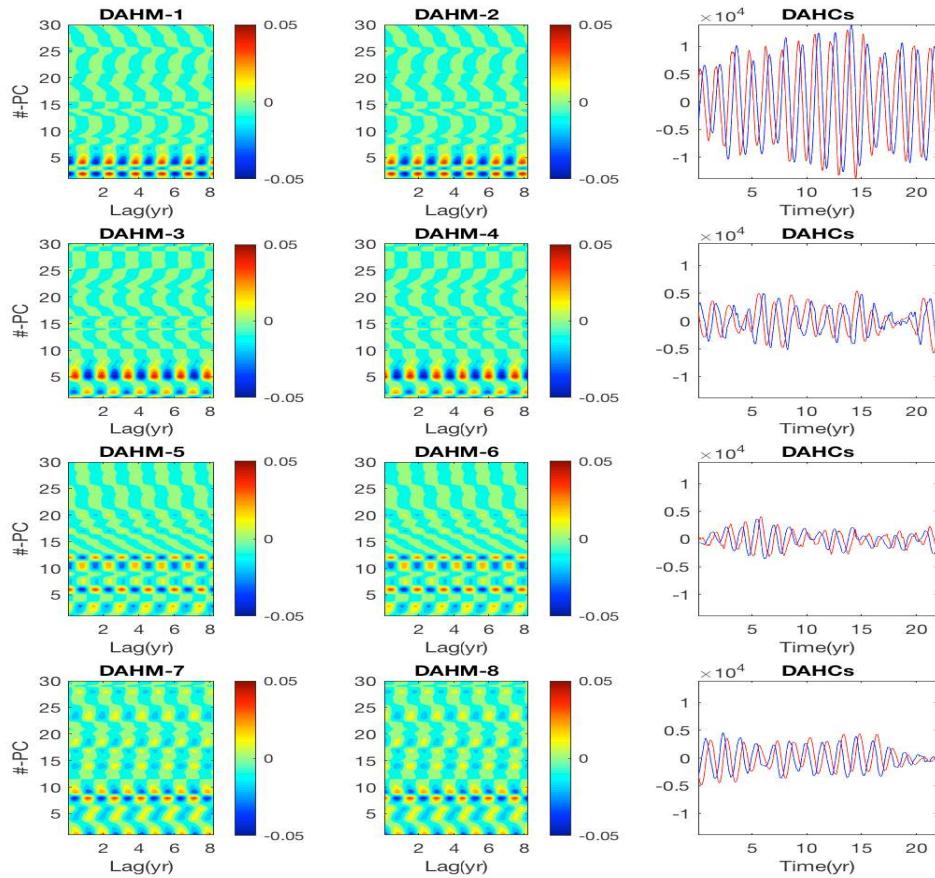
Recall from [55] that a separation in the DAH power spectrum between the DAH pairs at a given frequency  $f$  reflects how the singular values of the cross-spectral matrix  $\mathfrak{S}(f)$  defined by (13) are distributed. For the dataset analyzed here, while there is a large gap between dominant power

values at decadal peak, other frequencies tend to have several pairs in one group that is separated above the continuous background.

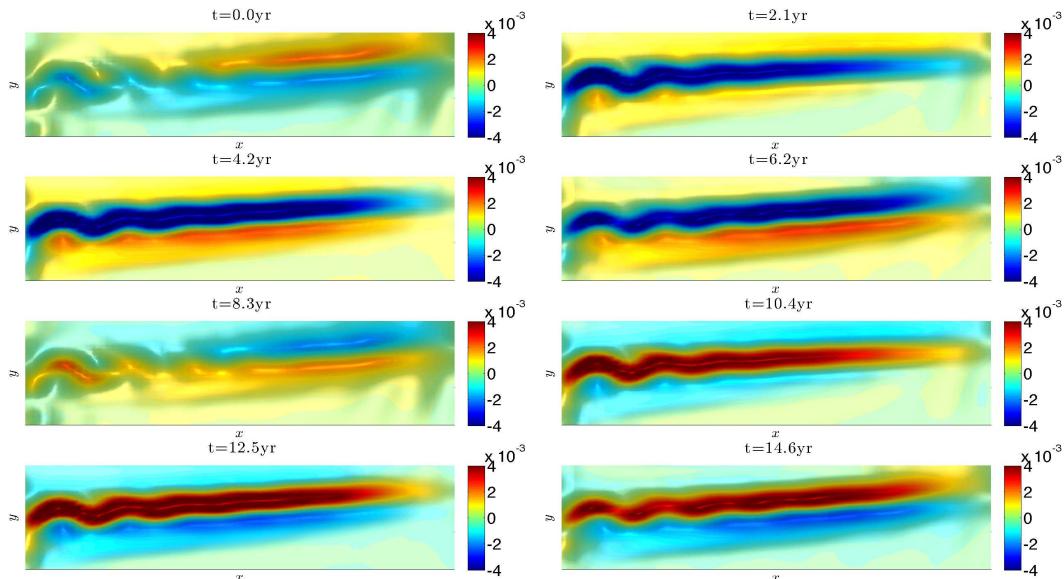
Figures 3 and 4 show space-time patterns (with the “space-coordinate” taken in the EOF-phase space) for the four pairs of DAHMs ( $\mathbf{W}_j$ ; see (17) and (18)), corresponding to the largest DAH spectral power at decadal and interannual frequency, respectively (cyan and green dots in Figure 2). The dominant modes, i.e., those corresponding to the pair having largest  $|\lambda_j|$  at these frequencies, have largest magnitude in the channels corresponding to leading PCs. As  $|\lambda_j|$  decreases, channels of higher-ranked PCs prevail with complex and non-trivial patterns. Figures 5 and 6 show the oscillation phases of the dominant (associated with the largest  $|\lambda_j|$ ) DAHM patterns, when transformed back to the physical space by using EOFs. Since the DAHMs shown on the left and center columns of Figures 5 and 6 have a time-component corresponding to the embedding dimension ( $\approx 17$  y) [55], the transformation consists here of multiplying the 30 “spatial” channels of the DAHMs by the EOFs and then plotting in the physical space the resulting patterns as time evolves, within the embedding window. By doing so, the decadal mode reveals mostly coherent meridional shifts of the eastward jet extension, while the intraseasonal one represents largely standing oscillation of small eddies along the jet extension.



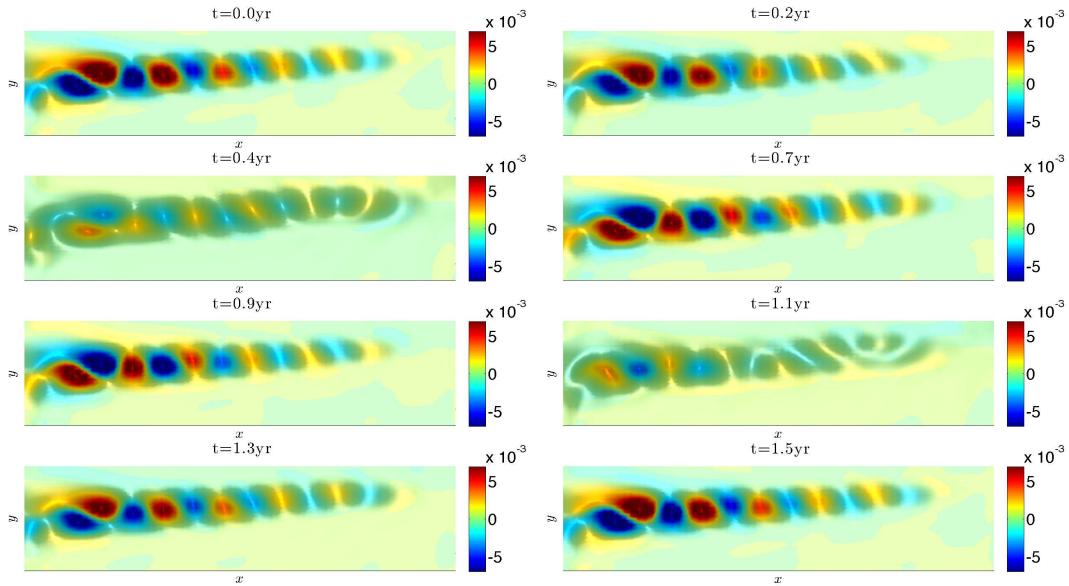
**Figure 3.** Left and center panels: Space-time patterns of data-adaptive harmonic mode (DAHM) pairs corresponding to the four largest  $|\lambda_j|$  (in descending order) at decadal frequency  $f = 0.061 \text{ y}^{-1}$  (cyan dots in the data-adaptive harmonic (DAH) power spectrum of Figure 2). Each of the modes in a pair is time-shifted by a quarter of a period, i.e., in exact phase quadrature;  $x$ -axis, time embedding dimension (in years);  $y$ -axis, spacial dimension (rank of principal component (PC)). Right panels: DAHCs obtained by projection of the input dataset of 30 PCs onto the DAHMs; see (20). A DAHC pair consists of narrowband time series at the same temporal frequency of the associated DAHMs, but modulated in amplitude.



**Figure 4.** Same as in Figure 3, but for  $f = 0.674 \text{ yr}^{-1}$ ; associated here with the green dots in the DAH power spectrum shown in Figure 2.



**Figure 5.** Manifestation in the physical domain of the leading pair of DAHMs (see Figure 3) at the decadal variability  $f = 0.061 \text{ yr}^{-1}$ ; i.e., corresponding to the largest  $|\lambda_j|$  (top cyan dot) in Figure 2. The resulting pattern is periodic (with period  $\approx 16.39 \text{ yr}$ ). Here, eight oscillation phases labeled by time are shown.



**Figure 6.** Same as in Figure 5, but for the pair of DAHMs associated with the largest  $|\lambda_j|$  (top green dot in Figure 2) at the interannual frequency  $f_I = 0.674 \text{ yr}^{-1}$  (with period  $\approx 1.5 \text{ yr}$ ) (cf. Figure 4).

### 2.3. DAH-MSLM Oceanic Emulators

According to (20), the DAHCs are obtained by projection of the input dataset of retained PCs onto DAHMs and thus have  $N' = 32,500 - 299 + 1 = 32,202$  data points (21). Right columns of Figures 3 and 4 show that the time series of DAHCs are narrowband at the characteristic frequency associated with the respective DAHM pairs, as predicted by the theory (Section 4.2.2). Interannual DAHCs (Figure 4) appear to be very well in phase-quadrature (i.e., having a shift of a quarter of their period), since window  $M'$  is sufficiently long to resolve this periodicity. On the other hand, the phase-quadrature relationship is less precise for decadal DAHCs (Figure 3) since  $M'$  is comparable to such periodicity. Furthermore, the magnitudes of DAHCs are higher for the dominant spectral pair and tend to diminish as  $|\lambda_j|$  decreases.

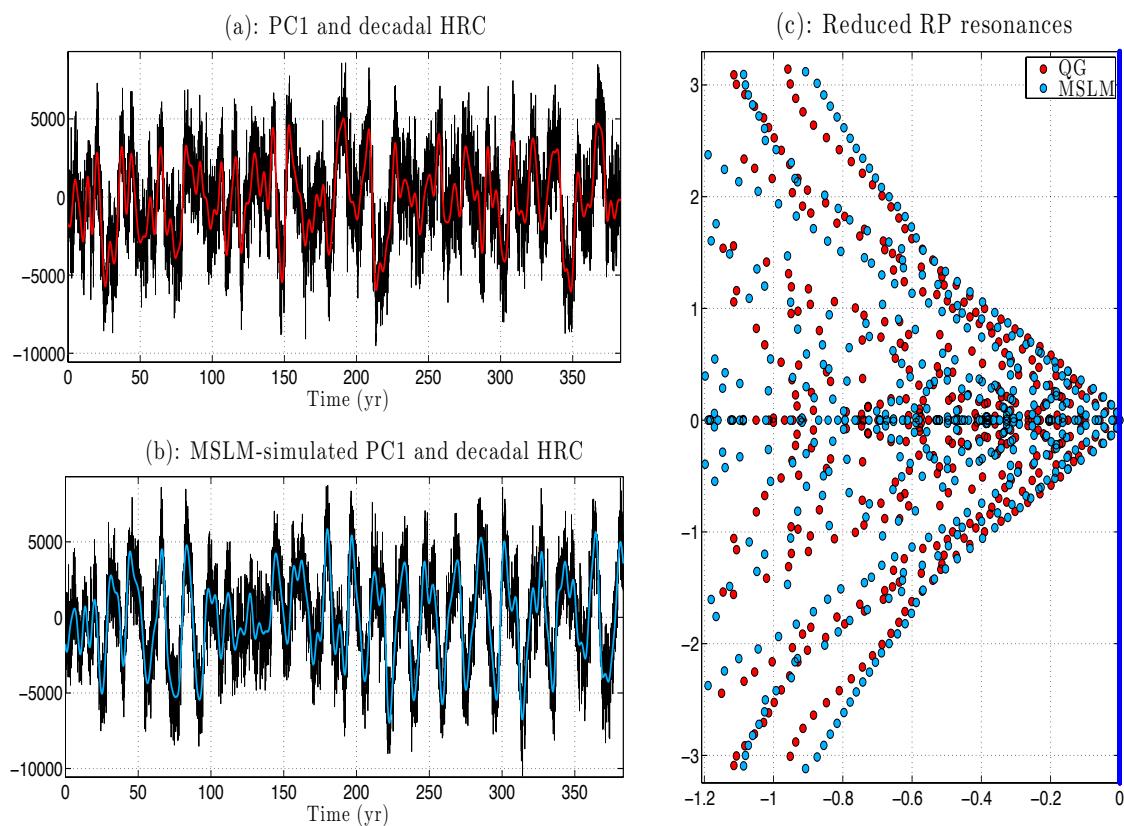
To model the time-evolution of the DAHCs, we apply now the MSLM modeling approach of ([55], Section VIII) and summarized in Section 4.3 below, for the reader's convenience. For each pair of DAHCs, the resulting MSLM model was trained to estimate  $3 + 4(d - 1) = 119$  coefficients for the main layer of Equations (26). The model is then integrated from initial conditions corresponding to the start of the record and is forced by a white noise realization to obtain an ensemble of simulated DAHCs. The latter are convolved with DAHMs according to (22) to obtain harmonic reconstruction components (HRCs) (23) that are added in the frequency bands of interest.

The DAH-MSLM model was estimated and run in parallel at  $B_f = 150$  frequencies in the full range  $f \leq 9.1 \text{ yr}^{-1}$  (see Section 4.3), taking  $\approx 30$  min of CPU time on a four-core 2.9 GHz Intel Core i7 MacBook Pro laptop to obtain a simulation of 1780 years long. On the other hand, the reference solution of the QG model (Section 4.1) takes about 1000 single-CPU hours to obtain the full data record [69].

Figures A7 and A8 of Appendix E below show good modeling skills of the resulting DAH-MSLM emulator in reproducing key statistical properties of leading PCs, namely their probability density functions (PDFs) and their autocorrelation function (ACFs), in a partial frequency band  $f \leq 5.27 \text{ yr}^{-1}$ . Furthermore, good statistical modeling skill is also obtained for the full range of temporal frequencies (i.e., in the Nyquist interval  $f \leq 9.1 \text{ yr}^{-1}$ ), and respective results are shown in Figures A5 and A6. Due in part to the long length of the simulation used here, these statistical skills do not show a marked dependence on the particular realization of the white noise used to drive our DAH-MSLM emulators. This observation is actually consistent with the ergodic property that an MSLM of the form (26) must

satisfy; a property that will be communicated elsewhere. Note that for an optimal simulation of the high-frequency band  $5.3 \text{ yr}^{-1} \leq f \leq 9.1 \text{ yr}^{-1}$ , it was beneficial to have  $b_{ij}^x$  and  $b_{ij}^y$  set to zero in (26).

While the PDFs are mostly Gaussian, the autocorrelation functions exhibit a complicated mixture of temporal scales and thus do not allow for a clear ranking between the different PCs in terms of variability. For example, a slower decay is dominant in ACFs of PC1, 6, 8 and 11, while other PCs exhibit fast decay with superimposed oscillations. This is where a key advantage of the DAH-MSLM approach lies, i.e., in its ability to distinguish distinct temporal scales from a mixture of frequencies embedded in the data allowing in turn for an effective modeling; the distinction being under the monitoring of the DAHD, while the modeling itself is made efficient by the use of MSLMs. PC1 is shown as the black curve in Panel (a) of Figure 7. There, one can observe already in this raw time series the mixture of frequencies superimposed on a modulated oscillation of decadal dominant period. PC1 as obtained from the DAH-MSLM emulator is also shown as the black curve in Panel (b) of Figure 7.



**Figure 7.** Decadal harmonic reconstruction component (HRC) and its reduced RP resonances for PC1: quasi-geostrophic (QG) model and multilayer Stuart-Landau model (MSLM). (a,b) show the sum of the first four HRCs on PC1 for the upper-layer stream function anomalies as simulated from the QG model (red curve) and its DAH-MSLM emulator (blue curve). In both panels, the PC1 is shown in black: in (b), PC1 is obtained after simulation of the DAH-MSLM emulator, whereas in (a), PC1 is obtained from simulation of the QG model. The HRCs are computed according to (23) in Section 4.2.2 below, from the corresponding simulated data. (c) shows the corresponding reduced RP resonances (see Appendix B) as estimated from HRCs shown in (a,b); colors match across panels (a-c).

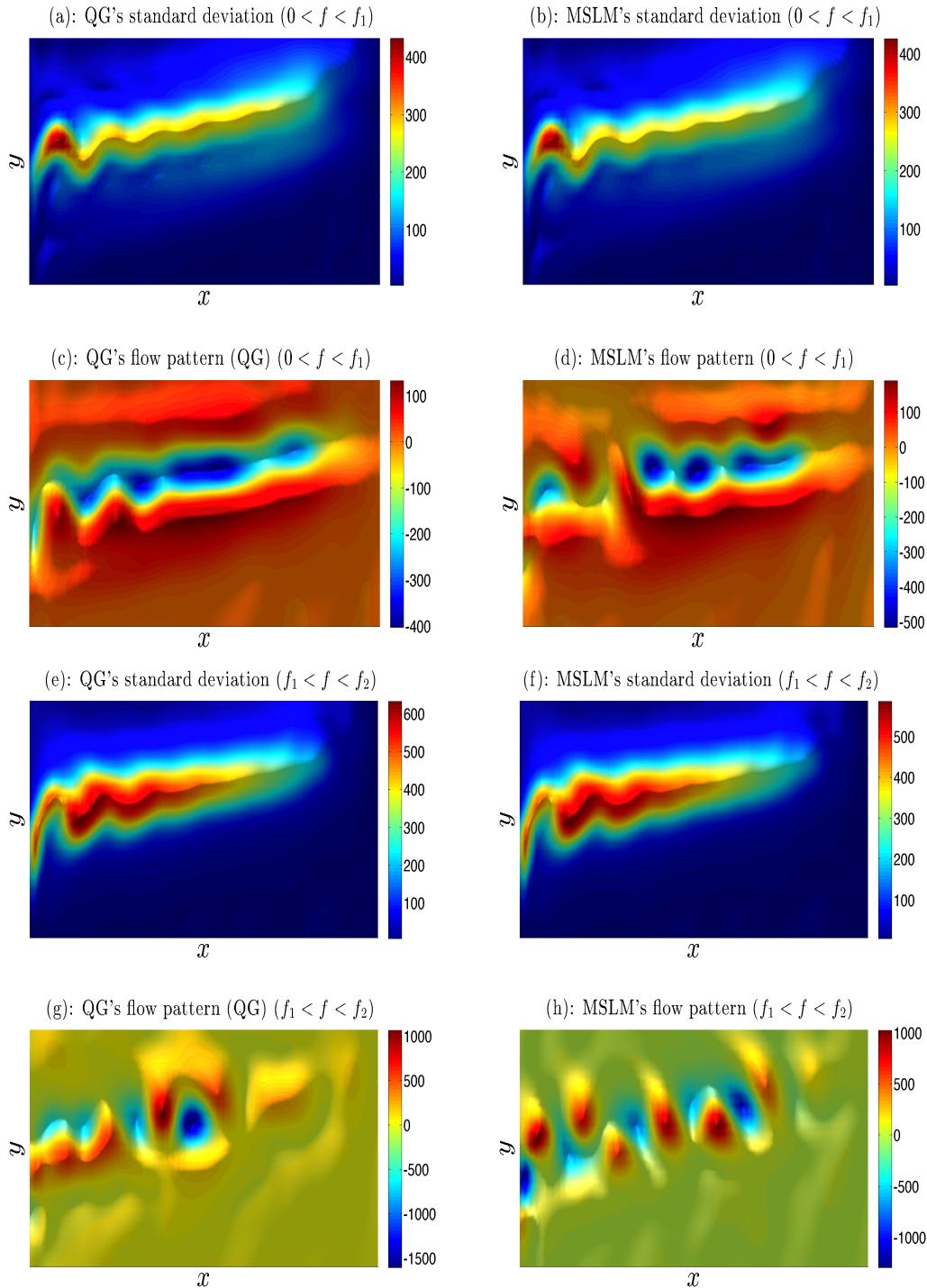
A comparison between these two curves with the naked eye shows already that the main dominant period, as well as the time series modulations and the higher frequency part of the signal are well reproduced. To better assess how the LFV associated with the decadal oscillation is captured, we proceeded in two steps. First, for each PC1, as simulated from the QG model and from the DAH-MSLM emulator, we calculated the first four HRCs that we sum up to span a range of frequencies corresponding to the decadal variability of the flow as identified by the DAH power spectrum. The latter sum of HRCs is shown as the red and light blue curves in Panels (a) and (b) for PC1 as simulated by the QG model and the DAH-MSLM emulator, respectively. In both cases, these curves provide a low-pass filter, smoothed version, of the signal (here PC1) corresponding to the decadal variability of the flow.

In a second step, we estimated the reduced Ruelle–Pollicott resonances (Appendices A and B) associated with the reduced phase space  $V$  constituted by these sums of HRCs and their lagged versions; the lag being taken here to be equal to  $1\text{ y}$ . Essentially, these resonances are obtained from the spectrum of the transition matrix associated with these sums of HRCs in the reduced phase space  $V$ ; see Appendix B for more details. As outlined in Appendices A and B (see also Appendix C), these resonances are characteristic of the underlying dynamics, and the arrangement of these resonances in the complex plane constitutes a signature of key dynamical features. In particular, they allow for precise decomposition formulas of correlation functions and power spectral densities (PSDs) in term of these resonances; see (A5) and (A6) in Appendix A below. See also [56,70,71] and the references therein. In this sense, their estimation goes beyond the estimation of ACFs and PSDs: the latter statistical quantities could indeed share similar features without sharing necessarily the same underlying resonances. In terms of modeling, an agreement of these resonances is thus more demanding, but also more indicative of a good capture of the dynamics than by looking simply at the reproduction of ACFs and PSDs.

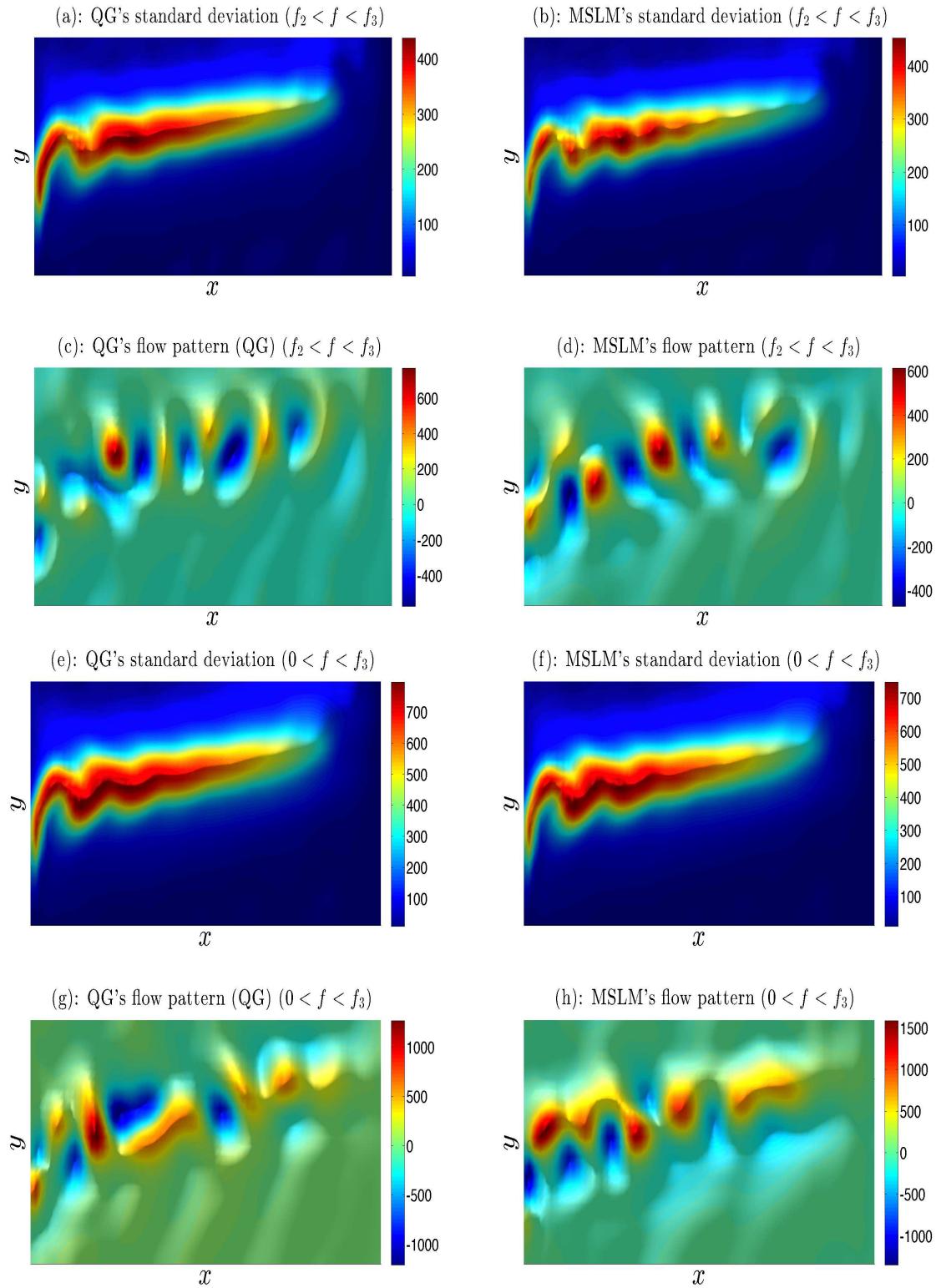
Panel (c) of Figure 7 shows a good agreement of the arrangements of the resulting reduced RP resonances estimated from the HRCs calculated from the QG simulations, on the one hand, and from the DAH-MSLM emulator, on the other. Such good modeling skills as diagnosed by (reduced) RP resonances are not only the attribute of the decadal variability, they are actually observed across other frequency ranges such as corresponding to interannual variability and even subannual and for most of the PCs (not shown). These results obtained in the EOF-phase space or its reduction indicate that a good agreement is expected to take place in the physical space, as well, between the simulations of the DAH-MSLM emulator and the original QG model from which it is derived.

To assess the modeling skills in the physical space, the DAH-MSLM simulations (in selected frequency bands) are transformed into the physical space by using EOFs. Comparison with harmonic reconstructions of QG data in low, intermediate, high and full frequency bands shows realistic instantaneous flow patterns obtained by the DAH-MSLM emulator (compare Panel (c) (resp. (g)) with Panel (d) (resp. (h)) of Figures 8 and 9), albeit this similarity is necessarily qualitative due to the stochastic nature of the latter.

In addition, the DAH-MSLM emulator reproduces very well both the spatial patterns and magnitude of variance in these key frequency bands; compare Panel (a) (resp. (e)) with Panel (b) (resp. (f)) of Figures 8 and 9. Despite the pronounced decadal peak, the LFV range  $f < 0.18\text{ y}^{-1}$  accounts only for 17% of the total variance in the full range of resolved frequencies  $f < 9.1\text{ y}^{-1}$  (in the truncated subspace of 30 leading EOFs), while most of the variance is captured by the intermediate range of temporal scales.



**Figure 8.** Upper-level stream function anomalies and standard deviation: QG vs. MSLM. The QG's standard deviation (SD) and its MSLM emulation are shown here in Panels (a,b), respectively, for the low-frequency range  $0 < f < f_1 = 0.18 \text{ } y^{-1}$ . Panels (c,d) depict typical flow patterns underlying the SD patterns for the QG model and MSLM, respectively. Panels (e,f) (resp. (g,h)): same as for Panels (a,b) (resp. (c,d)), but for the intermediate frequency range  $f_1 = 0.18 \text{ } y^{-1} < f < f_2 = 5.27 \text{ } y^{-1}$ .



**Figure 9.** Upper-level stream function anomalies and standard deviation: QG vs. MSLM. The QG's standard deviation (SD) and its MSLM emulation are shown here in Panels (a,b), respectively, for the high-frequency range  $f_2 = 5.27 \text{ } y^{-1} < f < f_3 = 9.1 \text{ } y^{-1}$ . Panels (c,d) depict typical flow patterns underlying the SD patterns for the QG model and MSLM, respectively. Panels (e,f) (resp. (g,h)): same as for Panels (a,b) (resp. (c,d)), but for the full range of frequencies  $0 < f < f_3 = 9.1 \text{ } y^{-1}$ .

### 3. Discussion

Thus, the DAHD of the QG model's stream function anomalies, after projection onto the space of EOFs, allows for the efficient modeling of the main statistical features of the flow and its variability across a broad range of frequencies, by a network of stochastic oscillators, namely the MLSMs. In mathematical terms, if one denotes by  $\psi_1$  the anomalies of the upper-layer stream function, the DAHD provides the following representation:

$$\psi_1(t, \mathbf{x}) = \sum_{k=1}^d \left( \sum_f R_k(t; f) \right) \Phi_k(\mathbf{x}), \quad \mathbf{x} = (x, y) \quad (1)$$

where  $\Phi_k$  denotes the  $k$ -th EOF and  $R_k(t; f)$  the harmonic reconstructed component, at the frequency  $f$  for this EOF; see (23) in Section 4.2.2 below. Since the  $R_k(t; f)$  are themselves linear combinations of the DAHC pairs at the same frequency (see (22)), the modeling of  $\psi_1$  boils down to the modeling of these pairs, due to (1). As shown above and further discussed from a theoretical viewpoint in Appendices B–D below, the efficient modeling of these DAHC pairs by MLSMs offers thus an interesting representation/modeling of a turbulent stream function in terms of stochastic oscillators. In a certain sense, the original quasiperiodic Landau view of turbulence [72,73], with the amendment of the inclusion of stochasticity, may be well suited to describe turbulence.

Unlike the data-driven emulator of [69], where the time-lagged multivariate singular spectrum analysis (M-SSA) basis and linear stochastic MSM formulation [54] are used to extract and simulate decadal LFV only, the DAH-MSLM approach allows us to capture the full spectrum of temporal scales, from decadal to mesoscale eddies. The key and unique features that make it possible and that are not available in other data-adaptive decompositions including M-SSA are the rigorous extraction of the spatio-temporal modes (i.e., DAHMs) unambiguously associated with a single temporal frequency (i.e., without mixing of scales) and a well-defined dynamical mechanism to describe their time evolution (i.e., respective DAHCs) using coupled Stuart–Landau stochastic oscillators. Both DAH power and phase spectra hold great promise for the dynamical diagnostics and intercomparison of multiscale datasets and models, focusing on specific frequency bands.

The future research directions include utilizing DAH-MSLM emulators for dynamical analyses, material transport and eddy parameterizations. For example, [5,8,74] approximated actual eddy effects in non-eddy-resolving models by applying transient forcings with spatio-temporal correlations. Such forcings excite flow fluctuations that evolve dynamically and eventually become rectified by the nonlinearity into the eddy-driven large-scale flow anomalies. A major interim weakness of this approach is that it treats transient forcing in the simplest way.

The DAH-MSLM modeling offers great advantages, because not only does it provide vehicles for statistically accurate approximations of the eddy field and its eddy forcing, but also it makes the above parameterization approach much more general and versatile. Instead of applying some random stochastic noise as a replacement for the eddy effects, DAH-MSLM can provide highly constrained forcing functions with spatio-temporal correlations fitted to the actual observed statistics of the eddies. The emulated eddy forcings can be embedded in the corresponding non-eddy-resolving ocean models as data-driven replacements for the eddy effects with direct [5–7] and indirect [8,75] implementations of the embedding.

## 4. Models and Methods

### 4.1. Mid-Latitude Ocean Model

In this study, the data are generated from the integration of the classical eddy-resolving double-gyre ocean model of [8]. The flow dynamics is governed by the quasi-geostrophic potential-vorticity (QG PV) equations for 3 stacked isopycnal layers:

$$\frac{\partial q_1}{\partial t} + J(\psi_1, q_1) + \beta \frac{\partial \psi_1}{\partial x} = \frac{1}{\rho_1 H_1} W(x, y) + \nu \nabla^4 \psi_1 \quad (2)$$

$$\frac{\partial q_2}{\partial t} + J(\psi_2, q_2) + \beta \frac{\partial \psi_2}{\partial x} = \nu \nabla^4 \psi_2 \quad (3)$$

$$\frac{\partial q_3}{\partial t} + J(\psi_3, q_3) + \beta \frac{\partial \psi_3}{\partial x} = -\gamma \nabla^2 \psi_3 + \nu \nabla^4 \psi_3 \quad (4)$$

where the layer index starts from the top;  $J(\cdot, \cdot)$  is the Jacobian operator;  $\rho_1 = 10^3 \text{ kg}\cdot\text{m}^{-3}$  is the upper layer density;  $\beta = 2 \times 10^{-11} \text{ m}^{-1}\cdot\text{s}^{-1}$  is the planetary vorticity gradient;  $\nu = 20 \text{ m}^2\cdot\text{s}^{-1}$  is the eddy viscosity coefficient; and  $\gamma = 4 \times 10^{-8} \text{ s}^{-1}$  is the bottom friction parameter. The basin size is  $2L = 3840 \text{ km}$ , so that  $-L \leq x \leq L$  and  $-L \leq y \leq L$ . The isopycnal layer depths are  $H_1 = 250$ ,  $H_2 = 750$  and  $H_3 = 3000 \text{ m}$ . The PV anomalies  $q_i$  and the velocity stream functions  $\psi_i$  are related as:

$$q_1 = \nabla^2 \psi_1 + S_1 (\psi_2 - \psi_1) \quad (5)$$

$$q_2 = \nabla^2 \psi_2 + S_{21} (\psi_1 - \psi_2) + S_{22} (\psi_3 - \psi_2) \quad (6)$$

$$q_3 = \nabla^2 \psi_3 + S_3 (\psi_2 - \psi_3) \quad (7)$$

where the stratification parameters  $S_1, S_{21}, S_{22}, S_3$  are such that the first and second Rossby deformation radii are  $Rd_1 = 40 \text{ km}$  and  $Rd_2 = 20.6 \text{ km}$ , respectively. The flow is forced by the prescribed wind stress curl  $W(x, y)$ . The double-gyre Ekman pumping  $W(x, y)$  is asymmetric in order to avoid artificial symmetrization of the gyres:

$$W(x, y) = -\frac{\pi \tau_0 A}{L} \sin \left[ \frac{\pi (L + y)}{L + Bx} \right], \quad y \leq Bx \quad (8)$$

$$W(x, y) = +\frac{\pi \tau_0}{LA} \sin \left[ \frac{\pi (y - Bx)}{L - Bx} \right], \quad y > Bx \quad (9)$$

where the asymmetry parameter is  $A = 0.9$ , the non-zonal tilt parameter is  $B = 0.2$ , and the wind stress amplitude is  $\tau_0 = 0.8 \text{ N}\cdot\text{m}^{-2}$ .

There are no-flow-through and partial-slip boundary conditions on the lateral walls, augmented by the integral mass conservation constraints. The model is solved on the uniform  $513^2$  grid with 7.5-km nominal resolution. The solution is saved every 20 days over 1780 years; this time interval is unprecedentedly long for an eddy-resolving ocean model, but our study requires this length to obtain highly accurate statistics of the flow. Each snapshot of the solution was coarse-grained by a local running-window averaging with a 120-km width to a  $65^2$  grid with 60-km spacing. We checked that the outcome is not sensitive to moderate variations of the running-window width.

### 4.2. Data-Adaptive Harmonic Decomposition

The data-adaptive harmonic decomposition (DAHD) [55] is a signal-processing technique that allows for a decomposition of the power and phase spectra via data-adaptive modes within a time-embedded phase space. Unlike other techniques exploiting time-embedding, such as M-SSA [25] or its nonlinear version [28], DAHD exploits a combination of integral operator and semigroup techniques [76] that help decompose the original signal into elementary signals that are narrowband for each separate discrete Fourier frequency, while being data-adaptive.

The mathematical details of the approach are provided in [55] within a general framework, including the case of multivariate time series issued from a mixing dynamical system, either stochastic or deterministic. Central to the approach is the spectral analysis of a class of integral operators whose kernels are built from correlation functions in a quite different way than found in principal component analysis (PCA) and its generalizations. For the sake of simplicity, we recall first from [55] how such an integral operator is constructed in the case of a one-dimensional time series  $X(t)$ . Given the two-sided autocorrelation function (ACF),  $\rho$  (of  $X(t)$ ), estimated on the interval  $I = [-\tau/2, \tau/2]$ , such an operator takes the form:

$$\mathcal{L}_\rho(\Psi)(r) = \frac{1}{\tau} \left( \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}-r} \rho(s+r)\Psi(s) ds + \int_{\frac{\tau}{2}-r}^{\frac{\tau}{2}} \rho(r+s-\tau)\Psi(s) ds \right), \quad r \in I \quad (10)$$

and acts on any square-integrable function  $\Psi$  on the interval  $I$ . The parameter  $\tau > 0$  characterizes the embedding window, but is chosen in practice so that  $\rho(t)$  sufficiently decays over  $[-\tau/2, \tau/2]$ .

At a practical level, the discretization of the operator  $\mathcal{L}_\rho$  defined by (10) leads to Hankel matrices built from temporal correlations fundamentally different than in M-SSA and alike. For multivariate time series, the ACF,  $\rho$ , is replaced by time-lagged cross-correlations, and operators such as given by (10) are grouped into a block operator whose discretization results into block-Hankel matrices; see (12) below and ([55], Sec. VI-D). The aforementioned DAHMs are then obtained as eigenvectors of such a block-Hankel matrix, while the corresponding eigenvalues provide a notion of energy contained into the signal that, while allowing for a reconstruction of the signal, is not equivalent to variance; see ([55], Remark V.1-(ii)). The main properties of DAH spectral elements are given below.

#### 4.2.1. DAH Eigenelements and Power Spectrum

Given a multivariate time series formed of  $d$  “channels” sampled uniformly at a unit of time  $\delta t$ , i.e.,  $X(t_n) = (X_1(t_n), \dots, X_d(t_n))$ , with  $t_n = n\delta t$ ,  $n = 1, \dots, N$ , the first step to determine the DAH spectra consists of estimating the two-sided cross-correlation coefficients  $\rho_k^{(p,q)}$  between channels  $p$  and  $q$  at lag  $k$  up to a maximum lag  $M - 1$ , i.e.,  $-M + 1 \leq k \leq M - 1$ .

As shown in ([55], Sec. VI-D), the discretization of the operator  $\mathcal{L}_\rho$  given by (10) with  $\rho = \rho^{(p,q)}$  leads to the following Hankel matrix  $\mathbf{H}^{(p,q)}$ :

$$\mathbf{H}^{(p,q)} = \begin{pmatrix} \rho_{-M+1}^{(p,q)} & \rho_{-M+2}^{(p,q)} & \cdots & \rho_0^{(p,q)} & \rho_1^{(p,q)} & \cdots & \rho_{M-1}^{(p,q)} \\ \rho_{-M+2}^{(p,q)} & \ddots & \ddots & \ddots & \ddots & \ddots & \rho_{-M+1}^{(p,q)} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \rho_{-M+2}^{(p,q)} \\ \rho_0^{(p,q)} & \ddots & \ddots & \ddots & \rho_{-M+1}^{(p,q)} & \ddots & \vdots \\ \rho_1^{(p,q)} & \ddots & \ddots & \ddots & \rho_{-M+2}^{(p,q)} & \ddots & \rho_0^{(p,q)} \\ \vdots & \rho_{M-1}^{(p,q)} & \rho_{-M+1}^{(p,q)} & \ddots & \ddots & \ddots & \vdots \\ \rho_{M-1}^{(p,q)} & \rho_{-M+1}^{(p,q)} & \rho_{-M+2}^{(p,q)} & \cdots & \rho_0^{(p,q)} & \cdots & \rho_{M-2}^{(p,q)} \end{pmatrix} \quad (11)$$

By forming such a Hankel matrix for each  $(p, q)$  in  $\{1, \dots, d\}^2$ , one can assemble the following block-Hankel matrix  $\mathbf{C}$  composed by  $d^2$  blocks of size  $(2M - 1) \times (2M - 1)$ , each given according to:

$$\begin{aligned} \mathbf{C}^{(p,q)} &= \mathbf{H}^{(p,q)}, \text{ if } 1 \leq p \leq q \leq d \\ \mathbf{C}^{(p,q)} &= \mathbf{H}^{(q,p)}, \text{ else} \end{aligned} \quad (12)$$

Note that because each building block,  $\mathbf{H}^{(p,q)}$ , is symmetric and because  $\mathbf{C}^{(p,q)} = \mathbf{C}^{(q,p)}$ , the grand matrix  $\mathbf{C}$  is itself symmetric, and therefore, its eigenvalues are necessarily real while the eigenvectors form an orthogonal set.

Furthermore, Theorem V.1 of [55] shows that the corresponding eigenvalues of  $\mathfrak{C}$  come in pairs of equal absolute values, but with the opposite sign. The eigenvalues can be grouped per Fourier frequency  $f$  ( $\neq 0$ ) and are actually determined by the singular values of a cross-spectral matrix at each frequency. In particular, denoting by  $\widehat{\rho^{p,q}}(f)$  the Fourier transform at the frequency  $f$  of the cross-correlation function  $\rho^{p,q}$ , we consider the following  $d \times d$  cross-spectral matrix  $\mathfrak{S}(f)$  whose entries are given by:

$$\mathfrak{S}_{p,q}^k = \begin{cases} \widehat{\rho^{p,q}}(f) & \text{if } q \geq p \\ \widehat{\rho^{q,p}}(f) & \text{if } q < p \end{cases} \quad (13)$$

Then, Theorem V.1 of [55] shows that for each singular value  $\sigma_k(f)$  of  $\mathfrak{S}(f)$  there exists, when  $f \neq 0$ , a pair of negative-positive eigenvalues  $(\lambda_-^k(f), \lambda_+^k(f))$  of  $\mathfrak{C}$  such that:

$$\lambda_+^k(f) = -\lambda_-^k(f) = \sigma_k(f), \quad 1 \leq k \leq d \quad (14)$$

i.e.,  $2d$  eigenvalues are associated with each Fourier frequency  $f \neq 0$ . The same theorem shows that  $d$  (but not paired) eigenvalues are associated with the frequency  $f = 0$ .

This property allows one to rearrange the eigenvalues into the DAH power spectrum that consists of forming, for each  $\ell$  ranging from  $1-(M'+1)/2$ , the discrete set:

$$\mathcal{P}_\ell = \{|\lambda_j|, : j \in \mathcal{J}(f_\ell)\} \quad (15)$$

where:

$$f_\ell = \frac{2\pi(\ell-1)}{M'-1}, \quad \ell = 1, \dots, \frac{M'+1}{2} \quad (16)$$

Hereafter, we use  $M' = 2M - 1$  for concision, re-indexing the string  $\{-M+1, \dots, M-1\}$  to run from  $1-M'$  as necessary.

Furthermore, Theorem V.1 of [55] shows that the eigenvectors  $\mathbf{W}_j$  of  $\mathfrak{C}$  can also be grouped per Fourier frequency  $f$  by using the following representation:

$$\mathbf{W}_j = (\mathbf{E}_1^j, \dots, \mathbf{E}_d^j)^{\text{tr}} \quad (17)$$

where DAHM snippet  $\mathbf{E}_k^j$  is a  $M'$ -long row vector that is explicitly associated with a Fourier frequency  $f$  according to:

$$\mathbf{E}_k^j(s) = B_k^j \cos(fs + \theta_k^j), \quad 1 \leq s \leq M' \quad (18)$$

where the amplitudes  $B_k^j$  and the phases  $\theta_k^j$  are both data-dependent, for each  $k$  in  $\{1, \dots, d\}$ . Thus, we can easily sort the eigenvectors according to a given Fourier frequency  $f_\ell$  in (16), by determining the following subset of indices in  $\{1, \dots, dM'\}$ :

$$\mathcal{J}(f_\ell) = \{j : \text{s.t. (18) holds with } f = f_\ell\} \quad (19)$$

Note that  $\mathcal{J}(f_\ell)$  is composed of  $2d$  indices when  $\ell \neq 0$  and of  $d$  indices if  $\ell = 0$ , and therefore,  $\mathcal{J}(f_\ell)$ 's form a partition of the total set of indices,  $\{1, \dots, dM'\}$ .

#### 4.2.2. DAH Coefficients

Another useful property concerns the pair of DAHMs associated with a pair of DAH eigenvalues  $(\lambda_j, \lambda_{j'})$ , such that  $\lambda_{j'} = -\lambda_j$  with  $j$  and  $j'$  that belong thus to the same subset  $\mathcal{J}(f)$ . For such a DAHM pair, the theory shows indeed that their corresponding phases satisfy  $\theta_k^{j'} = \theta_k^j + \pi/2$ , i.e., in each DAHM pair, the modes are shifted by one fourth of the period; see Theorem IV.1 of [55]. The DAHMs are thus always in exact phase quadrature, as for a sine-and-cosine pair in Fourier analysis, but in a data-adaptive fashion as encapsulated in the  $\theta_k^j$ 's and the  $B_k^j$ 's.

By analogy with M-SSA [25], the multivariate dataset  $\mathbf{X}$  can be projected onto the orthogonal set formed by the  $\mathbf{W}_j$ 's, in order to obtain the following DAH expansion coefficients (DAHCs):

$$\xi_j(t) = \sum_{s=1}^{M'} \sum_{k=1}^d X_k(t+s-1) \mathbf{E}_k^j(s) \quad (20)$$

where  $t$  varies from one to:

$$N' = N - M' + 1 \quad (21)$$

Although the DAHCs are not formally orthogonal in time, the DAHC pair  $(\xi_j(t), \xi_{j'}(t))$  associated with a DAHM pair  $(\mathbf{W}_j, \mathbf{W}_{j'})$ , is composed of time series that are nearly in phase quadrature; a property that is all the more pronounced when the embedding window parameter  $M$  can be made sufficiently large to resolve the decay of temporal correlations contained in  $\mathbf{X}$ ; see [55]. In other words, the larger  $M$  (subject to the length of the record), the more apparent is the phase quadrature property exhibited by  $\xi_j(t)$  and  $\xi_{j'}(t)$  constituting a DAHC pair.

Furthermore, any subset of DAHCs, as well as the their full set, can be convolved with its corresponding set of DAHMs, to produce a partial or full reconstruction of the original dataset, respectively. Thus, the following  $j$ -th reconstructed component (RC) at time  $t$  and for channel  $k$  is defined as:

$$R_k^j(t) = \frac{1}{M_t} \sum_{s=L_t}^{U_t} \xi_j(t-s+1) \mathbf{E}_k^j(s), \quad 1 \leq s \leq M' \quad (22)$$

where  $L_t$  (resp.  $U_t$ ) is a lower (resp. upper) bound in  $\{1, \dots, M'\}$ , that is allowed to depend on time. The normalization factor  $M_t$  equals  $M'$ , except near the ends of the time series, as in M-SSA [25], and the sum of all the RCs recovers the original time series.

Due to the ordering of the DAHMs in terms of Fourier frequency, the harmonic reconstruction component (HRC) can be formed from the sum of the RC pairs associated with a same Fourier frequency  $f \neq 0$ , namely:

$$R_k(t; f) = \sum_{j \in \mathcal{J}(f)} R_k^j(t) \quad (23)$$

where  $\mathcal{J}(f)$  denotes the set of indices given in (19). HRC provides an unambiguous and natural way to determine how variability at a particular frequency  $f$  is expressed in the time domain. For instance, Panels (a) and (b) of Figure 7 show the sum of the first four HRCs on PC1 for the upper-layer stream function anomalies as simulated from the QG model and its DAH-MSLM emulator, respectively.

#### 4.3. Multilayer Stuart–Landau Models

DAHD facilitates efficient inverse modeling of a given multivariate dataset  $\mathbf{X}$  in the transformed coordinates of DAHCs (20) and then utilizes the reconstruction Formulas (22) and (23) for transformation into the space of the original variables of  $\mathbf{X}$ . As we explain below, the modeling of DAHCs is simplified due to (i) the near-phase quadrature property satisfied by any DAHC pair and (ii) its narrowband character.

Let us consider a DAHC pair  $(\xi_j(t), \xi_{j'}(t))$  associated with a pair of DAH eigenvalues  $(\lambda_j, \lambda_{j'})$ , such that  $\lambda_{j'} = -\lambda_j$  with  $j$  and  $j'$  that belong thus to the same subset  $\mathcal{J}(f)$  associated with a frequency  $f$ . Moreover, without loss of generality, we can assume that  $\lambda_j$  is positive and  $\lambda_{j'}$  is negative. Hereafter, we also assume the time  $t$  to be a continuous parameter. For such a DAHC pair, we form the complex time series,  $\zeta_j(t) = \xi_j(t) + i\xi_{j'}(t)$  where  $i^2 = -1$ . As shown in Section VII of [55], we can infer that:

$$\xi_j(t) = \Re(\zeta_j(0)e^{-ift}) + \sum_{k=1}^d B_k^j \int_0^t \cos\left(-(t-r)f + \theta_k^j\right) \left(X_k\left(r + \frac{\tau}{2}\right) - X_k\left(r - \frac{\tau}{2}\right)\right) dr \quad (24)$$

Since the convolution term in the RHS of (24) involves a cosine function oscillating at the frequency  $f$ , the frequencies  $g \neq f$  contained in the time series  $X_k(t + \tau/2) - X_k(t - \tau/2)$  are filtered out, and the resulting DAHC,  $\xi_j(t)$ , is narrowband about the frequency  $f$ . This latter property is shared by all DAHCs,  $\xi_j(t)$ , with  $j$  in  $\mathcal{J}(f)$ , and in particular for  $\xi_{j'}(t)$ . Although the RHS of (24) provides an exact representation of the DAHC (in the case of an infinite time series), its usage in practice is limited since we desire a closure model of the  $\xi_j(t)$ 's that avoids an explicit dependence on the data,  $X_k$ , as found in (24).

Guided by the fact that a DAHC pair  $(\xi_j(t), \xi_{j'}(t))$  is composed of narrowband time series that are nearly in phase quadrature, Chekroun and Kondrashov [55] have shown that the class of Stuart–Landau (SL) oscillators driven by an additive noise [77,78] represent a natural class of closure models to capture amplitude modulations of  $\xi_j(t)$  and  $\xi_{j'}(t)$ :

$$\dot{z} = (\mu + i\gamma)z - (1 + i\beta)|z|^2z + \epsilon_t, \quad z \in \mathbb{C} \quad (25)$$

where  $\mu, \gamma$  and  $\beta$  are real parameters,  $\epsilon_t$  is a noise term and  $z(t) = \xi_j(t) + i\xi_{j'}(t)$ . Complementarily, Appendices A–D below justify the modeling of DAHCs within the class of SL oscillators by the theory of Ruelle–Pollicott (RP) resonances and their estimation from the time series. Furthermore, the collective behavior of all DAHC pairs associated with the same frequency  $f$  must be also taken into account by using an appropriate dynamical coupling between the corresponding individual SL oscillators, as well as by considering temporal and spatial cross-pair correlations in the noise term  $\epsilon_t$ .

The multilayer MSM framework of [54] is particularly suited to deal with these issues and applied to (25), it leads to MSLMs such as introduced in [55]. For this study, the optimal MSLM has only a main layer according to the stopping criterion described in Appendix A of [54], and it is given as the following system of SDEs driven by pairwise-correlated white noise:

$$\begin{aligned} dx_j &= \left( \beta_j(f)x_j - \alpha_j(f)y_j - \sigma_j(f)x_j(x_j^2 + y_j^2) + \sum_{\substack{i \neq j \\ i \in \mathcal{J}_d(f)}} b_{ij}^x(f)x_i + \sum_{\substack{i \neq j \\ i \in \mathcal{J}_d(f)}} a_{ij}^x(f)y_i \right) dt + d\epsilon_j^x \\ dy_j &= \left( \alpha_j(f)x_j + \beta_j(f)y_j - \sigma_j(f)y_j(x_j^2 + y_j^2) + \sum_{\substack{i \neq j \\ i \in \mathcal{J}_d(f)}} a_{ij}^y(f)x_i + \sum_{\substack{i \neq j \\ i \in \mathcal{J}_d(f)}} b_{ij}^y(f)y_i \right) dt + d\epsilon_j^y \\ d\epsilon_j^x &= Q_{11}^{j,j}(f) dW_1^j + Q_{12}^{j,j}(f) dW_2^j + \sum_{\substack{i \neq j \\ i \in \mathcal{J}_d(f)}} \sum_{k=1}^2 Q_{1k}^{j,j}(f) dW_k^i \\ d\epsilon_j^y &= Q_{21}^{j,j}(f) dW_1^j + Q_{22}^{j,j}(f) dW_2^j + \sum_{\substack{i \neq j \\ i \in \mathcal{J}_d(f)}} \sum_{k=1}^2 Q_{2k}^{j,j}(f) dW_k^i \end{aligned} \quad (26)$$

Appendices C and D below provide, in the context of the data collected from the oceanic model [8], numerical evidences and theoretical arguments for the modeling of the corresponding DAHCs by MSLMs such as (26). As mentioned above, the theoretical apparatus is in particular based on the theory of Ruelle–Pollicott resonances and their estimation from time series [56], recalled in Appendices A and B below.

Here,  $(x_j(t), y_j(t))$  is aimed at approximating a DAHC pair  $(\xi_j(t), \xi_{j'}(t))$  associated with a frequency  $f = f_\ell$ , and the index  $j$  belongs to the subset of indices  $\mathcal{J}(f_\ell)$  associated with  $d$  distinct pairs. The  $W_k^j$ 's with  $k = 1$  or  $k = 2$ , and  $1 \leq j \leq d$ , form  $2d$  independent Brownian motions.

Following Kondrashov et al. [54], all model coefficients are estimated for each  $(x_j, y_j)$ -pair by (multiple) linear regression (MLR). Linear constraints on  $\alpha_j(f), \beta_j(f)$  and  $\sigma_j(f)$  are imposed to ensure antisymmetry for the linear part of each  $(x_j, y_j)$ -pair, as well as equal and positive values  $\sigma_j(f) > 0$  to ensure asymptotic stability. The resulting regression residuals yield time series of  $(d\epsilon_j^x, d\epsilon_j^y)$  that are well approximated by pairwise-correlated white noise in the sums involving the  $W_k^j$ 's; see Appendix D for more details. Note that an MSLM, as any MSM, may include more layers to accurately model the

regression residual; see Section VIII-B of [55]. It turned out that for the simulations reported in this study, an MSLM of the form (26) was sufficient to reach satisfactory modeling skills; see Figures 8 and 9 above, as well as Appendix E. Note also that the DAHCs associated with  $f \equiv 0$  are not paired and resemble red noises in practice (see [55]); thus, only a linear MSM is used to model the DAHCs associated with the zero-frequency; see [54].

Because the SL oscillators are uncoupled across the frequencies, the DAH-MSLM approach is computationally efficient, totally parallelizable and, for a variety of datasets, laptop-enabled in practice. First, the model coefficients can be estimated in parallel for each frequency, and the overall number of independent coefficients to estimate remains small and fixed for each  $(x_j, y_j)$ -pair. Once all the resulting (few) MSLM coefficients have been estimated, for the simulation, as well, no extra coupling across the frequencies is needed other than running the MSLMs across the frequencies by the same noise realization, which can be also done in parallel.

To simulate the DAHC pairs, Equations (26) are discretized in time and integrated numerically forward using a Euler–Maruyama scheme from initial states obtained according to the initialization procedure described in Appendix B of [54], followed by a convolution with the associated DAHMs according to (22) to transform into the space of original variables. The simulated RCs are then added in each frequency bin to yield the corresponding HRCs given by (23) that, in turn, are added across the frequencies in the range of interest. For instance, Panel (b) of Figure 7 shows the sum of the first four HRCs on PC1 for the upper-layer stream function anomalies as simulated for the DAH-MSLM emulator of the QG model considered here.

**Acknowledgments:** This research was supported by the National Science Foundation grants OCE-1243175 and OCE-1658357 (Dmitri Kondrashov and Mickaël D. Chekroun) and DMS-1616981 (Mickaël D. Chekroun). Pavel Berloff was supported by the NERC grant NE/R011567/1.

**Author Contributions:** Dmitri Kondrashov, Mickaël D. Chekroun and Pavel Berloff conceived of and designed the experiments. Dmitri Kondrashov and Mickaël D. Chekroun performed the experiments. Dmitri Kondrashov and Mickaël D. Chekroun analyzed the data. Dmitri Kondrashov, Mickaël D. Chekroun and Pavel Berloff wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

DAH	Data-adaptive harmonic
DAHD	Data-adaptive harmonic decomposition
DAHC	Data-adaptive harmonic coefficient
DAHM	Data-adaptive harmonic mode
EOF	Empirical orthogonal function
HRC	Harmonic reconstructed component
LFV	Low frequency variability
MSLM	Multilayer Stuart–Landau model
PCA	Principal component analysis
PSD	Power spectral density
RC	Reconstructed component
RP	Ruelle–Pollicott
SDE	Stochastic differential equation
SL	Stuart–Landau

## Appendix A. Time Variability of Stochastic Systems and Ruelle–Pollicott Resonances

We communicate in this Appendix and the subsequent ones the mathematical foundations for the modeling of DAHCs by the class of Stuart–Landau models (25) and their generalization given by (26). The theoretical apparatus is based on the theory of Ruelle–Pollicott (RP) resonances and their numerical estimation from time series, following [56]. We recall below the main ingredients of this

theory, in particular regarding the fundamental role that RP resonances play in the decomposition of power spectral density (PSD) and correlation functions and that we apply to the case of time series constituted by the dominant pairs of DAHCs for the three frequency bands of interest in this study, namely decadal, interannual and subannual.

Given a system of stochastic differential equations (SDEs) in  $\mathbb{R}^p$ ,

$$dX = F(X) dt + G(X) dW_t. \quad (\text{A1})$$

Here,  $W_t = (W_t^1, \dots, W_t^q)$  denotes an  $\mathbb{R}^q$ -valued Wiener process whose components are mutually independent standard Brownian motions.

In Equation (A1), the drift part is provided by a (possibly nonlinear) vector field  $F$  of  $\mathbb{R}^p$ , and the (also possibly nonlinear) stochastic diffusion in its Itô version, given by  $G(X) dW_t$ , has its  $i$ -th-component ( $1 \leq i \leq p$ ) given by:

$$\sum_{j=1}^q G_{ij}(X) dW_t^j, \quad q \geq 1 \quad (\text{A2})$$

In what follows, we assume that the vector field  $F$  and the matrix-valued function:

$$G : \mathbb{R}^p \rightarrow \text{Mat}_{\mathbb{R}}(p \times q)$$

satisfy regularity conditions that guarantee the existence and the uniqueness of mild solutions, as well as the continuity of the trajectories; see, e.g., [79,80] for such conditions in the case of locally Lipschitz coefficients.

It is well known that the evolution of the probability density of the random  $\mathbb{R}^p$ -valued variable,  $X_t$  (solving Equation (A1)), is governed by the Fokker–Planck equation:

$$\partial_t \rho(X, t) = \mathcal{A} \rho(X, t) = -\text{div}(\rho(X, t) F(X)) + \frac{1}{2} \text{div} \nabla(\Sigma(X) \rho(X, t)), \quad X \in \mathbb{R}^p \quad (\text{A3})$$

with  $\Sigma(X) = G(X) G(X)^T$ .

What is less-known however is that the spectral properties of the second-order differential operator,  $\mathcal{A}$ , informs about fundamental objects such as the power spectra or correlation functions computed typically along a (random) trajectory of Equation (A1). We briefly recall the main elements hereafter and refer to [56,71] for more details.

First, given an observable  $\varphi : \mathbb{R}^p \rightarrow \mathbb{R}$  for the system (A1), we recall that the correlation spectrum  $S_\varphi(f)$  is obtained by taking the Fourier transform of the correlation function  $C_\varphi(\tau)$ , namely:

$$S_\varphi(f) = \widehat{C_\varphi(\tau)}, \quad \text{with } C_\varphi(\tau) = \lim_{T \rightarrow \infty} \int_0^T \varphi(X_{t+\tau}^x) \varphi(X_t^x) dt \quad (\text{A4})$$

where  $X_t^x$  denotes the stochastic process solving (A1) and satisfying  $X_0^x = x$  at time  $t = 0$ .

As shown in [71], for a broad class of SDEs that possess an ergodic probability distribution  $\mu$ , the spectrum,  $\sigma(\mathcal{A})$ , of the Fokker–Planck operator,  $\mathcal{A}$ , is typically contained in the left-half complex plane,  $\{z \in \mathbb{C} \text{ s.t. } \Re(z) \leq 0\}$ , and its resolvent  $R(z) = (z \text{Id} - \mathcal{A})^{-1}$ , is a well-defined linear operator that satisfies:

$$S_\varphi(f) = \int_{\mathbb{R}^p} \varphi(X) [R(if)\varphi](X) d\mu \quad (\text{A5})$$

Here,  $f$  lies in the complex plane  $\mathbb{C}$ , and the poles of the resolvent  $R(if)$ , which correspond to the RP resonances, introduce singularities into  $S_\varphi(f)$ . Once the PSD is calculated, i.e., once  $|S_\varphi(f)|$  is computed with  $f$  taken to be real, these poles manifest themselves as peaks that stand out over a continuous background at the frequency  $f$  if the corresponding RP resonances with imaginary part  $f$  (or nearby) are close enough to the imaginary axis. The continuous background has its origin in

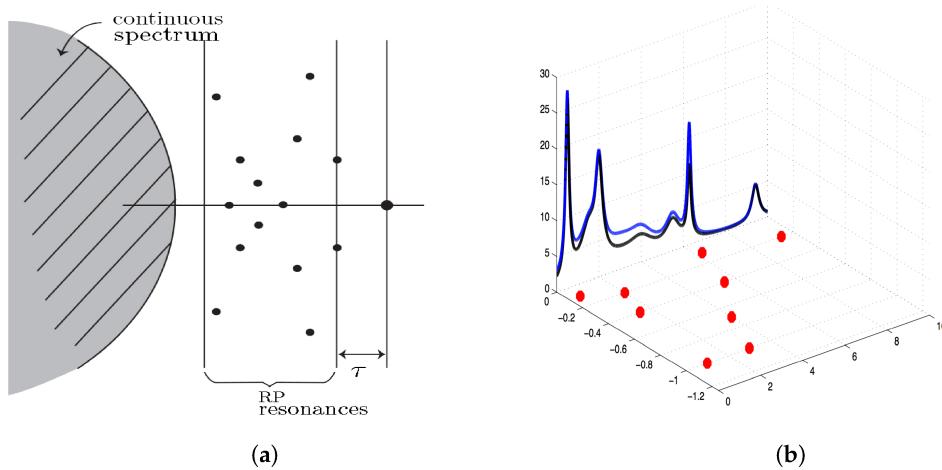
the continuous part of  $\sigma(\mathcal{A})$  lying typically in a sector  $\{z \in \mathbb{C} \text{ s.t. } \Re(z) \leq -\gamma\}$  (for some  $\gamma > 0$ ), while the RP resonances are the isolated eigenvalues of finite multiplicity of  $\mathcal{A}$ , lying within a strip  $-\gamma < \Re(z) \leq 0$ ; see Panel (A) of Figure A1.

Denoting by  $\lambda_j$ 's the  $N_p$  poles of the resolvent  $R(z)$  of  $\mathcal{A}$ , i.e., the Ruelle–Pollicott resonances, and by  $m_1, \dots, m_N$  their corresponding orders, the correlation function,  $C_\varphi(\tau)$ , possesses then the following expansion:

$$C_\varphi(\tau) = \sum_{j=1}^{N_p} \left[ \sum_{k=0}^{m_j-1} \frac{\tau^k}{k!} (\mathcal{A} - \lambda_j \text{Id})^k \right] a_j(\varphi) e^{\lambda_j \tau} + \mathcal{Q}(\tau) \quad (\text{A6})$$

where  $\mathcal{Q}(\tau)$  exhibits typically a decay property associated with properties of the essential spectrum of  $\mathcal{A}$ . Note that even if the  $\lambda_j$ 's do not depend on the observable  $\varphi$ , the coefficients,  $a_j(\varphi)$ 's, in the expansion (A6), do.

Historically, the resonances have been introduced by Ruelle [81] and Pollicott [82] for discrete and continuous chaotic deterministic systems (see [83] for the case of Anosov flows), but can be framed easily into a stochastic context, which benefits then from the tools of stochastic analysis, which allows, e.g., for justifying decomposition formula such as (A6) for a broad class of SDEs; see [70,71,84].



**Figure A1.** (a) The Ruelle–Pollicott (RP) resonances are the isolated eigenvalues of the Fokker–Planck operator,  $\mathcal{A}$ , defined in (A3); they are represented by red dots in (b) and by black dots here. The rightmost vertical line represents the imaginary axis above which the power spectrum lies; see (a) for another perspective. The rate of decay of correlations is controlled by the spectral gap  $\tau$  (not to be confused with  $\tau$  in (A6)); see [56,85]. (b) The imaginary part of the RP resonances corresponds to the location of a peak in the PSD (black curve lying above the imaginary axis) and the real part to its width. In blue is represented a reconstruction of the PSD based on RPs; a discrepancy is shown here to emphasize that in practice, the RPs are very often only estimated/approximated; see [56,71] (courtesy of Maciej Zworski). (a) Schematic of the spectrum of  $\mathcal{A}$  given in (A3). (b) Correspondence between the PSD and RP resonances according to (A5).

## Appendix B. Estimating Resonances from Time Series: The Reduced RP Resonances

Because they are related to the Fokker–Planck operator,  $\mathcal{A}$ , the RP resonances inform about the structure of the underlying SDE. A natural question then arises: is it possible to infer the “shape” of an SDE (i.e.,  $F$  and  $G$  in (A1)) from the observation of its solutions? The short answer to this question is “no” in general, except in certain cases (see [86]), but as we will see, a reduced spectrum, related to RP resonances, can be estimated from time series and depending on its geometry may inform us greatly in turn about the sought structural ingredients. In that respect, the approach of [56] paved a roadmap for addressing this question from a practical viewpoint; we refer to [71] for theoretical foundations.

In practice, only partial observations of the solutions to (A1) are available, e.g., few solution components. Speaking roughly, it is shown in [56] that given partial observations of a complex system that lie in a reduced phase space  $V$ , a Markov operator  $\mathfrak{T}$  with state space  $V$  can be inferred from these observations such that (i)  $\mathfrak{T}$  characterizes the coarse-graining in  $V$  of the transition probabilities in the full state space and (ii) the spectrum of  $\mathfrak{T}$  relates to the RP resonances, but in an averaged sense; see also [29,87]. In practice, the dimension of  $V$  is kept low so that  $\mathfrak{T}$  can be efficiently estimated via a maximum likelihood estimator (MLE). We detail below this procedure and what (ii) means, in the context of DAHCs.

Our standing assumption is that for each frequency  $f \neq 0$ , there exists a model of the form (A1) of which a solution is given by the set of DAHC-pairs,  $(\xi_j(t), \xi_{j'}(t))$ , at frequency  $f$ . Under this working assumption, since there is a total of  $d$  pairs, our phase space is here  $\mathbb{R}^p$  with  $p = 2d$ . Our observations are made out of a single pair of DAHCs,  $(\xi_j(t), \xi_{j'}(t))$ , so that  $\dim(V) = 2$  here. We denote by  $\varphi : \mathbb{R}^p \rightarrow V$  the corresponding mapping and take  $t = t_n$ , i.e., discrete time instants, as given by the multiple of the sampling time at which the DAHCs are computed, here 20 days.

The Markov operator  $\mathfrak{T}$  is approximated by the transition matrix  $P$  whose entries are given:

$$P_{kl} = \frac{\#\left\{(\xi_j(t_n), \xi_{j'}(t_n)) \in J_k \wedge (\xi_j(t_{n+1}), \xi_{j'}(t_{n+1})) \in J_l\right\}}{\#\left\{(\xi_j(t_n), \xi_{j'}(t_n)) \in J_k\right\}} \quad (\text{A7})$$

where the  $J_k$ 's form a partition (of size  $M \times M$ ) of a domain  $\mathcal{D}$  in  $\mathcal{V}$  containing the set of discrete points  $(\xi_k(t_n), \xi_{k'}(t_n))$ , for  $n = 1, \dots, N$ ; see, e.g., [56,87,88].

Assuming that the sought system of SDEs, (A1), possesses an ergodic statistical equilibrium,  $\mu$ , one can show that in the limit  $N \rightarrow \infty$ , the spectrum of  $P$  (after application of the (principal value of) logarithm) provides an approximation of the point spectrum of the following averaged Fokker–Planck operator, given formally for all  $\Psi$  in  $C^2(V)$  by:

$$\overline{\mathcal{A}}\Psi(v) = \int_Z \left( -\text{div}(\Psi(v)F(v, z)) + \frac{1}{2}\text{div}\nabla(\Sigma(v, z)\Psi(v)) \right) d\mu_v(z) \quad (\text{A8})$$

see [89]. The point spectrum  $\sigma_p(\overline{\mathcal{A}})$  of  $\overline{\mathcal{A}}$  provides what we call hereafter the reduced RP resonances of  $\mathcal{A}$ . In [71], it is called the reduced mixing spectrum, terminology that we have purposely changed here to avoid any confusion with the notion of mixing in the physical space; the mixing referred to in [71] takes place in the phase space and is typically manifested by decay of correlations; see also [56]. The point spectrum  $\sigma_p(\overline{\mathcal{A}})$  characterizes the mixing properties in the reduced phase space, once the effects of the variables lying in the unobserved factor  $Z$  have been averaged out. Thus, in short,

$$\log(\sigma(P)) \text{ approximates } \sigma_p(\overline{\mathcal{A}}) \quad (\text{A9})$$

More precisely, in (A8), the space  $Z$  denotes the supplement of  $V$  in  $\mathbb{R}^p$ , i.e.,  $\mathbb{R}^p = V \oplus Z$ , and  $\mu_v$  denotes the disintegration of  $\mu$  above  $v$  in  $V$ , along the unobserved factor space  $Z$ . Mathematically, for all Borel sets  $B$  and  $F$  of  $Z$  and  $V$ , respectively,

$$\mu(B \times F) = \int_{v \in F} \mu_v(B) dm(v) \quad (\text{A10})$$

with  $m = \Pi_V \mu$ ; see [54] and the references therein. Thus, the operator  $\overline{\mathcal{A}}$  corresponds to the conditional expectation of  $\mathcal{A}$  obtained after averaging along the unobserved variables lying in  $Z$ . Once the Markov matrix  $P$  is estimated from time series via (A7), its spectrum provides in turn an estimation of the reduced RP resonances and thus may inform us about the nature of the coefficients of  $\overline{\mathcal{A}}$  provided that the spectrum of  $P$  exhibits interpretable features in terms of known SDEs.

To illustrate this point, we focus first on the dominant pair of DAHCs for the interannual frequency, namely  $f = 0.67 \text{ yr}^{-1}$ . The time series constituting this pair of DAHCs are shown in the left panel of Figure A2. The corresponding transition matrix  $P$  is estimated by using (A7) for this pair of DAHCs and the state space  $\mathcal{D}$  over which this matrix is estimated and taken to be  $\mathcal{D} = [m_1^\epsilon, M_1^\epsilon] \times [m_2^\epsilon, M_2^\epsilon]$  with:

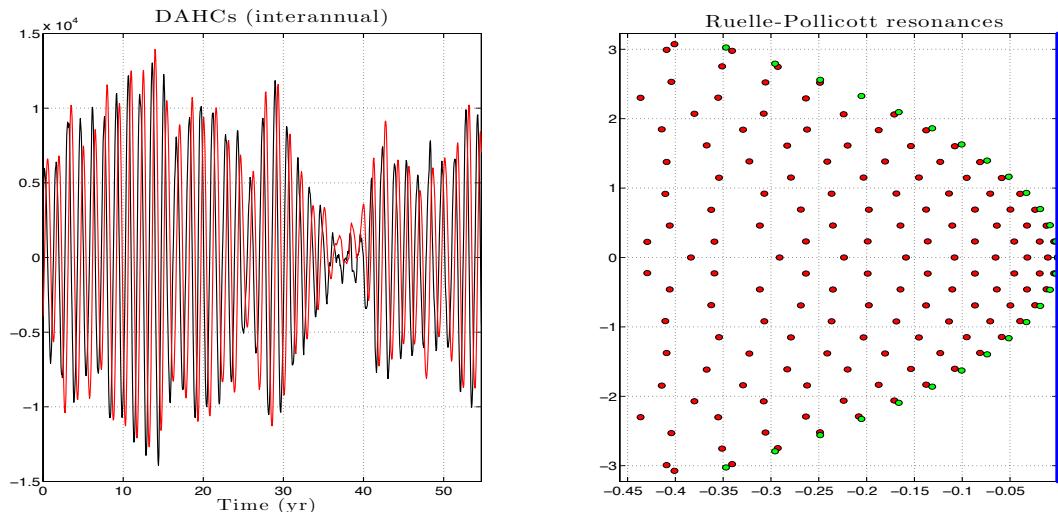
$$m_1^\epsilon = (1 + \epsilon) \min \{ \xi_j(t_n), n = 1, \dots, N \} \quad (\text{resp. } m_2^\epsilon = (1 + \epsilon) \min \{ \xi_{j'}(t_n), n = 1, \dots, N \}) \quad (\text{A11})$$

and:

$$M_1^\epsilon = (1 + \epsilon) \max \{ \xi_j(t_n), n = 1, \dots, N \} \quad (\text{resp. } M_2^\epsilon = (1 + \epsilon) \max \{ \xi_{j'}(t_n), n = 1, \dots, N \}) \quad (\text{A12})$$

for some  $\epsilon > 0$ . Here,  $N = 32,200$ , and a (non-uniform) partition of  $\mathcal{D}$  constituted by  $10^3$  cells  $J_k$ 's is considered.

The spectrum of  $P$  thus estimated is shown in the right panel of Figure A2 (in red), after application of the logarithm. It is striking to observe that the corresponding eigenvalues are organized into an array of shifted parabolas, at least for most of these eigenvalues. This special geometry reflects actually key structural properties about the averaged Fokker–Planck operator  $\bar{A}$  as we explain hereafter.



**Figure A2.** Reduced RP resonances and their analytic approximation. Left panel: The pair of DAHCs analyzed. Right panel: Corresponding eigenvalues of  $P$  as estimated by using (A7), after application of the logarithm (in red). The blue vertical line corresponds to the imaginary axis. These eigenvalues correspond to approximations of the point spectrum of the averaged Fokker–Planck operator given in (A8). The analytic approximations provided by (A17) are shown as green dots.

### Appendix C. RP Resonances of Stuart–Landau Models: Analytic Approximations

Let us consider the following special case of a Stuart–Landau model Equation (25) given in Cartesian coordinates by:

$$\begin{aligned} dx &= \left[ \beta x - \omega y - x(x^2 + y^2) \right] dt + \sigma dW_t^1 \\ dy &= \left[ \omega x + \beta y - y(x^2 + y^2) \right] dt + \sigma dW_t^2 \end{aligned} \quad (\text{A13})$$

In polar coordinates  $(r, \theta)$  with  $x = r \cos \theta$  and  $y = r \sin \theta$ , this system becomes by applying the Itô's formula:

$$\begin{aligned} dr &= \left( \beta r - r^3 + \frac{\sigma^2}{2r} \right) dt + \sigma dW_r \\ d\theta &= \omega dt + \frac{\sigma}{r} dW_\theta \end{aligned} \quad (\text{A14})$$

where  $W_r$  and  $W_\theta$  are two Wiener processes satisfying the relationships

$$\begin{aligned} dW_r &= \cos \theta dW_t^1 + \sin \theta dW_t^2 \\ dW_\theta &= -\sin \theta dW_t^1 + \cos \theta dW_t^2 \end{aligned}$$

The Fokker–Planck equation associated with (A14) is given by:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial r} \left[ \left( \beta r - r^3 + \frac{\sigma^2}{2r} \right) \rho \right] - \omega \frac{\partial}{\partial \theta} (\rho) + \frac{\sigma^2}{2} \frac{\partial^2 \rho}{\partial r^2} + \frac{1}{2} \left( \frac{\sigma}{r} \right)^2 \frac{\partial^2 \rho}{\partial \theta^2} \quad (\text{A15})$$

We assume now that  $\beta$  is sufficiently largely positive and  $\sigma$  is sufficiently small, so that roughly speaking, the dynamics of  $(x(t), y(t))$  settles near a circle of radius  $R_*$ , corresponding to an average radius of the stochastic limit cycle [90]. Under this assumption, it is expected thus that the azimuthal component of the solutions of (A15) can be well approximated by the solutions of the following advection-diffusion equation with periodic boundary conditions:

$$\frac{\partial u}{\partial t} = -\omega \frac{\partial u}{\partial \theta} + \frac{1}{2} \left( \frac{\sigma}{R_*} \right)^2 \frac{\partial^2 u}{\partial \theta^2} \quad (\text{A16})$$

The solutions of (A16) are formally given by:

$$u(t, \theta) = \sum_{k=-\infty}^{+\infty} a_k \exp \left( -k^2 \frac{\sigma^2}{2R_*^2} + ik\omega t \right) e^{-ik\theta}, \quad a_k \in \mathbb{C}$$

It is then easy to obtain the following approximation formula (up to a rescaling) for the eigenvalues,  $\lambda_k$ , of the Fokker–Planck operator's azimuthal part associated with Equation (A15):

$$\lambda_k \approx -2\pi^2 k^2 \sigma^2 + i \frac{2\pi k}{T}, \quad \text{and } \overline{\lambda_k}, \quad k \in \mathbb{Z} \quad (\text{A17})$$

We refer to [89] for more details; see also [91]. Note that here,  $T$  in (A17) denotes the period associated with the azimuthal velocity  $\omega$  such that  $\omega = 2\pi/T$ .

The analytic approximations given by (A17) are shown by the green dots in the right panel of Figure A2 for  $\sigma = 1.02 \times 10^{-2}$  and  $T = 540$  days. To such analytic approximation of the RP resonances associated with (A13), the parabolas of resonances that are shifted to the left of the complex plane by  $\gamma > 0$ , namely:

$$\lambda_k \approx -2\pi^2 k^2 \sigma^2 - \gamma + i \frac{2\pi k}{T}, \quad \text{and } \overline{\lambda_k}, \quad k \in \mathbb{Z} \quad (\text{A18})$$

constitute parabolas of harmonics; see [71]. These parabolas of harmonics, shown in green here for one of them to avoid overload, provide good approximations of most of the resonances shown (in red) in the right panel of Figure A2.

Note that in practice, the shift  $\gamma$  relates to the eigenvalues of the radial part  $\mathcal{A}_r$  of the Fokker–Planck operator associated with (A15), namely:

$$\mathcal{A}_r \Psi = \frac{\partial}{\partial r} \left[ \left( \beta r - r^3 + \frac{\sigma^2}{2r} \right) \Psi \right] + \frac{\sigma^2}{2} \frac{\partial^2 \Psi}{\partial r^2} \quad (\text{A19})$$

In particular, the smallest  $\gamma$  can be measured from the estimated resonances and informs on the value of  $\beta$  ( $> 0$  by assumption) via the formula:

$$\gamma = 2\sqrt{\beta^2 + 2\sigma^2} \quad (\text{A20})$$

while the other constants  $\gamma > 0$  appearing in (A18) are given by  $2k\sqrt{\beta^2 + 2\sigma^2}$  with  $k$  in  $\mathbb{Z}^+$ ; see again [71] and also [92].

Thus, the numerical estimation of the eigenvalues of the transition matrix  $P$  for the pair of DAHCs associated with the interannual frequency,  $f = 0.67 \text{ y}^{-1}$ , combined with the analytic description outlined above reveal that the Fokker–Planck operator associated with an SL model of the form (A13) constitutes a very good approximation of the (abstract) averaged Fokker–Planck operator  $\bar{\mathcal{A}}$  given by (A8). Such a spectral analysis strongly supports the idea of modeling DAHCs within the class of SL models, at least for the interannual variability of the flow. The next section of this Appendix discusses briefly the case of other frequency bands. As emphasized below, although more ingredients may be required for their good modeling, SL models still form a natural class for a coherent modeling of the subset of DAHC pairs associated with a given frequency  $f \neq 0$ .

#### Appendix D. Multilayer Stuart–Landau Modeling of DAHCs

After the material and numerical results presented in Appendices A–C above, we are in a position to explain the motivations behind the form (26) of MSLMs given in the main text. To understand these motivations, we formalize a bit more the writing of (26) by introducing the notations  $X_j = (x_j, y_j)^T$  to denote a DAHC pair and  $B_t^j = (W_{1,t}^j, W_{2,t}^j)^T$  to denote a two-dimensional Brownian motion.

The system (26) is written then as the following system of SDEs:

$$dX_j = \left( L_j(f)X_j + F_j(f; X_j) + \sum_{i=1}^d C_{ji}(f)X_i \right) dt + \sum_{i=1}^d M_{ji}(f)dB_t^i, \quad 1 \leq j \leq d \quad (\text{A21})$$

with:

$$L_j(f) = \begin{pmatrix} \beta_j(f) & -\alpha_j(f) \\ \alpha_j(f) & \beta_j(f) \end{pmatrix} \text{ and } F_j(f; X_j) = \begin{pmatrix} -\sigma_j(f)x_j(x_j^2 + y_j^2) \\ -\sigma_j(f)y_j(x_j^2 + y_j^2) \end{pmatrix} \quad (\text{A22})$$

$$M_{ji}(f) = \begin{pmatrix} Q_{11}^{ij}(f) & Q_{12}^{ij}(f) \\ Q_{21}^{ij}(f) & Q_{22}^{ij}(f) \end{pmatrix}, \text{ and } C_{ji}(f) = \begin{pmatrix} b_{ij}^x(f) & a_{ij}^x(f) \\ a_{ij}^y(f) & b_{ij}^y(f) \end{pmatrix} \quad (\text{A23})$$

Let us introduce the  $n$ -dimensional variable  $\xi$  such that  $\xi_j = x_j$ ,  $\xi_{j+1} = y_j$  and  $\xi = (\xi_1, \dots, \xi_n)$  with  $n = 2d$ . Then, (A21) can be re-written itself as:

$$d\xi = (L(f)\xi + F(f; \xi) + C(f)\xi) dt + G(f)dB_t, \quad 1 \leq j \leq d \quad (\text{A24})$$

Here,  $L(f)$  is a  $n \times n$  block-diagonal matrix, where each non-zero  $2 \times 2$  block is given by the  $L_j(f)$ 's,  $F(f; \xi)$  is the  $n$ -dimensional vector formed by the collection of two-dimensional vectors  $F_j(f; X_j)$  written in the  $\xi$ -variable and  $C(f)$  (resp.  $G(f)$ ) is the  $n \times n$  matrix whose  $2 \times 2$  blocks are given by the  $C_{ji}(f)$ 's (resp.  $M_{ji}(f)$ 's).

In practice, it is under the form (A24) that the coefficients are estimated from the time series of (all) the DAHC pairs associated with a common frequency  $f$ . This estimation is made in two steps: first, the coefficients in  $L$ ,  $F$  and  $C$  (respecting the constraints on  $L$  and  $F$ ) are estimated via, e.g., regression techniques, from which the  $2d$ -dimensional time series of residual  $\mathbf{r}_t$  is stored. By computing the  $2d \times 2d$  covariance matrix  $\mathbf{C}_r$  of  $\mathbf{r}_t$ , the grand matrix  $G(f)$  is then obtained via the Cholesky decomposition of  $\mathbf{C}_r$ , namely  $\mathbf{C}_r = G(f)G(f)^T$ .

The coupling terms are of two types: those brought by the matrix  $C(f)$  and those brought by the noise matrix  $G(f)$ . In the absence of these coupling terms, we are just left with a collection of

uncoupled SL models of the form (A13). Even for frequencies  $f$  (such as in Appendix B) for which individual SL models provide very plausible SDE generators of any given pair of DAHCs at the frequency  $f$  (such as explained in Appendix C), it is unlikely to obtain in this simplified fashion a coherent simulation of the DAHCs (associated with frequency  $f$ ) that would share phase relationships consistent with those exhibited by the original DAHCs (at the same frequency). It is one of the main motivations for the presence of such coupling terms in (A21) and thus in (A24).

In the presence of coupling terms, it is noteworthy that the averaged Fokker–Planck operator associated with (A24) differs from that associated with an individual SL model, i.e., associated with:

$$dX_j = \left( L_j(f)X_j + F_j(f; X_j) \right) dt + M_{jj}(f)dB_t^j \quad (\text{A25})$$

by first-order and second-order partial differential operators involving averaging over a combination of coefficients in  $C(f)$  and  $G(f)$ , respectively.

The analysis conducted in Appendix C suggests thus that any “deviation” from parabolas of the arrangement of resonances could serve as a good indication that the presence of such derivatives in the averaged Fokker–Planck operator is non-negligible (emphasizing thus the presence of coupling terms), although such a statement should be taken with care due not only to possible estimation errors that could lie behind such discrepancies, but also to other factors governed e.g., by the ratio between  $\beta$  and  $\sigma$ , adopting the notations of Appendix C; see [89] for more details.

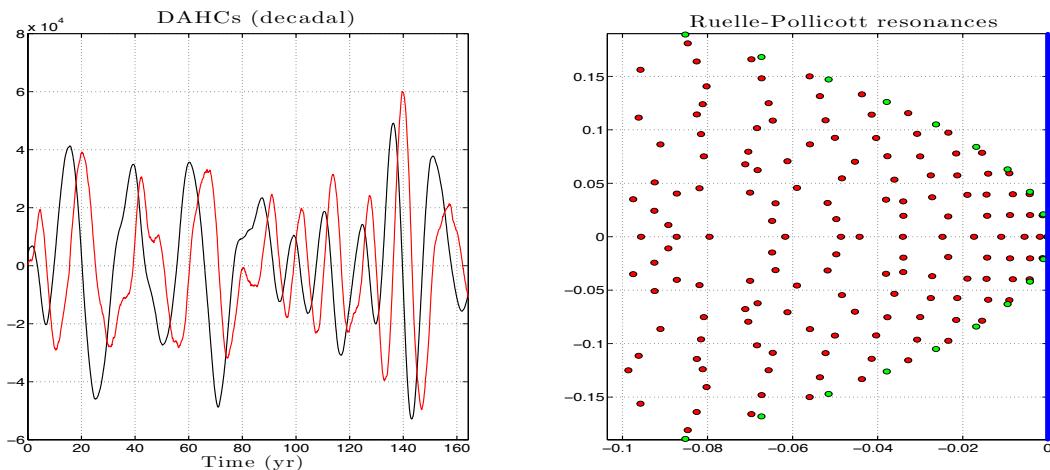
Discrepancies with respect to an array of parabolas are observed in the right panel of Figure A3 for the RP resonances estimated from the dominant pair of DAHCs shown in the left panel of Figure A3, for the frequency  $f = 0.061$  cycle/ $y^{-1}$ , i.e., corresponding to decadal variability ( $T \approx 16.39y$ ). The reason here lies presumably in the less precise phase-quadrature relationship that the time series composing such a pair satisfies, as pointed out already in Section 2.3 for the pairs of DAHCs shown in Figure 3 associated with decadal variability. Actually, not only the phase-quadrature relationship among the time series constituting a DAHC pair is altered at low-frequency, but also their narrowband character unlike for the interannual and subannual cases (not shown). This alteration of such narrowband features is actually rooted into windowing effects since the 16.4-y-period is actually pretty close to the size of the window ( $\approx 17$  y) used to estimate the correlations underlying the DAH grand matrix,  $\mathfrak{C}$ , in (12).

Thus, we may naturally expect that a single SL model (for which in particular such phase-quadratures are typically observed) is insufficient for modeling purposes. Nevertheless, the outer envelope of these resonances still follows closely a parabola as shown by the green dots in Figure A3 obtained by application of the analytic Formula (A18) with  $T = 17$  y,  $\gamma = 0$  and  $\sigma = 3 \times 10^{-3}$ . The coupling terms  $C(f)$  in (A24) and the noise matrix  $G(f)$  are aimed at modeling the differences from an array of parabolas within the envelop’s convex hull.

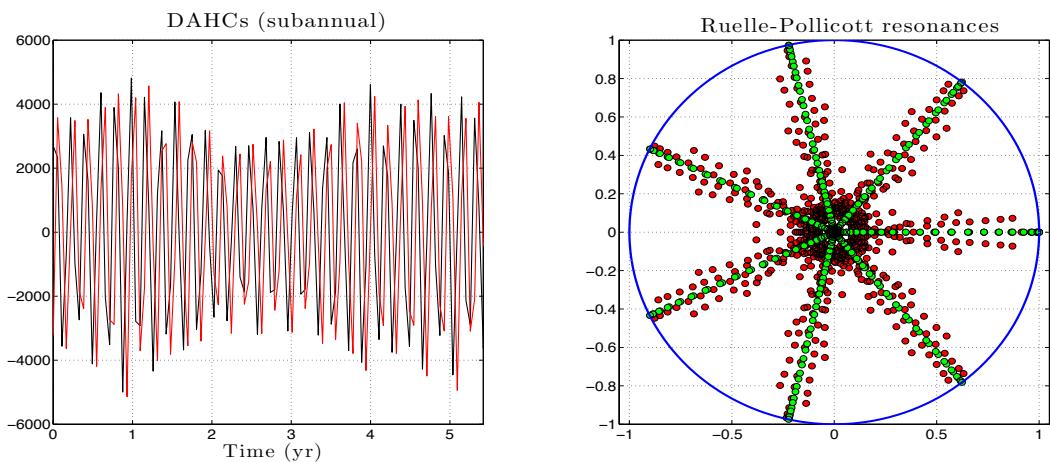
The subannual case supports also the relevance of the analytic prediction (A17) and thus of the class of Stuart–Landau models used to model the DAHCs for the corresponding frequencies, albeit in a less direct way. Here, the frequency associated with the pair of DAHCs analyzed hereafter is determined via the DAH spectrum given in Figure 2. The corresponding value of this frequency is  $f \approx 5.27$  cycle/ $y^{-1}$ . It lies close to the Nyquist frequency, and therefore, sampling effects are expected to manifest. For instance, whereas the dominant period should be thus  $T_p \approx 69$  days as predicted by the theory, an estimation of the resonances as eigenvalues of the Markov matrix  $P$  with entries given by (A7) shows an estimated period  $T_s$  of  $1/7$   $y^{-1}$  corresponding approximately to 52 days. The latter period is clearly visible by displaying the eigenvalues of  $P$  without applying the logarithm, unlike in Figures A2 and A3: seven strings of resonances are indeed apparent in the right panel of Figure A4 within the unit disc of the complex plane, i.e., by adopting the same mode of representation as in [56]. These resonances strings are very well approximated by rays of resonances obtained by application of the analytic Formula (A17), this time with  $T = T_s$  (and not with  $T = T_p$ ),  $\gamma = 0$  and  $\sigma = 3 \times 10^{-3}$ ; see the right panel of Figure A4. Note that a cluster of resonances has formed near

the center, this cluster corresponding to an approximation of the essential spectrum of the Markov operator  $\mathfrak{T}$  (see Appendix B) that  $P$  approximates; see [56].

In summary, the theory of Ruelle–Pollicott resonances and their estimation via time series [56] provide strong theoretical elements for the numerical modeling of DAHCs build up on MSLMs such as given by (26); see Section VIII of [55] for a generalization of (26). Note that the discussion conducted above about the subannual and decadal cases emphasizes that for practical purposes, it is not necessarily a good idea (for the estimation stage of an MSLM’s coefficients) to constrain the coefficients  $\alpha_j(f)$  in (A22) to be given by  $2\pi f$ , due to sampling or windowing effects such as manifested here for the subannual and decadal cases, respectively.



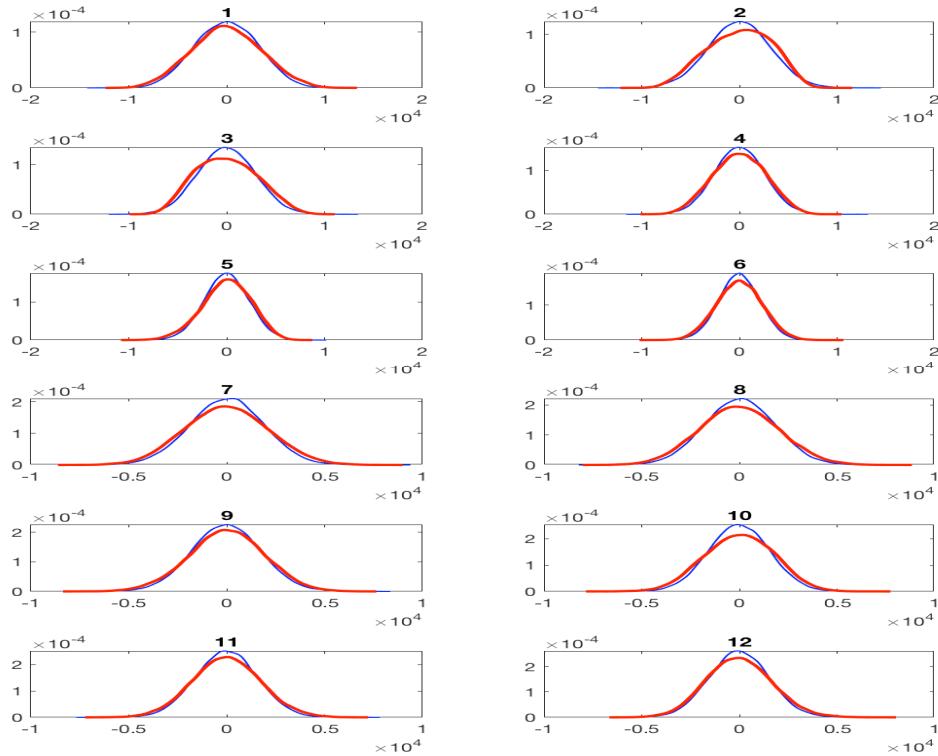
**Figure A3.** Reduced RP resonances and their analytic approximation. Same as in Figure A2, but for the frequency  $f = 0.061$  cycle/yr $^{-1}$ .



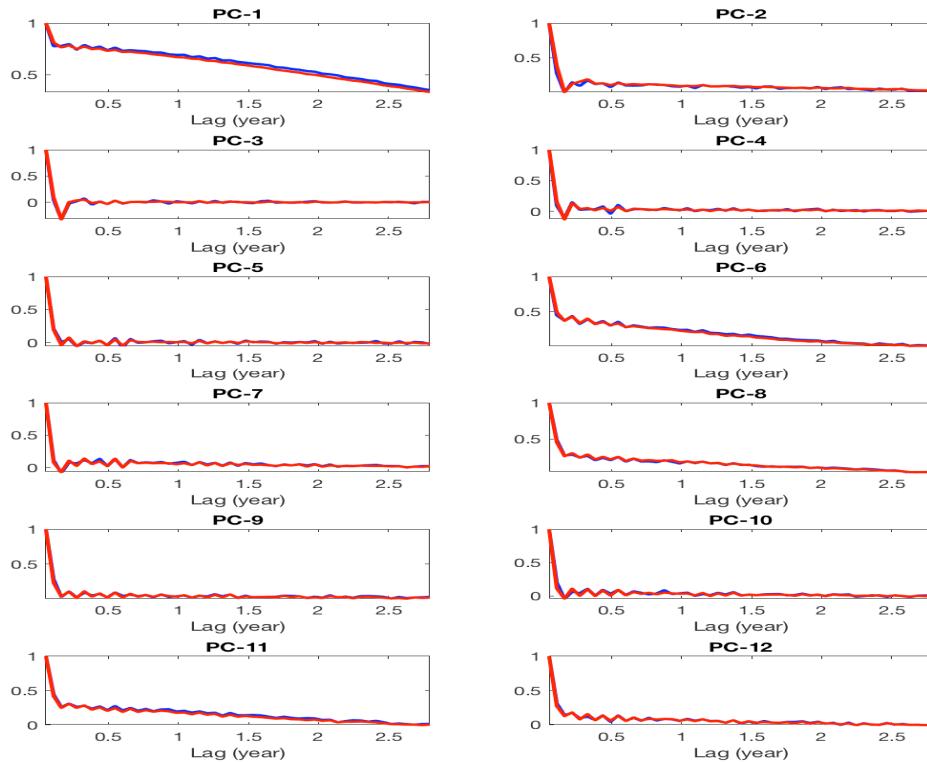
**Figure A4.** Reduced RP resonances and their analytic approximation. Same as in Figure A2, but for the frequency  $f = 5.27$  cycle/yr $^{-1}$ .

## Appendix E. Modeling Skills in the EOF Space

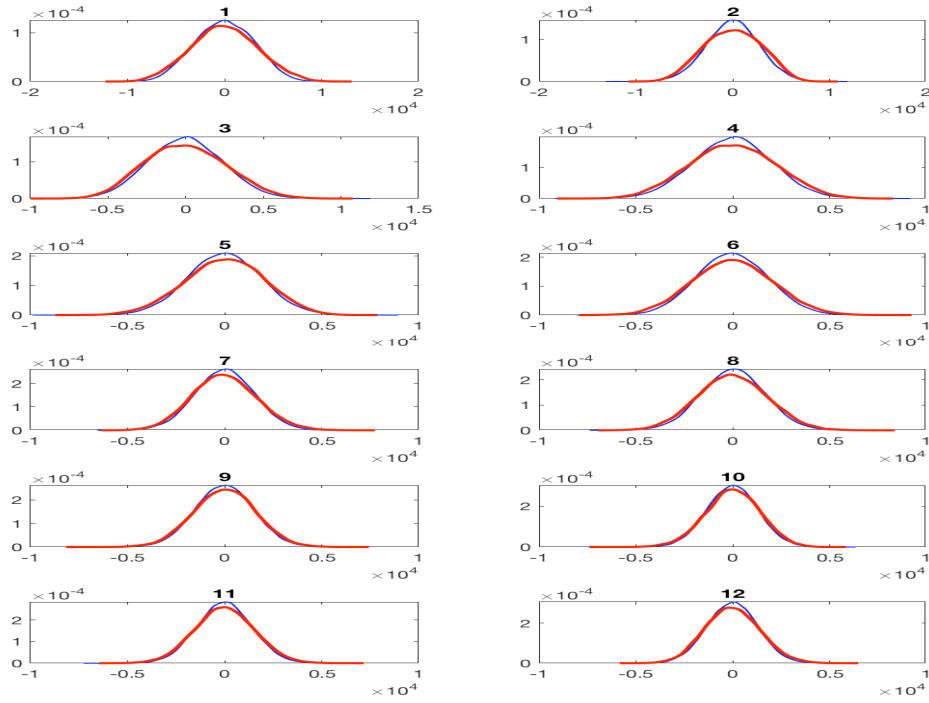
In this section, we communicate on the modeling skills in the space of EOFs. These skills are assessed by comparing the probability density functions (PDFs) of the twelve leading PCs of upper-layer stream function anomalies in Figures A5 and A7, as well as their autocorrelation functions (ACFs); see Figures A6 and A8.



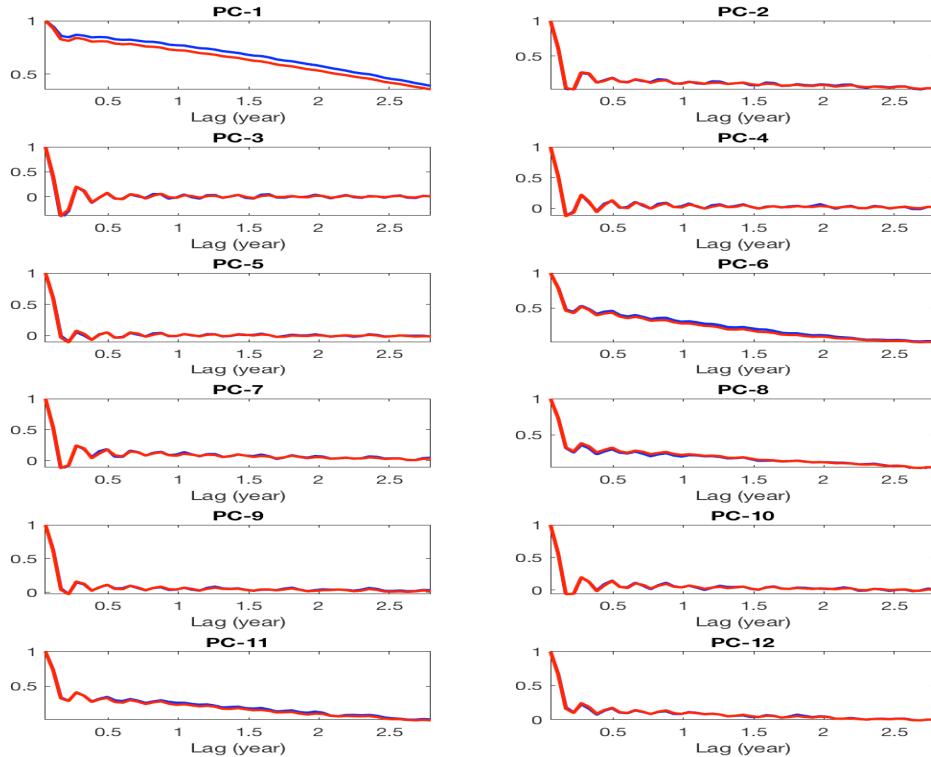
**Figure A5.** Same format as in Figure A7, but the comparison of harmonic reconstruction (red) and DAH-MSLM stochastic simulation (blue) in a full frequency band  $f < 9.1 \text{ } y^{-1}$ .



**Figure A6.** Same format as in Figure A8, but the comparison of harmonic reconstruction (red) and DAH-MSLM stochastic simulation (blue) in a full frequency band  $f < 9.1 \text{ } y^{-1}$ .



**Figure A7.** Probability density function (PDF) of the twelve leading PCs of upper-layer stream function anomalies, DAH-filtered in a frequency band  $f < 5.27 y^{-1}$ : Red, harmonic reconstruction of QG data (HRC; cf. (23)); blue, DAH-MSLM stochastic simulation.



**Figure A8.** Autocorrelation function (ACF) of the twelve leading PCs of upper-layer stream function anomalies, DAH-filtered in a frequency band  $f < 5.27 y^{-1}$ : Red, harmonic reconstruction of QG data (HRC; cf. (23)); blue, DAH-MSLM stochastic simulation.

## References

1. Gent, P.; McWilliams, J. Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.* **1990**, *20*, 150–155.
2. Eden, C. A closure for meso-scale eddy fluxes based on linear instability theory. *Ocean Model.* **2011**, *39*, 362–369.
3. Leith, C.E. Stochastic backscatter in a subgrid-scale model: Plane shear mixing layer. *Phys. Fluids A Fluid Dyn.* **1990**, *2*, 297–299.
4. Berloff, P.S.; McWilliams, J.C. Material transport in oceanic gyres. Part III: Randomized stochastic models. *J. Phys. Oceanogr.* **2003**, *33*, 1416–1445.
5. Berloff, P.S. Random-forcing model of the mesoscale oceanic eddies. *J. Fluid Mech.* **2005**, *529*, 71–95.
6. Porta Mana, P.; Zanna, L. Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Model.* **2014**, *79*, 1–20.
7. Jansen, M.; Held, I. Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Model.* **2014**, *80*, 36–48.
8. Berloff, P. Dynamically consistent parameterization of mesoscale eddies. Part I: Simple model. *Ocean Model.* **2015**, *87*, 1–19.
9. Chorin, A.J.; Hald, O.H.; Kupferman, R. Optimal prediction with memory. *Physica D* **2002**, *166*, 239–257.
10. Givon, D.; Kupferman, R.; Stuart, A. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **2004**, *17*, R55–R127.
11. Chorin, A.J.; Hald, O.H.; Kupferman, R. Prediction from partial data, renormalization, and averaging. *J. Sci. Comput.* **2006**, *28*, 245–261.
12. Hald, O.H.; Stinis, P. Optimal prediction and the rate of decay for solutions of the Euler equations in two and three dimensions. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 6527–6532.
13. Wouters, J.; Lucarini, V. Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach. *J. Stat. Phys.* **2013**, *151*, 850–860.
14. Temam, R. *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, 2nd ed.; Springer: New York, NY, USA, 1997; p. 648.
15. Ma, T.; Wang, S. *Bifurcation Theory and Applications*; World Scientific: Singapore, 2005; Volume 53.
16. Ma, T.; Wang, S. *Phase Transition Dynamics*; Springer: Berlin/Heidelberg, Germany, 2016.
17. Chekroun, M.; Liu, H.; McWilliams, J. The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories. *Comput. Fluids* **2017**, *151*, 3–22.
18. Chekroun, M.D.; Liu, H.; Wang, S. *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I*; Springer Briefs in Mathematics; Springer: New York, NY, USA, 2015.
19. Chekroun, M.D.; Liu, H.; Wang, S. *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*; Springer Briefs in Mathematics; Springer: New York, NY, USA, 2015.
20. Jolliffe, I. *Principal Component Analysis*; Wiley: Hoboken, NJ, USA, 2002.
21. Preisendorfer, R.W. *Principal Component Analysis in Meteorology and Oceanography*; Elsevier: New York, NY, USA, 1988; p. 425.
22. Tipping, M.E.; Bishop, C.M. Probabilistic principal component analysis. *J. R. Stat. Soc. B* **1999**, *61*, 611–622.
23. Schölkopf, B.; Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **1998**, *10*, 1299–1319.
24. Mukhin, D.; Gavrilov, A.; Feigin, A.; Loskutov, E.; Kurths, J. Principal nonlinear dynamical modes of climate variability. *Sci. Rep.* **2015**, *5*, 15510.
25. Ghil, M.; Allen, M.R.; Dettinger, M.D.; Ide, K.; Kondrashov, D.; Mann, M.E.; Robertson, A.W.; Saunders, A.; Tian, Y.; Varadi, F.; et al. Advanced spectral methods for climatic time series. *Rev. Geophys.* **2002**, *40*, doi:10.1029/2000RG000092.
26. Coifman, R.R.; Lafon, S. Diffusion maps. *Appl. Comput. Harmon. Anal.* **2006**, *21*, 5–30.
27. Coifman, R.R.; Kevrekidis, I.G.; Lafon, S.; Maggioni, M.; Nadler, B. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Model. Simul.* **2008**, *7*, 842–864.
28. Giannakis, D.; Majda, A.J. Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2222–2227.

29. Froyland, G.; Gottwald, G.A.; Hammerlindl, A. A computational method to extract macroscopic variables and their dynamics in multiscale systems. *SIAM J. Appl. Dyn. Syst.* **2014**, *13*, 1816–1846.

30. Schmid, P.J. Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **2010**, *656*, 5–28.

31. Budišić, M.; Mohr, R.; Mezić, I. Applied Koopmanism. *Chaos* **2012**, *22*, 047510.

32. Tu, J.H.; Rowley, C.W.; Luchtenburg, D.M.; Brunton, S.L.; Kutz, J.N. On dynamic mode decomposition: Theory and applications. *J. Comput. Dyn.* **2014**, *1*, 391–421.

33. Williams, M.O.; Kevrekidis, I.G.; Rowley, C.W. A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlinear Sci.* **2015**, *25*, 1307–1346.

34. Klus, S.; Koltai, P.; Schütte, C. On the numerical approximation of the Perron-Frobenius and Koopman operator. *J. Comput. Dyn.* **2016**, *3*, 51–79.

35. Klus, S.; Nüske, F.; Koltai, P.; Wu, H.; Kevrekidis, I.; Schütte, C.; Noé, F. Data-driven model reduction and transfer operator approximation. *J. Nonlinear Sci.* **2017**, *1*–26, doi:10.1007/s00332-017-9437-7.

36. Chorin, A.J.; Lu, F. Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9804–9809.

37. Lu, F.; Lin, K.; Chorin, A. Comparison of continuous and discrete-time data-based modeling for hypoelliptic systems. *Commun. Appl. Math. Comput. Sci.* **2016**, *11*, 187–216.

38. Lu, F.; Lin, K.; Chorin, A. Data-based stochastic model reduction for the Kuramoto–Sivashinsky equation. *Phys. D Nonlinear Phenom.* **2017**, *340*, 46–57.

39. Khashei, M.; Bijari, M. An artificial neural network  $(p, d, q)$ -model for time series forecasting. *Expert Syst. Appl.* **2010**, *37*, 479–489.

40. Hsu, K.L.; Gupta, H.V.; Sorooshian, S. Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.* **1995**, *31*, 2517–2530.

41. Mukhin, D.; Kondrashov, D.; Loskutov, E.; Gavrilov, A.; Feigin, A.; Ghil, M. Predicting critical transitions in ENSO models. Part II: Spatially dependent models. *J. Clim.* **2015**, *28*, 1962–1976.

42. Wang, Z.; Akhtar, I.; Borggaard, J.; Iliescu, T. Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison. *Comput. Methods Appl. Mech. Eng.* **2012**, *237*, 10–26.

43. Iliescu, T.; Wang, Z. Variational multiscale proper orthogonal decomposition: Navier–stokes equations. *Numer. Methods Partial Differ. Equ.* **2014**, *30*, 641–663.

44. Kwasniok, F. Empirical low-order models of barotropic flow. *J. Atmos. Sci.* **2004**, *61*, 235–245.

45. Sapsis, T.P.; Dijkstra, H.A. Interaction of additive noise and nonlinear dynamics in the double-gyre wind-driven ocean circulation. *J. Phys. Oceanogr.* **2013**, *43*, 366–381.

46. Penland, C. Random forcing and forecasting using principal oscillation pattern analysis. *Month. Wea. Rev.* **1989**, *117*, 2165–2185.

47. Penland, C.; Sardeshmukh, P.D. The optimal growth of tropical sea-surface temperature anomalies. *J. Clim.* **1995**, *8*, 1999–2024.

48. Kravtsov, S.; Kondrashov, D.; Ghil, M. Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J. Clim.* **2005**, *18*, 4404–4424.

49. Kravtsov, S.; Kondrashov, D.; Ghil, M. Empirical model reduction and the modeling hierarchy in climate dynamics. In *Stochastic Physics and Climate Modeling*; Palmer, T.N., Williams, P., Eds.; Cambridge Univ. Press: Cambridge, UK, 2009; pp. 35–72.

50. Kondrashov, D.; Kravtsov, S.; Robertson, A.W.; Ghil, M. A hierarchy of data-based ENSO models. *J. Clim.* **2005**, *18*, 4425–4444.

51. Kondrashov, D.; Chekroun, M.D.; Robertson, A.W.; Ghil, M. Low-order stochastic model and “past-noise forecasting” of the Madden-Julian oscillation. *Geophys. Res. Lett.* **2013**, *40*, 5305–5310.

52. Majda, A.J.; Harlim, J. Physics constrained nonlinear regression models for time series. *Nonlinearity* **2012**, *26*, 201.

53. Strounine, K.; Kravtsov, S.; Kondrashov, D.; Ghil, M. Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance. *Physica D* **2010**, *239*, 145–166.

54. Kondrashov, D.; Chekroun, M.D.; Ghil, M. Data-driven non-Markovian closure models. *Physica D* **2015**, *297*, 33–55.

55. Chekroun, M.D.; Kondrashov, D. Data-adaptive harmonic spectra and multilayer Stuart-Landau models. *Chaos* **2017**, *27*, doi:10.1063/1.4989400.

56. Chekroun, M.D.; Neelin, J.D.; Kondrashov, D.; McWilliams, J.C.; Ghil, M. Rough parameter dependence in climate models: The role of Ruelle-Pollicott resonances. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1684–1690.

57. Kondrashov, D.; Chekroun, M.D.; Yuan, X.; Ghil, M. Data-adaptive harmonic decomposition and stochastic modeling of Arctic sea ice. In *Advances in Nonlinear Geosciences*; Tsonis, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2018.

58. Kondrashov, D.; Chekroun, M.D.; Ghil, M. Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent. *Dyn. Stat. Clim. Syst. Interdiscip. J.* **2018**, in press.

59. Kondrashov, D.; Chekroun, M.D. Data-adaptive harmonic analysis and modeling of solar wind-magnetosphere coupling. *J. Atmos. Solar Terr. Phys.* **2018**, in press.

60. Berloff, P.; Hogg, A.M.C.; Dewar, W. The turbulent oscillator: A mechanism of low-frequency variability of the wind-driven ocean gyres. *J. Phys. Oceanogr.* **2007**, *37*, 2363–2386.

61. Shevchenko, I.V.; Berloff, P.; Guerrero-Lopez, D.; Roman, J.E. On low-frequency variability of the midlatitude ocean gyres. *J. Fluid Mech.* **2016**, *795*, 423–442.

62. Jiang, S.; Jin, F.F.; Ghil, M. Multiple equilibria, periodic, and aperiodic solutions in a wind-driven, double-gyre, shallow-water model. *J. Phys. Oceanogr.* **1995**, *25*, 764–786.

63. Nadiga, B.T.; Luce, B.P. Global bifurcation of Shilnikov type in a double-gyre ocean model. *J. Phys. Oceanogr.* **2001**, *31*, 2669–2690.

64. Simonnet, E.; Dijkstra, H.A. Spontaneous generation of low-frequency modes of variability in the wind-driven ocean circulation. *J. Phys. Oceanogr.* **2002**, *32*, 1747–1762.

65. Simonnet, E.; Ghil, M.; Ide, K.; Temam, R.; Wang, S. Low-frequency variability in shallow-water models of the wind-driven ocean circulation. Part II: Time-dependent solutions. *J. Phys. Oceanogr.* **2003**, *33*, 729–752.

66. Dijkstra, H.A.; Ghil, M. Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys.* **2005**, *43*, doi:10.1029/2002RG000122.

67. Dijkstra, H. A normal mode perspective of intrinsic ocean-climate variability. *Ann. Rev. Fluid Mech.* **2016**, *48*, 341–363.

68. Ghil, M. The wind-driven ocean circulation: Applying dynamical systems theory to a climate problem. *Discret. Contin. Dyn. Syst. A* **2017**, *37*, 189–228.

69. Kondrashov, D.; Berloff, P. Stochastic modeling of decadal variability in ocean gyres. *Geophys. Res. Lett.* **2015**, *42*, 1543–1553.

70. Gaspard, P. Trace formula for noisy flows. *J. Stat. Phys.* **2002**, *106*, 57–96.

71. Chekroun, M.; Tantet, A.; Dijkstra, H.; Neelin, J.D. Mixing spectrum in reduced phase spaces of stochastic differential equations. Part I: Theory. *arXiv* **2017**, arXiv:1705.07573.

72. Landau, L.D.; Lifshits, E.M. *Fluid Mechanics*; Pergamon Press: Oxford, UK, 1959.

73. Ruelle, D.; Takens, F. On the nature of turbulence. *Commun. Math. Phys.* **1971**, *20*, 167–192.

74. Berloff, P.; Dewar, W.; Kravtsov, S.; McWilliams, J. Ocean eddy dynamics in a coupled ocean–atmosphere model. *J. Phys. Oceanogr.* **2007**, *37*, 1103–1121.

75. Berloff, P. Dynamically Consistent Parameterization of Mesoscale Eddies. Part II: Eddy Fluxes and and diffusivity from transient impulses. *Fluids* **2016**, *1*, 22.

76. Engel, K.J.; Nagel, R. *A Short Course on Operator Semigroups*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2006.

77. Zakharova, A.; Loos, S.; Siebert, J.; Gjurchinovski, A.; Claussen, J.C.; Schöll, E. Controlling chimera patterns in networks: interplay of structure, noise, and delay in Control of Self-Organizing Nonlinear Systems. In *Control of Self-Organizing Nonlinear Systems*; Schöll, E., Klapp, S.H.L., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 35–72.

78. Selivanov, A.A.; Lehnert, J.; Dahms, T.; Hövel, P.; Fradkov, A.L.; Schöll, E. Adaptive synchronization in delay-coupled networks of Stuart-Landau oscillators. *Phys. Rev. E* **2012**, *85*, 016201.

79. Cerrai, S. *Second-Order PDE's in Finite and Infinite Dimension: A Probabilistic Approach*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2001; Volume 1762.

80. Flandoli, F.; Gubinelli, M.; Priola, E. Flow of diffeomorphisms for SDEs with unbounded Holder continuous drift. *Bull. Sci. Math.* **2010**, *134*, 405–422.

81. Ruelle, D. Locating resonances for Axiom A dynamical systems. *J. Stat. Phys.* **1986**, *44*, 281–292.

82. Pollicott, M. Meromorphic extensions of generalised zeta functions. *Invent. Math.* **1986**, *85*, 147–164.

83. Butterley, O.; Liverani, C. Smooth Anosov flows: Correlation spectra and stability. *J. Mod. Dyn.* **2007**, *1*, 301–322.
84. Dyatlov, S.; Zworski, M. Stochastic stability of Pollicott–Ruelle resonances. *Nonlinearity* **2015**, *28*, 3511.
85. Froyland, G. Computer-assisted bounds for the rate of decay of correlations. *Commun. Math. Phys.* **1997**, *189*, 237–257.
86. Crommelin, D.; Vanden-Eijnden, E. Reconstruction of diffusions using spectral data from time series. *Commun. Math. Sci.* **2006**, *4*, 651–668.
87. Schütte, C.; Fischer, A.; Huisings, W.; Deuflhard, P. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.* **1999**, *151*, 146–168.
88. Crommelin, D.; Vanden-Eijnden, E. Fitting time series by continuous-time Markov chains: A quadratic programming approach. *J. Comput. Phys.* **2006**, *217*, 782–805.
89. Tantet, A.; Chekroun, M.; Dijkstra, H.; Neelin, J.D. Mixing Spectrum in Reduced Phase Spaces of Stochastic Differential Equations. Part II: Stochastic Hopf Bifurcation. *arXiv* **2017**, arXiv:1705.07573.
90. Chekroun, M.D.; Simonnet, E.; Ghil, M. Stochastic climate dynamics: Random attractors and time-dependent invariant measures. *Physica D* **2011**, *240*, 1685–1700.
91. Bagheri, S. Effects of weak noise on oscillating flows: Linking quality factor, Floquet modes, and Koopman spectrum. *Phys. Fluids* **2014**, *26*, 094104.
92. Gaspard, P.; Nicolis, G.; Provata, A.; Tasaki, S. Spectral signature of the pitchfork bifurcation: Liouville equation approach. *Phys. Rev. E* **1995**, *51*, 74.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).