

Completing partially known space-time flow statistics: A resolvent-based approach

By A. Towne

1. Introduction

Practical limitations in both experiments and simulations can lead to partial knowledge of flow statistics. For example, an array of probes in an experiment provides information at a limited number of spatial locations and for a single flow quantity, e.g., velocity from hot-wires or pressure from microphones. Similarly, particle image velocimetry might provide velocity data, but not thermodynamic quantities, in a limited field of view. In simulations, one may wish to know flow statistics in a region that is not adequately resolved by the computational grid, such as unresolved near-wall regions or locations outside of the main computational domain. This report develops a method for estimating unknown space-time flow statistics from a limited set of known values of these statistics.

The method is based on the resolvent formalism of McKeon & Sharma (2010). The resolvent operator is derived from the Navier-Stokes equations linearized about the turbulent mean flow and constitutes a transfer function in the frequency domain between terms that are nonlinear and linear with respect to fluctuations to the mean. The basic idea of the method developed in this report is to use a limited set of known space-time flow statistics to infer the statistics of the nonlinear terms, which in turn imply values for the unknown flow statistics.

This approach builds on the work of Beneddine *et al.* (2016) and Zare *et al.* (2017). These authors developed methods for completing unknown one-point temporal and two-point spatial statistics, respectively. These are both subsets of two-point space-time correlations, so our method can be thought of as a generalization or union of these previous approaches. This is an important step since two-point space-time statistics contain more information about the flow and constitute a starting point for identifying and modeling coherent structures (Towne *et al.* 2017b).

Beneddine *et al.* (2016) proposed a method for estimating unknown power spectral densities (PSDs) using knowledge of the mean flow field and power spectra at a few locations. This is accomplished using a least-squares fit at each frequency between the known power spectra and the leading singular response mode obtained from the resolvent operator. This strategy explicitly assumes that the spectral content at frequencies of interest is dominated by the leading resolvent mode, and the method performs well when the matching points are located in regions where this hypothesis is valid. Specifically, excellent PSD estimates were obtained for the flow over a backward-facing step (Beneddine *et al.* 2016) and a transitional jet (Beneddine *et al.* 2017). As pointed out by Beneddine *et al.* (2017), this assumption will typically not hold in fully turbulent flows. For example, Schmidt *et al.* (2017a,b) showed that the leading resolvent modes in turbulent jets account for a modest fraction of the total flow energy at each frequency (at most $\sim 20\%$). In this case, attempting to fit the leading resolvent mode to the PSD will lead to large errors. Our method builds on this rank-1 approach of Beneddine *et al.*

(2016) by relaxing this a priori assumption and allowing the known data to self-select the relevant portions of the resolvent operator.

Zare *et al.* (2017) developed a method that uses arbitrary known entries in the spatial covariance tensor to estimate the remaining unknown entries. Their approach is also based on linearized flow equations and entails solving a convex optimization problem that determines a matrix controlling the structure and statistics of the associated nonlinear forcing terms. The optimization problem is subject to two constraints on the estimated covariance tensor: it must reproduce the known entries and obey a Lyapunov equation that relates the forcing and flow statistics. The constrained optimization problem is difficult to solve and requires a customized algorithm (Zare *et al.* 2015, 2017). Our approach follows these authors' underlying strategy of using the known data to infer the statistics of the nonlinear forcing terms. However, our frequency-domain formulation is algorithmically simple, requiring only basic linear algebra manipulations and avoiding Lyapunov equations and the need for external optimization routines.

A limitation of our approach compared to these other methods is that two-point space-time statistics, rather than one-point temporal or spatial statistics, are required as input. This could be problematic in some applications, but in many cases the same data used to compute the simpler statistics can be used to obtain the two-point space-time statistics as well. This is the case any time measurements at multiple locations are obtained simultaneously, e.g., by an array of probes or by sampling numerical data, and two-point space-time statistics can always be obtained using just two concurrent measurements.

The remainder of this report is organized as follows. The method is derived and described in Section 2 and demonstrated in Section 3 using a simple model problem given by the Ginzburg-Landau equation. Finally, Section 4 summarizes the report and discusses further improvements and applications of the method.

2. Method

2.1. Objective

Consider a state vector of flow variables $\mathbf{q}(\mathbf{x}, t)$ that describe a flow, e.g., velocities and thermodynamic variables. The independent variables \mathbf{x} and t represent the spatial dimensions of the problem and time, respectively. Now suppose that the two-point space-time statistics are known for a reduced set of variables

$$\mathbf{y} = \mathcal{C}\mathbf{q}, \quad (2.1)$$

where the linear operator $\mathcal{C}(\mathbf{x})$ selects any desired subset or linear combination of \mathbf{q} . The problem objective can now be precisely stated in terms of two-point space-time correlation tensors:

$$\text{given} \quad \mathbf{C}_{yy}(\mathbf{x}, \mathbf{x}', \tau) = E \{ \mathbf{y}(\mathbf{x}, t) \mathbf{y}^*(\mathbf{x}', t + \tau) \}, \quad (2.2a)$$

$$\text{estimate} \quad \mathbf{C}_{qq}(\mathbf{x}, \mathbf{x}', \tau) = E \{ \mathbf{q}(\mathbf{x}, t) \mathbf{q}^*(\mathbf{x}', t + \tau) \}. \quad (2.2b)$$

Here, $E \{ \cdot \}$ is the expectation operator over time and the asterisk superscript indicates a Hermitian transpose.

Using the relationship between space-time correlation tensors and the cross-spectral density (CSD) tensors

$$\mathbf{S}(\mathbf{x}, \mathbf{x}', \omega) = \int_{-\infty}^{\infty} \mathbf{C}(\mathbf{x}, \mathbf{x}', \tau) e^{i\omega\tau} d\tau, \quad (2.3)$$

this objective can be equivalently stated in the frequency domain:

$$\text{given} \quad \mathbf{S}_{yy}(\mathbf{x}, \mathbf{x}', \omega) = E \{ \hat{\mathbf{y}}(\mathbf{x}, \omega) \hat{\mathbf{y}}^*(\mathbf{x}', \omega) \}, \quad (2.4a)$$

$$\text{estimate} \quad \mathbf{S}_{qq}(\mathbf{x}, \mathbf{x}', \omega) = E \{ \hat{\mathbf{q}}(\mathbf{x}, \omega) \hat{\mathbf{q}}^*(\mathbf{x}', \omega) \}, \quad (2.4b)$$

where $\hat{\mathbf{y}}(\mathbf{x}, \omega)$ and $\hat{\mathbf{q}}(\mathbf{x}, \omega)$ are the Fourier transforms of \mathbf{y} and \mathbf{q} , respectively.

2.2. Approach

Our approach to this problem relies on the resolvent operator obtained from the linearized flow equations and its connection with the remaining nonlinear terms (McKeon & Sharma 2010). Begin with nonlinear flow equations of the form

$$\frac{\partial \mathbf{q}}{\partial t} = \mathcal{F}(\mathbf{q}). \quad (2.5)$$

The compressible Navier-Stokes equations are naturally written in this form, and the incompressible equations can be written this way by projecting into a divergence-free basis to eliminate the continuity equation. Additional transport equations can also be included.

Applying the Reynolds decomposition

$$\mathbf{q}(\mathbf{x}, t) = \bar{\mathbf{q}}(\mathbf{x}) + \mathbf{q}'(\mathbf{x}, t) \quad (2.6)$$

to Eq. (2.5) and isolating the terms that are linear in \mathbf{q}' yields an equation of the form

$$\frac{\partial \mathbf{q}'}{\partial t} - \mathcal{A}(\bar{\mathbf{q}}) \mathbf{q}' = \mathbf{f}(\bar{\mathbf{q}}, \mathbf{q}'), \quad (2.7)$$

where

$$\mathcal{A}(\bar{\mathbf{q}}) = \frac{\partial \mathcal{F}}{\partial \mathbf{q}}(\bar{\mathbf{q}}) \quad (2.8)$$

is the linearized Navier-Stokes operator and \mathbf{f} contains the remaining nonlinear terms. Similarly, Eq. (2.1) becomes

$$\mathbf{y}' = \mathcal{C} \mathbf{q}'. \quad (2.9)$$

In the frequency domain, Eqs. (2.7) and (2.9) can be manipulated to give

$$\hat{\mathbf{y}} = \mathcal{R}_y \hat{\mathbf{f}}, \quad (2.10a)$$

$$\hat{\mathbf{q}} = \mathcal{R}_q \hat{\mathbf{f}}, \quad (2.10b)$$

where

$$\mathcal{R}_y(\mathbf{x}, \omega) = \mathcal{C} (i\omega I - \mathcal{A})^{-1}, \quad (2.11a)$$

$$\mathcal{R}_q(\mathbf{x}, \omega) = (i\omega I - \mathcal{A})^{-1} \quad (2.11b)$$

are resolvent operators associated with $\hat{\mathbf{y}}$ and $\hat{\mathbf{q}}$, respectively, and I is the identity matrix of appropriate dimension.

Using Eqs. (2.4) and (2.10), the CSD tensors can be written in terms of these resolvent operators as

$$\mathbf{S}_{yy} = \mathcal{R}_y \mathbf{S}_{ff} \mathcal{R}_y^*, \quad (2.12a)$$

$$\mathbf{S}_{qq} = \mathcal{R}_q \mathbf{S}_{ff} \mathcal{R}_q^*, \quad (2.12b)$$

where $\mathbf{S}_{ff}(\mathbf{x}, \mathbf{x}', \omega) = E \{ \hat{\mathbf{f}}(\mathbf{x}, \omega) \hat{\mathbf{f}}^*(\mathbf{x}', \omega) \}$ is the CSD tensor of the nonlinear term \mathbf{f} (Semeraro *et al.* 2016; Towne *et al.* 2016, 2017b).

To obtain an approximation of the desired statistics \mathbf{S}_{qq} , we use the known statistics \mathbf{S}_{yy} to estimate \mathbf{S}_{ff} . The salient question then becomes: how much can we learn about \mathbf{S}_{ff} from \mathbf{S}_{yy} ? An answer is provided by examining the singular value decomposition (SVD)

$$\mathcal{R}_y = \mathbf{U}_y \mathbf{\Sigma}_y \mathbf{V}_y^* \quad (2.13a)$$

$$= \mathbf{U}_y \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}^*. \quad (2.13b)$$

The columns of the matrices \mathbf{V}_y and \mathbf{U}_y correspond to input and output modes that form orthonormal bases for $\hat{\mathbf{f}}$ and $\hat{\mathbf{y}}$, respectively. The rectangular matrix $\mathbf{\Sigma}_y$ determines the gain of each of the input modes to the output. Since the rank of \mathcal{R}_y can be no greater than the number of entries in \mathbf{y} , i.e., the number of locations/quantities for which the statistics are known, many of the input modes have no impact on the output. Accordingly, the SVD can be written in the form of Eq. (2.13b), where the diagonal $\mathbf{\Sigma}_1$ contains the non-zero singular values and the blocks \mathbf{V}_1 and \mathbf{V}_2 contain input modes that have non-zero and zero gain, respectively. It is important to note that these resolvent modes are different from those usually studied, which are given by the SVD of \mathcal{R}_q .

The distinction between input modes that do or do not impact the output can be used to isolate the part of \mathbf{S}_{ff} that can be deduced from knowledge of \mathbf{S}_{yy} . Since \mathbf{V}_y provides a complete basis for $\hat{\mathbf{f}}$, \mathbf{S}_{ff} can be expanded as

$$\mathbf{S}_{ff} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}^*, \quad (2.14)$$

where the matrices \mathbf{E}_{ij} represent correlation between expansion coefficients associated with each input mode (see Towne *et al.* 2017b). Inserting this expansion into Eq. (2.12a) and using Eq. (2.13b) to simplify the expression gives rise to the equation

$$\mathbf{S}_{yy} = \mathbf{U}_y \mathbf{\Sigma}_1 \mathbf{E}_{11} \mathbf{\Sigma}_1 \mathbf{U}_y^*. \quad (2.15)$$

This means that only the part of \mathbf{S}_{ff} associated with \mathbf{E}_{11} impacts the observed statistics \mathbf{S}_{yy} ; the remaining \mathbf{E}_{ij} terms have no impact and are thus unobservable from these known data. Consequently, \mathbf{E}_{11} contains all of the information about \mathbf{S}_{ff} that can be inferred from \mathbf{S}_{yy} . Using the orthonormality of \mathbf{U}_y , Eq. (2.15) gives

$$\mathbf{E}_{11} = \mathbf{\Sigma}_1^{-1} \mathbf{U}_y^* \mathbf{S}_{yy} \mathbf{U}_y \mathbf{\Sigma}_1^{-1}. \quad (2.16)$$

The remaining terms \mathbf{E}_{22} and $\mathbf{E}_{12} = \mathbf{E}_{21}^*$ (this equality is required to make \mathbf{S}_{ff} Hermitian) can be arbitrarily chosen without impacting \mathbf{S}_{yy} , but these terms will impact \mathbf{S}_{qq} and therefore must be modeled. The simplest choice, and the one used in the remainder of this report, is to set these unknown terms to zero, leading to the approximation

$$\mathbf{S}_{ff} \approx \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{E}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{bmatrix}^* = \mathbf{V}_1 \mathbf{E}_{11} \mathbf{V}_1^*. \quad (2.17)$$

Inserting this into Eq. (2.12b) gives the corresponding approximation of the desired flow statistics

$$\mathbf{S}_{qq} \approx \mathcal{R}_q \mathbf{V}_1 \mathbf{E}_{11} \mathbf{V}_1^* \mathcal{R}_q^*. \quad (2.18)$$

By construction, the known statistics used as input are exactly recovered, ensuring that the approximation converges in the limit of full knowledge of the flow statistics. Other approximations can be obtained by choosing the unknown \mathbf{E}_{ij} terms differently; a few possibilities are discussed in Section 4.

2.3. Remarks

The method has been described so far in terms of statistics of the nonlinear forcing terms. Alternatively, it is instructive to think of the method in terms of statistics of coefficients in a resolvent-mode expansion of $\mathbf{S}_{\mathbf{q}\mathbf{q}}$. The standard resolvent modes associated with the linearized flow equations are defined by the SVD $\mathcal{R}_{\mathbf{q}} = \mathbf{U}_{\mathbf{q}}\mathbf{\Sigma}_{\mathbf{q}}\mathbf{V}_{\mathbf{q}}^*$. Eq. (2.18) can then be written as

$$\mathbf{S}_{\mathbf{q}\mathbf{q}} \approx \mathbf{U}_{\mathbf{q}}\mathbf{\Sigma}_{\mathbf{q}}\mathbf{S}_{\beta\beta}\mathbf{\Sigma}_{\mathbf{q}}\mathbf{U}_{\mathbf{q}}^*, \quad (2.19)$$

where

$$\mathbf{S}_{\beta\beta} = \mathbf{V}_{\mathbf{q}}^*\mathbf{V}_1\mathbf{E}_{11}\mathbf{V}_1^*\mathbf{V}_{\mathbf{q}} \quad (2.20)$$

is the CSD of the expansion coefficients in a resolvent-mode expansion of $\hat{\mathbf{q}}$ (Towne *et al.* 2017b). In general, $\mathbf{S}_{\beta\beta}$ can project onto any of the resolvent output modes in $\mathbf{U}_{\mathbf{q}}$. Thus, the known statistics $\mathbf{S}_{\mathbf{y}\mathbf{y}}$, through their influence on \mathbf{E}_{11} , determine which resolvent modes participate in the estimate of $\mathbf{S}_{\mathbf{q}\mathbf{q}}$. This can be contrasted with the rank-1 model described earlier, in which only the leading resolvent mode is allowed to contribute.

3. Example: Ginzburg-Landau equation

In this section, the method described above is demonstrated using the Ginzburg-Landau equation, which has been used by several previous authors (e.g., Hunt & Crighton 1991; Bagheri *et al.* 2009; Chen & Rowley 2011; Towne *et al.* 2017b) as a convenient one-dimensional model that mimics key properties of the linearized Navier-Stokes operator for real flows. The linearized operator takes the form

$$\mathcal{A} = -\nu \frac{\partial}{\partial x} + \gamma \frac{\partial^2}{\partial x^2} + \mu(x). \quad (3.1)$$

Several variants of the function $\mu(x)$ have been used in the literature; here the quadratic form

$$\mu(x) = (\mu_0 - c_\mu^2) + \frac{\mu_2}{2}x^2 \quad (3.2)$$

is adopted (Hunt & Crighton 1991; Bagheri *et al.* 2009; Chen & Rowley 2011). All of the parameters in Eqs. (3.1) - (3.2) are set to the values used by Towne *et al.* (2017b). With these parameters, the leading singular value of $\mathcal{R}_{\mathbf{q}}$ at its peak frequency is 10 times larger than the second singular value, which is a typical value for real flows. Following Bagheri *et al.* (2009), the equations are discretized with a pseudo-spectral approach using $N = 220$ Hermite polynomials.

The discretized equations are stochastically excited in the time domain using forcing terms with prescribed statistics identical to those used by Towne *et al.* (2017b). In particular, the forcing is spatially correlated and qualitatively similar to the nonlinear terms in real flows, such as a turbulent jet (Towne *et al.* 2017a). Although these forcing statistics are prescribed in this model problem and therefore known, this knowledge is not made available to the estimation procedure. The equations are integrated using a fourth-order embedded Runge-Kutta method (Shampine & Reichelt 1997), and a total of 10000 snapshots of the solution are collected with spacing $\Delta t = 0.5$, leading to a Nyquist frequency of $\omega_{Nyquist} = 2\pi$. The CSD of the solution is computed from these data using Welch's (1967) method.

For the majority of the following analysis, \mathbf{y} is defined to correspond to data obtained from three probes located at $x = -10, 0$ and 10 . Other choices are considered in Section 3.4.

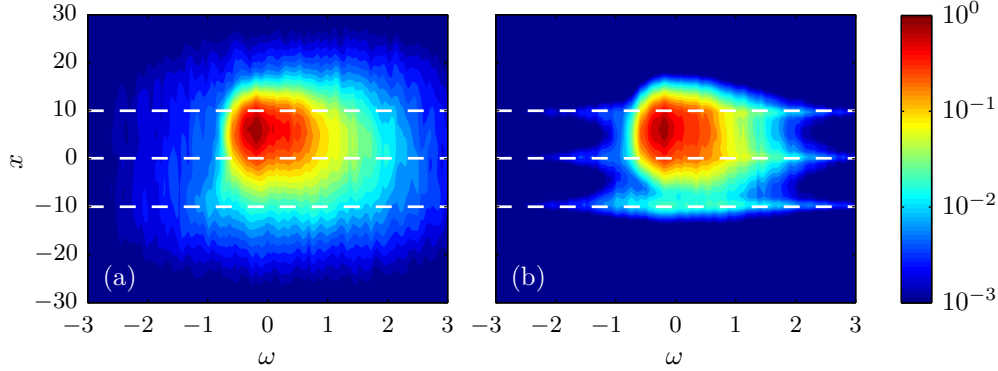


FIGURE 1. Power spectral density (PSD) as a function of ω and x for the Ginzburg-Landau model problem: (a) true PSD and (b) estimated PSD using three probes at the locations of the dashed lines.

3.1. Power spectra

The PSD is contained in the diagonal entries of \mathbf{S}_{qq} . The true power-spectral density for the Ginzburg-Landau model problem is shown as a function of ω and x in Figure 1(a). A single peak is observed at $\omega \approx -0.2$ and $x \approx 5$, and the amplitude remains above 1% of the peak over a range of about $-0.75 < \omega < 2$ and $-20 < x < 15$. The dashed lines show the x locations where the data is taken as known, and the estimation procedure will attempt to reconstruct the PSD elsewhere.

The approximation of the PSD obtained using these three probes is shown in Figure 1(b). By construction, the approximation is exact at the probe locations. The peak is well captured and the agreement is good in high-energy regions. In the lower-energy regions, the PSD is under-predicted away from the probe locations. This is a consequence of neglecting the undetermined portions of the forcing. It is likely that additional improvements could be obtained by modeling these undetermined portions of the forcing, as discussed in Section 4.

3.2. Cross-spectra

The CSD estimates are evaluated next. Figure 2 compares the real part of the true and modeled CSD at eight frequencies, which are listed in the caption. The contour levels are the same for the true and estimated data at each frequency and range from the minimum to maximum values of the true CSD. The circles indicate the locations where the CSD is known and the remaining values are to be estimated. The first six frequencies (panels a-l) fall within the high-energy region observed in Figure 1. In these cases, the estimates are accurate and track the length scales and shape of the CSD as a function of frequency. The final two frequencies (panels m-p) fall in low-energy regions. The basic trends in the length scales and shape are still captured, but the estimates are not quantitatively accurate.

3.3. Space-time correlations

The space-time correlation tensor \mathbf{C}_{qq} can be recovered from the CSD \mathbf{S}_{qq} using the inverse Fourier transform of Eq. (2.3). As an example, Figure 3 shows the true (solid lines) and estimated (dashed lines) correlations as a function of time lag τ for three spatial locations, $x = -5, 0$ and 5 . These locations correspond to a low-energy region,

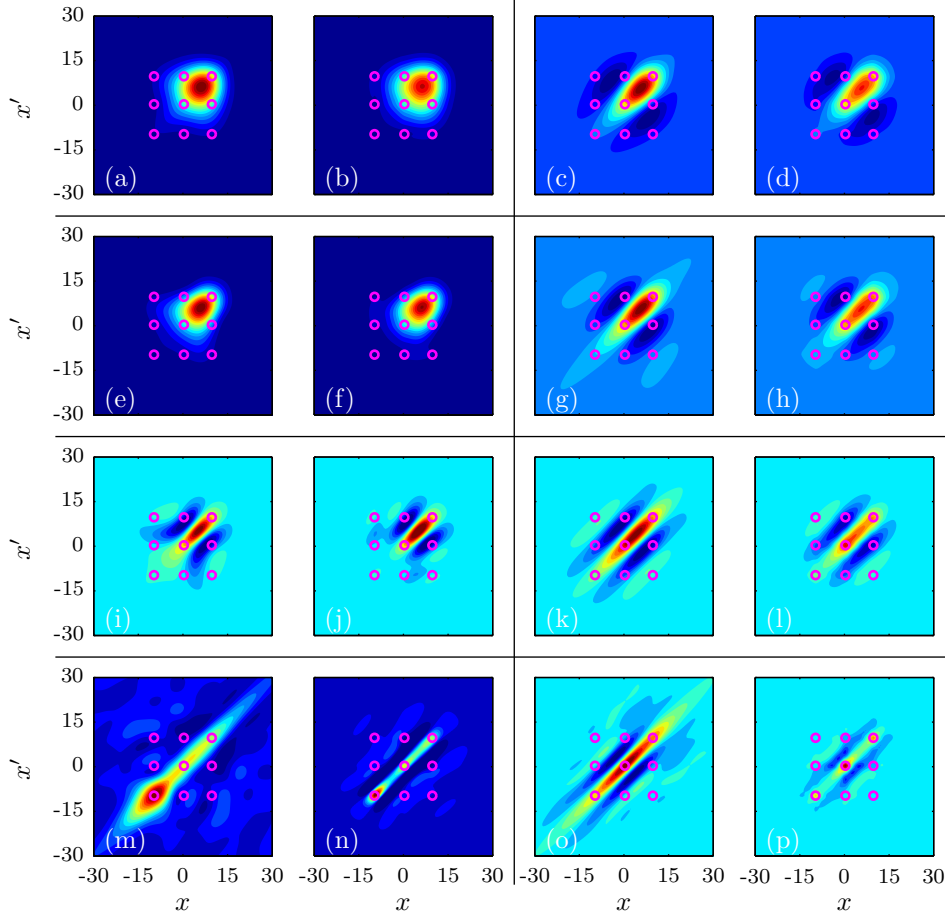


FIGURE 2. Cross-spectral density (CSD) for the frequencies (a-b) $\omega = 0$; (e-f) $\omega = -0.2$; (i-j) $\omega = -0.6$; (m-n) $\omega = -1$; (c-d) $\omega = 0.4$; (g-h) $\omega = 0.6$; (k-l) $\omega = 1$; (o-p) $\omega = 2$. In each case, the left-hand plot shows the true CSD and the right-hand plot shows the estimated values using three probes, which lead to known CSD values at the locations indicated by the small circles.

a probe position and the energy peak, respectively. Each curve has been scaled by the maximum value of the corresponding true correlation.

The one-point autocorrelation for each point is shown in Figure 3(a-c). The amplitude of the autocorrelation for the low-energy point at $x = -5$ (panel a) is significantly under-predicted, but the correlation length scale is well captured. The estimated autocorrelation at $x = 0$ (panel b) is exact since this point corresponds to one of the probe locations. The autocorrelation at $x = 5$ (panel c) is accurately estimated apart from a small under-prediction of the peak, which corresponds to an under-prediction of the variance.

The cross-correlations between these three points are shown in Figure 3(d-f). The estimates are quite good in all cases, including those involving the low-energy point (panels d and f) and two unknown points (panel e). It is interesting that the cross-correlations involving the low-energy point are more accurate than the autocorrelation

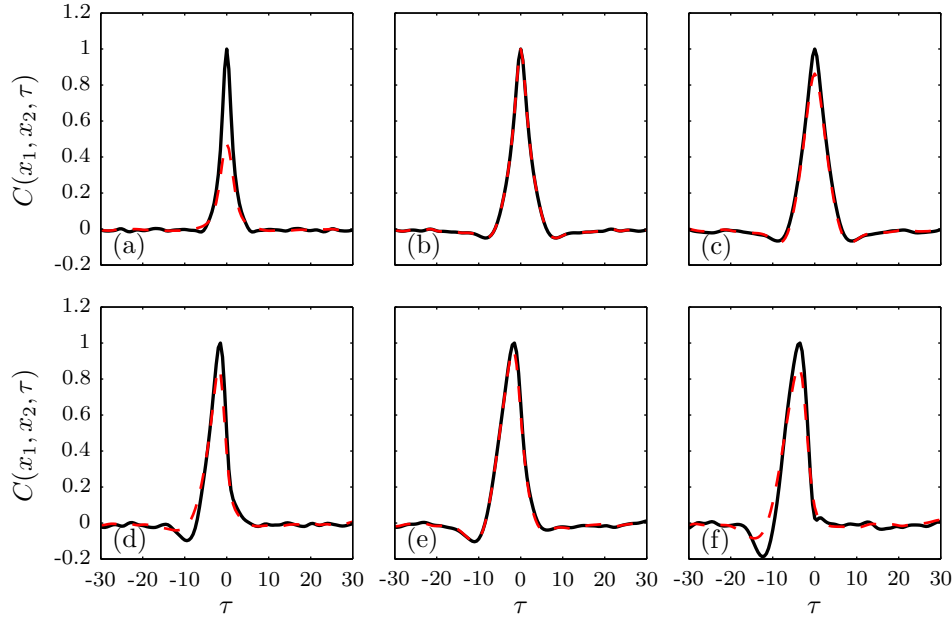


FIGURE 3. Cross-correlation as a function of time lag τ for (a) $x_1 = x_2 = -5$; (b) $x_1 = x_2 = 0$; (c) $x_1 = x_2 = 5$; (d) $x_1 = -5, x_2 = 0$; (e) $x_1 = 0, x_2 = 5$; (f) $x_1 = -5, x_2 = 5$. The solid lines show the true values, and the dashed lines show the estimates values using three probes at $x = -10, 0$ and 10 . Both the true and estimated curves in each plot have been scaled by the maximum value of the true correlation.

at this point. The agreement for the cross-correlation between the known and high-energy points (panel e) is almost perfect.

The spatial distribution of the true and estimated cross-correlation tensors at fixed values of the time lag τ is shown in Figure 4. The plotted time lag values range from $\tau = 0$ to 20 ; negative values need not be considered due to the symmetry

$$C_{qq}(x_1, x_2, -\tau) = C_{qq}^*(x_2, x_1, \tau). \quad (3.3)$$

The contour levels are the same for the true and estimated statistics at each value of τ (but different for each τ) and range from zero to the maximum values of the true correlations. Again, the circles indicate the locations where the correlations are known and the remaining values are to be estimated.

The spatial correlation tensor is obtained for $\tau = 0$ and is shown in Figure 4(a). As already observed in Figure 3, the amplitudes of the correlations at zero time lag are slightly underpredicted, but the spatial shape is well captured. As the time lag τ is increased, the estimates faithfully track the changing shape of the true correlations up to at least $\tau = 10$ (panels a-h). By this point, the magnitudes of the correlations are already very small (see Fig 3). Even so, the qualitative shape of the estimates is reasonable all the way up to $\tau = 20$.

3.4. Effect of probe location and comparisons with the rank-1 model

In this section, comparisons are made between the new method described in this report and the rank-1 method of Beneddine *et al.* (2016) that was discussed in Section 1. Par-

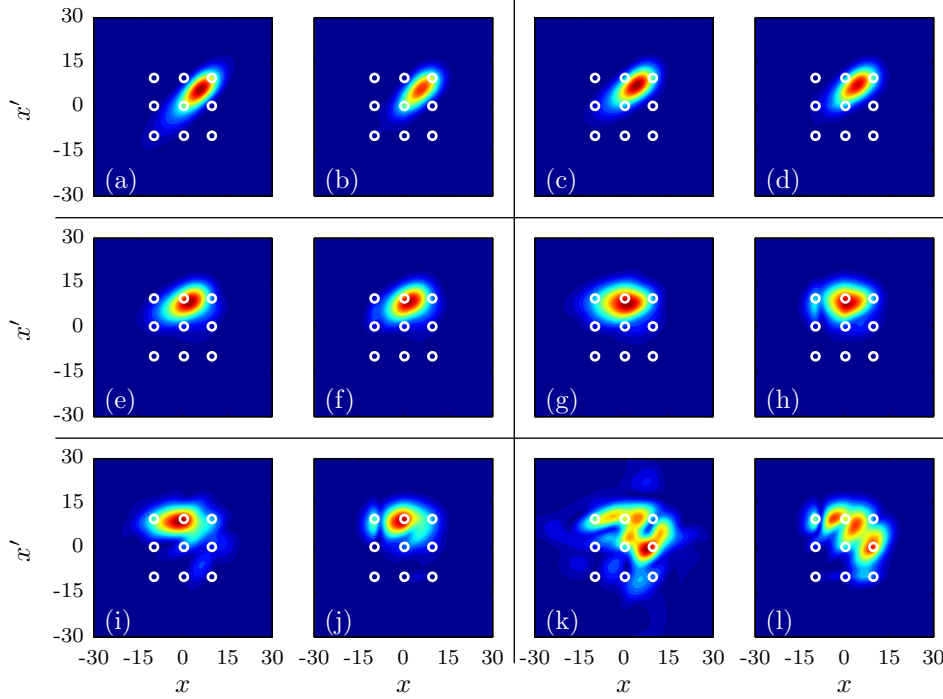


FIGURE 4. Cross-correlation as a function of x and x' for fixed time-lag values (a-b) $\tau = 0$; (c-d) $\tau = 2$; (e-f) $\tau = 5$; (g-h) $\tau = 10$; (i-j) $\tau = 15$; (k-l) $\tau = 20$. In each case, the left-hand plot shows the true correlations, and the right-hand plot shows the estimated values using three probes, which lead to known cross-correlations at the locations indicated by the small circles.

ticular attention is given to the impact of the probe location(s) on the accuracy of the estimates provided by these two methods.

The PSD, which is the target quantity of the rank-1 model, is considered first. Figure 5 compares the true PSD (top row) to the estimates from the rank-1 method (second row) and the new model (third row) for four different sets of probe locations (columns).

First, a single probe is placed at $x = 0$. In this case, the methods provide similar estimates, but the peak amplitude is slightly better predicted by the new method while higher frequencies are better captured by the rank-1 method. Adding two more probes at $x = \pm 10$ (second column) has little impact on the rank-1 estimate. In contrast, the new method is able to use this additional information to improve its estimate, particularly in regard to the shape of the moderately energetic region surrounding the peak and in the vicinity of the new probes.

Next, a single probe is placed at the low-energy location $x = -10$, well away from the peak. At this point, the underlying assumption of the rank-1 model – that the solution is dominated by the leading resolvent mode – is false. This is representative of the situation that will be encountered in real turbulent flows. Because of this, the rank-1 method leads to large over-predictions of the PSD. In contrast, the new method yields a moderate under-prediction of the PSD, which can be attributed to neglecting the unobservable portions of \mathbf{S}_{ff} .

In the final case (fourth column), the flow statistics are known in a continuous region,

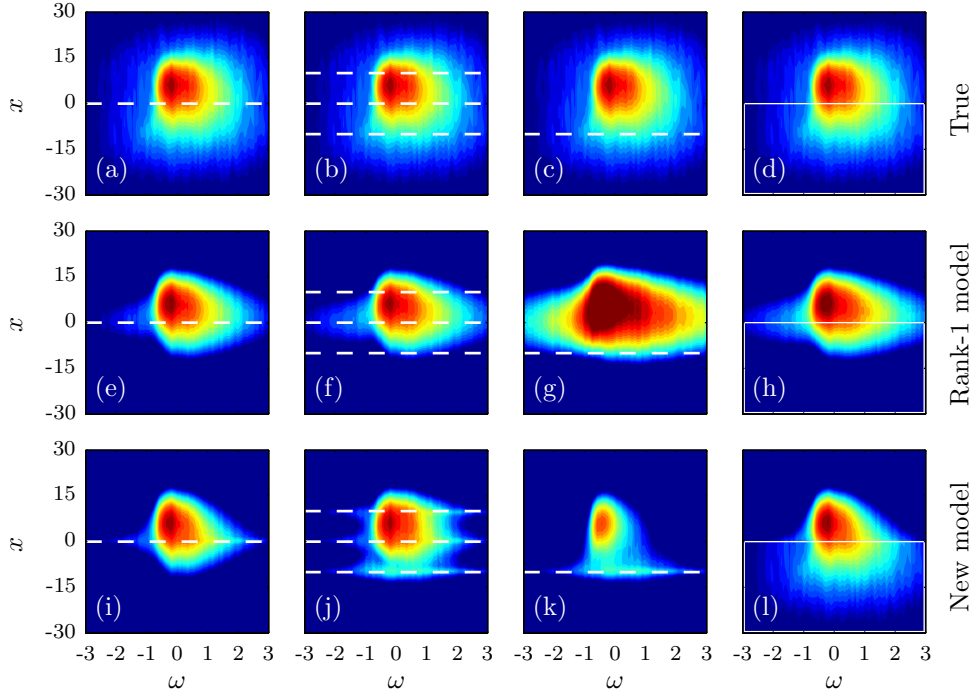


FIGURE 5. Power spectral density as a function of ω and x : (a-d) true values repeated for ease of comparison; (e-h) estimated values from the rank-1 model of Beneddine *et al.* (2016); (i-l) estimated values from the new model presented in this report. The estimates are based on the following probe locations: (e,i) $x = 0$; (f,j) $x = -10, 0$ and 10 ; (g,k) $x = -10$; (h,l) $x \leq 0$.

$x \leq 0$, rather than at an isolated set of points. These results can be compared to the first case in which statistics were known only at the boundary of this region, $x = 0$. Including the additional data for $x < 0$ leads to worse results for the rank-1 model (notice the significant over-prediction of the peak). This is an undesirable property; it means that poorly placed probes (where the rank-1 assumption is invalid) can obscure the information provided by well-placed probes (where the assumption is valid). More generally, this is a manifestation of the fact that the rank-1 method does not necessarily converge with increasing input information, even in the limit of complete knowledge of the flow statistics, in contrast to the new method. In the current example, it is clear that the additional information for $x < 0$ does not degrade the estimate of the new method as it did for the rank-1 model, but instead leads to small improvements in the estimated PSD.

Next, comparisons are made between the the CSD estimates provided by the two methods for the same four sets of probes. While the rank-1 method was not specifically designed to estimate cross-spectra, its form nevertheless implies values for these two-point statistics, and their accuracy is important if the method is to be used for time-domain modeling, as in Beneddine *et al.* (2017). Comparisons are made for the frequency of maximum gain, $\omega = -0.6$, where the rank-1 model is expected to be most appropriate.

The overall conclusions regarding the CSD estimates are similar to those just discussed for the PSD. The two methods yield equivalent results for the single probe at $x = 0$. The

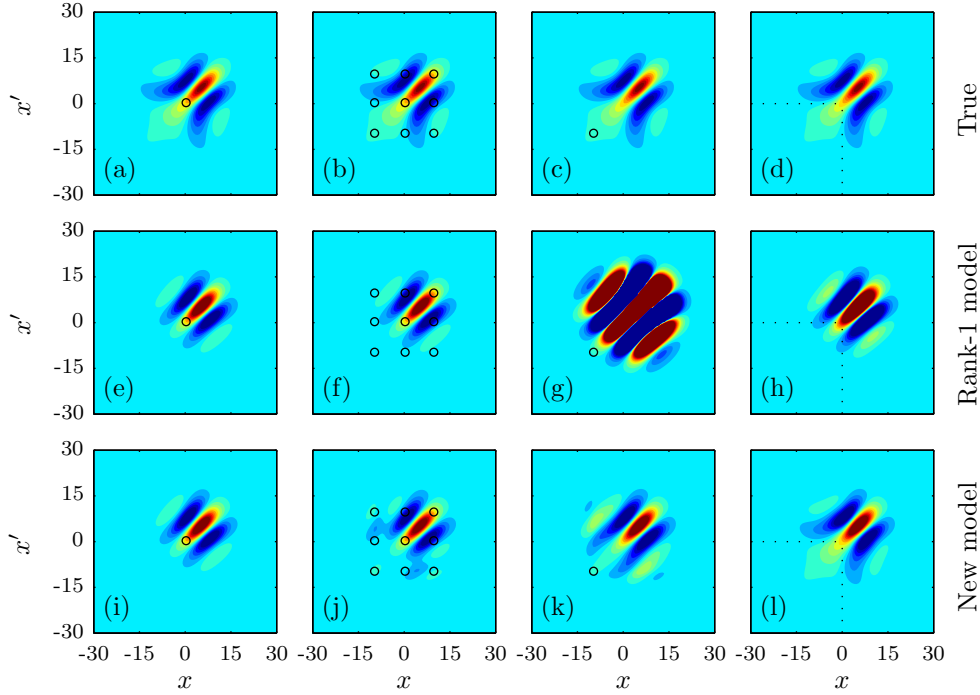


FIGURE 6. Cross-spectral density as a function of x and x' as the frequency of maximum gain, $\omega = -0.6$: (a-d) true values repeated for ease of comparison; (e-h) estimated values from the rank-1 model of Beneddine *et al.* (2016); (i-l) estimated values from the new model presented in this report. The estimates are based on the following probe locations: (e,i) $x = 0$; (f,j) $x = -10, 0$ and 10 ; (g,k) $x = -10$; (h,l) $x \leq 0$.

estimates for the new model are improved by adding two more probes at $x = \pm 10$, while they have little impact on the rank-1 results. Using a single probe at the low-energy point $x = -10$ leads to reasonable estimates using the new method but large errors for the rank-1 method. Finally, using data from the region $x \leq 0$ degrades the rank-1 estimates compared to using only the boundary point $x = 0$ but improves the estimates obtained using the new model. In this case, the CSD estimates from the new model are indistinguishable from the true CSD, even though the known region does not contain the peak PSD.

4. Conclusions

Building on the work of Beneddine *et al.* (2016) and Zare *et al.* (2017), this report introduces a method for completing partially known space-time flow statistics. The method is based on the resolvent methodology developed by McKeon & Sharma (2010) and the statistical interpretation of this theory proposed by Towne *et al.* (2016, 2017b). The central idea of our approach is to use known data to infer the statistics of the nonlinear terms that constitute a forcing on the linearized Navier-Stokes equations.

The method has been demonstrated using the Ginzburg-Landau equation. Using data at three probe locations, the method provides good estimates of the unknown power

spectra, cross-spectra, and space-time correlations within the energetic regions of ω - x or τ - x space. Comparisons are then made with the rank-1 model proposed by Beneddine *et al.* (2016). The two methods give similar results when the probes are placed at locations dominated by a single resolvent mode, but the new method gives superior results when the probes are placed at locations that violate this underlying assumption of the rank-1 model. The improved behavior in this case is important for turbulent flows. Furthermore, the estimates provided by the new method improve with the addition of more known input data, while the rank-1 method does not. While these results are encouraging, the performance of the method for real turbulent flows remains to be evaluated; this is the next step in this investigation.

There may also be opportunities to further improve the method by modeling the portions of the forcing CSD that cannot be observed using the known data. In the current formulation, these terms are simply set to zero, and there exist several possible alternatives. One is to assume that the unobserved forcing is uncorrelated with the observed part and with itself, leading to the approximation

$$\mathbf{S}_{ff} = [\mathbf{V}_1 \ \mathbf{V}_2] \begin{bmatrix} \mathbf{E}_{11} & 0 \\ 0 & a\mathbf{I} \end{bmatrix} [\mathbf{V}_1 \ \mathbf{V}_2]^*. \quad (4.1)$$

An appropriate value for the scalar amplitude a could be determined from the amplitudes of the known \mathbf{E}_{11} terms.

Another possibility is to choose the unobservable terms by insisting that the estimated \mathbf{S}_{ff} projects exclusively onto the first n singular modes of \mathcal{R}_y . The \mathbf{E}_{ij} values that achieve this can be obtained using simple linear algebra manipulations. This possibility is similar to a suggestion by Beneddine *et al.* (2017), except here the expansion coefficients are treated as statistical quantities rather than complex scalars. As shown by Towne *et al.* (2017b), this statistical treatment removes a fundamental accuracy restriction imposed by treating the expansion coefficients as deterministic scalars and allows for a convergent approximation.

Acknowledgments

This investigation was funded by the National Aeronautics and Space Administration, Grant #NNX15AU93A and by the Advanced Simulation and Computing (ASC) program of the US Department of Energys National Nuclear Security Administration via the PSAAP-II Center at Stanford, Grant #DE-NA0002373.

REFERENCES

- BAGHERI, S., HENNINGSON, D. S., HOEPFFNER, J. & SCHMID, P. J. 2009 Input-output analysis and control design applied to a linear model of spatially developing flows. *Appl. Mech. Rev.* **62**, 020803.
- BENEDDINE, S., SIPP, D., ARNAULT, A., DANDOIS, J. & LESSHAFFT, L. 2016 Conditions for validity of mean flow stability analysis. *J. Fluid Mech.* **798**, 485–504.
- BENEDDINE, S., YEGAVIAN, R., SIPP, D. & LECLAIRE, B. 2017 Unsteady flow dynamics reconstruction from mean flow and point sensors: an experimental study. *J. Fluid Mech.* **824**, 174–201.
- CHEN, K. K. & ROWLEY, C. W. 2011 H_2 optimal actuator and sensor placement in the linearised complex Ginzburg–Landau system. *J. Fluid Mech.* **681**, 241–260.

- HUNT, R. E. & CRIGHTON, D. G. 1991 Instability of flows in spatially developing media. *Proc. R. Soc. London Ser. A* **435**, 109–128.
- MCKEON, B. J. & SHARMA, A. S. 2010 A critical-layer framework for turbulent pipe flow. *J. Fluid Mech.* **658**, 336–382.
- SCHMIDT, O. T., COLONIUS, T. & BRÈS, G. A. 2017*a* Linear dynamics of large-scale structures in turbulent jets. In *10th International Symposium on Turbulence and Shear Flow Phenomena*.
- SCHMIDT, O. T., TOWNE, A., RIGAS, G., COLONIUS, T. & BRÈS, G. A. 2017*b* Spectral analysis of jet turbulence. In arXiv preprint, arXiv:1711.06296.
- SEMERARO, O., JAUNET, V., JORDAN, P., CAVALIERI, A. V. G. & LESSHAFFT, L. 2016 Stochastic and harmonic optimal forcing in subsonic jets. *AIAA Paper* #2016-2935.
- SHAMPINE, L. F. & REICHEL, M. W. 1997 The MATLAB ODE suite. *SIAM J. Sci. Comput.* **18**, 1–22.
- TOWNE, A., BRÈS, G. A. & LELE, S. K. 2016 Toward a resolvent-based statistical jet-noise model. *Annual Research Briefs*. Center for Turbulence Research, Stanford University, pp. 247–258.
- TOWNE, A., BRÈS, G. A. & LELE, S. K. 2017*a* A statistical jet-noise model based on the resolvent framework. *AIAA Paper* #2017-3406.
- TOWNE, A., SCHMIDT, O. T. & COLONIUS, T. 2017*b* Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis. in arXiv preprint, arXiv:1708.04393.
- WELCH, P. 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans. Acoust. Speech* **15**, 70–73.
- ZARE, A., JOVANOVIĆ, M. R. & GEORGIU, T. T. 2015 Alternating direction optimization algorithms for covariance completion problems. In *American Control Conference (ACC)*, 2015, pp. 515–520. IEEE.
- ZARE, A., JOVANOVIĆ, M. R. & GEORGIU, T. T. 2017 Colour of turbulence. *J. Fluid Mech.* **812**, 636–680.