Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

A Bayesian fusion model for space-time reconstruction of finely resolved velocities in turbulent flows from low resolution measurements

Linh Van Nguyen[1], Jean-Philippe Laval[1] and Pierre Chainais[2]

[1] USTL, LML, CNRS UMR 8107, F-59650 Villeneuve d'Ascq, France
[2] Ecole Centrale Lille, CRIStAL-CNRS UMR 9189, F-59650 Villeneuve d'Ascq, France

AQ1 E-mail: xxxx

Received 9 April 2015
Accepted for publication 18 August 2015
Published

CrossMark

Online at stacks.iop.org/JSTAT/2015/000000
doi:10.1088/1742-5468/2015/00/000000

Abstract. The study of turbulent flows calls for measurements with high resolution in both space and time. We propose a new approach to reconstruct high-temporal–high-spatial resolution velocity fields by combining two sources of information that are well resolved either in space or in time, the low-temporal–high-spatial (LTHS) and the high-temporal–low-spatial (HTLS) resolution measurements. In the framework of co-conception between sensing and data post-processing, this work extensively investigates a Bayesian reconstruction approach using a simulated database. A Bayesian fusion model is developed to solve the inverse problem of data reconstruction. The model uses a maximum *a posteriori* estimate, which yields the most probable field knowing the measurements. The direct numerical simulation (DNS) of a wall-bounded turbulent flow at moderate Reynolds number is used to validate and assess the performances of the present approach. Low-resolution measurements are subsampled in time and space from the fully resolved data. Reconstructed velocities are compared to the reference DNS to estimate the reconstruction errors. The model is compared to other conventional methods such as linear stochastic estimation and cubic spline interpolation. Results show the superior accuracy of the proposed method in

© 2015 IOP Publishing Ltd and SISSA Medialab srl

1742-5468/15/000000+19$33.00

JNL:JSTAT PIPS: 519267 TYPE: PAP TS: NEWGEN DATE:9/9/15 EDITOR: IOP SPELLING: US

all configurations. Further investigations of model performances on various scales demonstrate its robustness. Numerical experiments also permit one to estimate the expected maximum information level corresponding to limitations of experimental instruments.

Keywords: turbulence

J. Stat. Mech. (2015) 519267

Contents

1. Introduction

Turbulence, though governed by the Navier–Stokes equations, is extremely hard to predict due to its spatiotemporal intermittency as well as three-dimensional and irregular properties. It is also a multi-scale phenomenon where a very wide range of scales from the largest eddies to Kolmogorov microscales coexist and interact. Since the ratio between the largest and the smallest scales increases with Reynolds number as $\mathrm{Re}^{3/4}$, flows with high Reynolds number are the most challenging. Wall-bounded flows are particularly difficult to model due to the overlap of several scaling regions as a function of distance to the wall. Coherent structures in such flows can extend up to several boundary layers in thickness. The modeling of such structures and scales therefore requires extremely detailed flow information in both space and time.

Despite constant progress, none of the experimental techniques, even in academic research, is capable of providing spatiotemporally resolved information in sufficiently wide spatial domains and for diverse flow conditions. Particle image velocimetry (PIV)—the most advanced turbulence measurement technique—cannot measure space-time resolved velocities. Stereoscopic PIV measures three-component velocities at high spatial resolution and with a large field of view, but is limited to a low acquisition rate compared to the flow dynamics. High-repetition tomographic PIV and time-resolved PIV (TrPIV) are improving but still limited to small volumes and low-speed flows. Other point-measurement techniques such as hot wire anemometry (HWA) measure the full temporal dynamics. However, the combination of HWA devices to get better spatial resolution is not straightforward and remains intrusive.

Direct numerical simulation (DNS) can provide reliable and fully resolved velocities of turbulent flows. It simulates the flows by directly solving the Navier–Stokes equations. The computational cost of such a numerical approach is very high since the number of simulated grid points increases as $\mathrm{Re}^{9/4}$. DNS therefore can simulate only flows with low to moderate Reynolds number and simple geometries.

To have fully resolved velocities, one idea is to measure and combine two types of complementary measurements in space and time: the low-temporal–high-spatial resolution (LTHS) and the high-temporal–low-spatial resolution (HTLS) measurements. One particular example of such an idea is presented in [1]. This joint experiment provides a database of boundary layer flows with high Reynolds number. The data are provided by stereoscopic PIV synchronized with a rake of HWA probes. PIV has a large field of view and a high spatial resolution but low acquisition frequency. HWA measurements have extremely high temporal resolution, but the spatial discretization of the rake of probes is very coarse compared to Kolmogorov scales.

Various methods have been proposed to combine such measured data of turbulent flows to recover the maximum information level. Linear stochastic estimation (LSE) is the most common one. Its introduction into the turbulence community dates back to the works by Adrian [2, 3] and it has been further investigated since [4–6]. These works use LSE as a tool to extract coherent structures from the measurements. Later works proposed various extensions such as multi-time, nonlinear or higher-order LSE [7–10]. In these works, unknown velocities are reconstructed from measurements of other quantities such as pressure or shear-stress. LSE can be also linked to proper orthogonal decomposition (POD) to reduce the order of reconstruction problems [11].

The idea of combining sparse velocity measurements to obtain fully resolved fields was not addressed until recently [12, 13]. In [12], 3D smoke intensity and 2D PIV measurements are combined using a POD-LSE model to get fully resolved 3D velocities of a flow over a flat plate. The POD-LSE estimation model has been developed further [13] with a reconstruction scheme based on a multi-time LSE. Either a Kalman filter or a Kalman smoother is used, depending on whether the problem is real-time estimation or data post-processing. The model is tested using TrPIV measurements of a bluff-body wake at a low Reynolds number. Sparse velocity measurements are extracted virtually from the high-resolution ones, while original data are used to estimate reconstruction errors.

LSE suffers from critical limitations though it is used extensively. First, as a conditional average, LSE estimates a set of coefficients that associate the so-called conditional eddies to one flow pattern [3]. Using these coefficients to reconstruct all velocity fields, LSE fails to capture coherent structures and misleads physical interpretations when particular patterns exist. Second, the reconstructed structures are independent of event magnitudes [14]. Reconstructed flows are associated with weak fluctuations only. Last, LSE as a low-pass filter reconstructs large scales only and loses flow details even at measured positions.

The present work proposes a novel model to reconstruct the fully resolved HTHS velocities from HTLS and LTHS measurements. This model is based on a Bayesian inference framework using a maximum *a posteriori* (MAP) estimate [15]. It is inspired by the multispectral image fusion problem with the limited resolution of image measurements in space-wavelength domains [16]. This framework was discussed early in communication problems [17, 18] and is used more extensively in image processing, remote sensing and data fusion [19–25]. The Bayesian fusion model takes benefit from both sources of information in space and time simultaneously by searching for the most probable flow for given measurements. Better performances are expected since space and time correlations are equally important. The model also recovers flow details that are inaccessible from single interpolations. By integrating the measurements directly, it proposes a compromise estimate such that detailed flow information close to the sensor positions is well preserved. This approach also overcomes the limitations of LSE, which acts as a low-pass filter due to the minimization of mean square error. To test the model, the DNS database of a turbulent wall-bounded flow is used. These space-time fully resolved data allow optimization and validation of the model. Sparse measurements of HTLS and LTHS are extracted from the full dataset, while the reference DNS data are used in the end to evaluate reconstruction errors. Performances are evaluated for various configurations with different subsampling ratios.

The paper is organized as follows. Section 2 presents the Bayesian model using a MAP estimate. Model simplification and estimation of statistical parameters are also discussed. Section 3 describes the DNS database used to test the model and also other reconstruction methods for comparison. Results for various configurations are presented. Conclusions and future works are discussed in section 4.

2. Bayesian fusion model

2.1. Bayesian model

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ denote LTHS and HTLS measurements, and $\boldsymbol{z}$ denote HTHS data to reconstruct. $\boldsymbol{z}$, $\boldsymbol{x}$ and $\boldsymbol{y}$ are random, zero-mean vectors of size $NP\times 1$, $NQ\times 1$ and $MP\times 1$ respectively. N and M are numbers of spatial points in each snapshot, while P and Q are numbers of snapshots. The present work is a challenging inverse problem since we consider $M\ll N$ and $Q\ll P$. Let the subscript 's' denote operators performing in space, and 't' be those in time; $\mathbb{I}$ is an interpolator; $\mathbb{S}$ is for subsampling; $\mathbb{L}$ is a low-pass filter

J. Stat. Mech. (2015) 519267

Table 1. Summary of notation.

$\boldsymbol{z}$	NP-dimensional vector of HTHS DNS data
$\boldsymbol{y}$	MP-dimensional vector of HTLS measurements
$\boldsymbol{x}$	NQ-dimensional vector of LTHS measurements
N	number of spatial points in each HTHS/LTHS snapshot
M	number of spatial points in each HTLS snapshot
P	number of HTHS or HTLS snapshots
Q	number of LTHS snapshots
$\mathbb{I}_t$	1D cubic spline interpolator in time
$\mathbb{I}_s$	2D cubic spline interpolator in space
$\mathbb{S}_t$	subsampling in time from P to Q snapshots, $\mathbb{S}_t\boldsymbol{z} = \boldsymbol{x}$
$\mathbb{S}_s$	subsampling in space from N to M points, $\mathbb{S}_s\boldsymbol{z} = \boldsymbol{y}$
$\mathbb{L}_s$	2D fifth-order least-squares spline filter in space
$\mathbb{L}_t$	1D fifth-order least-squares spline filter in time
$\Delta\kappa$	loss of kinetic energy computed using equation (28)
ϵ	normalized root mean square error (NRMSE) defined in equation (36)
$\lvert.\rvert$	determinant of a matrix
$\boldsymbol{u}^T$	transpose of vector $\boldsymbol{u}$
$\underset{\boldsymbol{z}}{\operatorname{argmax}}(.)$	argument of the maximum, which is $\boldsymbol{z}$ for which the function attains its maximum
$\Vert \boldsymbol{u} \Vert_{\Sigma}^2$	square of a Mahalanobis distance: $\Vert \boldsymbol{u} \Vert_{\Sigma}^2 = \boldsymbol{u}^T\Sigma^{-1}\boldsymbol{u}$
$\Vert \boldsymbol{u} \Vert_2^2$	square of a Euclidean distance: $\Vert \boldsymbol{u} \Vert_2^2 = \boldsymbol{u}^T\boldsymbol{u}$
$\boldsymbol{n} \sim \mathcal{N}(0, \Sigma_{\boldsymbol{n}})$	Gaussian noise vector of zero mean and covariance matrix $\Sigma_{\boldsymbol{n}}$
$\mathcal{N}(\boldsymbol{u}\vert\boldsymbol{\mu}_{\boldsymbol{u}}, \Sigma_{\boldsymbol{n}})$	Gaussian distribution of variable $\boldsymbol{u}$ that takes the mean $\boldsymbol{\mu}_{\boldsymbol{u}}$ and fluctuates due to $\boldsymbol{n}$ of covariance $\Sigma_{\boldsymbol{n}}$
$\mathcal{N}(\boldsymbol{u}\vert\boldsymbol{v})$	distribution of $\boldsymbol{u}$ knowing (or conditional on) $\boldsymbol{v}$

(LPF). The cubic spline interpolation, either 1D or 2D, is used as $\mathbb{I}$ for its state-of-the-art interpolation results and finite support [26, 27]. $\mathbb{L}$ is a fifth-order least-squares spline filter [28, 29] for its sharp cutoff response to better separate large scales from small scales. Table 1 lists all notations used in this paper.

The direct model of the measurement system involves the subsampling operator and some measurement noises:

$$\boldsymbol{x} = \mathbb{S}_t\boldsymbol{z} + \boldsymbol{b}_t \tag{1}$$

$$\boldsymbol{y} = \mathbb{S}_s\boldsymbol{z} + \boldsymbol{b}_s \tag{2}$$

where $\boldsymbol{b}_t$ and $\boldsymbol{b}_s$ are the (typically white Gaussian) measurement noises. Therefore, given sparse measurements of either $\boldsymbol{y}$ in space or $\boldsymbol{x}$ in time, two estimators $\hat{\boldsymbol{z}}_1$ and $\hat{\boldsymbol{z}}_2$ of the fully resolved vector $\boldsymbol{z}$ can be reconstructed by single interpolations. The 1D time interpolation goes from NQ- to NP-dimensional space, i.e. $\boldsymbol{x} \longmapsto \hat{\boldsymbol{z}}_1 = \mathbb{I}_t\boldsymbol{x}$, while the 2D space interpolation goes from MP- to NP-dimensional space, i.e. $\boldsymbol{y} \longmapsto \hat{\boldsymbol{z}}_2 = \mathbb{I}_s\boldsymbol{y}$.

J. Stat. Mech. (2015) 519267

Let NP -dimensional vectors $\boldsymbol{h}_s$ and $\boldsymbol{h}_t$ denote the information that cannot be recovered by simple interpolations; $\boldsymbol{z}$ can be modeled in two ways from these separate estimates:

$$\boldsymbol{z} = \hat{\boldsymbol{z}}_1 + \boldsymbol{h}_t = \mathbb{I}_t\boldsymbol{x} + \boldsymbol{h}_t \tag{3}$$

$$\boldsymbol{z} = \hat{\boldsymbol{z}}_2 + \boldsymbol{h}_s = \mathbb{I}_s\boldsymbol{y} + \boldsymbol{h}_s \tag{4}$$

Missing information $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$ essentially feature small scales. Using either $\boldsymbol{x}$ or $\boldsymbol{y}$, it is not possible to estimate $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$. The idea of Bayesian fusion is to combine the two models by using $\mathbb{I}_t\boldsymbol{x}$ in (3) to estimate the unknown $\boldsymbol{h}_s$ in (4) and vice versa.

Let $\mathcal{N}(\boldsymbol{u}|\boldsymbol{\mu}_{\boldsymbol{u}}, \Sigma_{\boldsymbol{u}})$ denote the multivariate Gaussian distribution of an NP-dimensional random vector $\boldsymbol{u}$ with mean value $\boldsymbol{\mu}_{\boldsymbol{u}}$ and covariance matrix $\Sigma_{\boldsymbol{u}}$. The $NP \times NP$ matrix is the expectation of $(\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}})(\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}})^T$. The probability density function (pdf) of $\boldsymbol{u}$ with a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{u}|\boldsymbol{\mu}_{\boldsymbol{u}}, \Sigma_{\boldsymbol{u}})$ is

$$p(\boldsymbol{u}) = \frac{1}{(2\pi)^{NP/2}|\Sigma_{\boldsymbol{u}}|^{1/2}} \exp\left(-\frac{1}{2}\|\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}}\|^2_{\Sigma_u}\right) \tag{5}$$

where |.| denotes the matrix determinant, and $\|\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}}\|^2_{\Sigma_u}$ is the Mahalanobis distance:

$$\|\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}}\|^2_{\Sigma_n} = (\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}})^T\Sigma_{\boldsymbol{n}}^{-1}(\boldsymbol{u}-\boldsymbol{\mu}_{\boldsymbol{u}}). \tag{6}$$

Let us assume that $\mathbb{I}_t\boldsymbol{x}$ and $\boldsymbol{h}_t$ are approximately independent; $\mathbb{I}_t\boldsymbol{x}$ captures large temporal scales of $\boldsymbol{x}$. Similarly, $\mathbb{I}_s\boldsymbol{y}$ and $\boldsymbol{h}_s$ are assumed to be approximately independent. Due to subsampling, aliasing terms are also present in each pair of $(\mathbb{I}_t\boldsymbol{x}, \boldsymbol{h}_t)$ and $(\mathbb{I}_s\boldsymbol{y}, \boldsymbol{h}_s)$. Assume also that $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$ are zero-mean Gaussian noises, i.e. $\boldsymbol{h}_t \sim \mathcal{N}(0, \Sigma_{\boldsymbol{h}_t})$ and $\boldsymbol{h}_s \sim \mathcal{N}(0, \Sigma_{\boldsymbol{h}_t})$. Pdfs of these unknowns are modeled as

$$p(\boldsymbol{h}_t) = \frac{1}{(2\pi)^{NP/2}|\Sigma_{\boldsymbol{h}_t}|^{1/2}} \exp\left(-\frac{1}{2}\|\boldsymbol{h}_t\|^2_{\Sigma_{h_t}}\right) \tag{7}$$

and similarly for $p(\boldsymbol{h}_s)$. Posterior distributions of $\boldsymbol{z}$ knowing either $\boldsymbol{x}$ or $\boldsymbol{y}$ are then modeled as

$$\mathcal{N}(\boldsymbol{z}|\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{z}|\mathbb{I}_t\boldsymbol{x}, \Sigma_{\boldsymbol{h}_t}) \tag{8}$$

$$\mathcal{N}(\boldsymbol{z}|\boldsymbol{y}) \sim \mathcal{N}(\boldsymbol{z}|\mathbb{I}_s\boldsymbol{y}, \Sigma_{\boldsymbol{h}_s}) \tag{9}$$

where $\mathcal{N}(\boldsymbol{z}|\boldsymbol{x})$ (respectively $\mathcal{N}(\boldsymbol{z}|\boldsymbol{y})$) is the posterior distribution of $\boldsymbol{z}$ knowing $\boldsymbol{x}$ (respectively knowing $\boldsymbol{y}$).

2.2. MAP estimation

The present Bayesian model aims to build an estimate of $\boldsymbol{z}$ given $\boldsymbol{x}$ and $\boldsymbol{y}$ using the probability models (8) and (9). The model uses a MAP estimate to search for the most probable $\hat{\boldsymbol{z}}$ given $\boldsymbol{x}$ and $\boldsymbol{y}$ such that $\hat{\boldsymbol{z}}$ maximizes the posterior pdf $p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y})$:

doi:10.1088/1742-5468/2015/00/000000

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\operatorname{argmax}}\; p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}). \tag{10}$$

Using Bayesian rules [30], one has

$$p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}.)p(\boldsymbol{z}) \tag{11}$$

Assuming that $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent conditional on $\boldsymbol{z}$, equation (11) becomes

$$p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{y}|\boldsymbol{z})p(\boldsymbol{z}). \tag{12}$$

In equation (12), the likelihood functions $p(\boldsymbol{x}|\boldsymbol{z})$, $p(\boldsymbol{y}|\boldsymbol{z})$ and the prior pdf $p(\boldsymbol{z})$ appear, while only the posterior probabilities $p(\boldsymbol{z}|\boldsymbol{x})$ and $p(\boldsymbol{z}|\boldsymbol{y})$ are available in (8) and (9).

To complete the model, the likelihood functions can be expressed in terms of the posterior pdfs and prior of $\boldsymbol{z}$ using Bayesian rules. section 2.3.3 in [31] introduces an alternative way to estimate these functions from posterior pdfs using a linear Gaussian model. Various tests of different Gaussian priors $p(\boldsymbol{z})$ lead to the use of a noninformative prior. This prior, referred to also as a vague or flat prior, assumes that all the values of $\boldsymbol{z}$ are equally likely [32]. The estimation of $\hat{\boldsymbol{z}}$ is now solely based on the measurements and not influenced by external information. The prior distribution therefore has no influence on the posterior pdfs.

With the assumption of a noninformative prior, $p(\boldsymbol{z})$ is constant. Using Bayesian rules, the relation between the likelihood function and the posterior pdf is

$$p(\boldsymbol{z}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z}). \tag{13}$$

Since $p(\boldsymbol{z})$ is replaced by a constant, one gets $p(\boldsymbol{x}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{x})$. $p(\boldsymbol{y}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{y})$. Equation (12) becomes

$$p(\boldsymbol{z}|\boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{z}|\boldsymbol{y}). \tag{14}$$

The MAP estimation is

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\operatorname{argmax}}\; p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{z}|\boldsymbol{y}). \tag{15}$$

Logarithms of $p(\boldsymbol{z}|\boldsymbol{x})$ and $p(\boldsymbol{z}|\boldsymbol{y})$ are

$$-\ln p(\boldsymbol{z}|\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{z} - \mathbb{I}_t\boldsymbol{x}\|^2_{\Sigma_{h_t}} + C_1 \tag{16}$$

$$-\ln p(\boldsymbol{z}|\boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{z} - \mathbb{I}_s\boldsymbol{y}\|^2_{\Sigma_{h_s}} + C_2 \tag{17}$$

where C_1 and C_2 are independent of $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$. Solving (15) is equivalent to minimizing the cost function

$$C(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{z} - \mathbb{I}_t\boldsymbol{x}\|^2_{\Sigma_{h_t}} + \frac{1}{2}\|\boldsymbol{z} - \mathbb{I}_s\boldsymbol{y}\|^2_{\Sigma_{h_s}}. \tag{18}$$

Computing the gradient of $C(\boldsymbol{z})$ and setting to zero:

$$\frac{\partial C(\boldsymbol{z})}{\partial \boldsymbol{z}} = \Sigma^{-1}_{\boldsymbol{h}_s}(\boldsymbol{z} - \mathbb{I}_s\boldsymbol{y}) + \Sigma^{-1}_{\boldsymbol{h}_t}(\boldsymbol{z} - \mathbb{I}_t\boldsymbol{x}) = 0, \tag{19}$$

the solution to the optimization problem (10) is

J. Stat. Mech. (2015) 519267

$$\hat{\boldsymbol{z}} = \left(\Sigma_{\boldsymbol{h}_t}^{-1} + \Sigma_{\boldsymbol{h}_s}^{-1}\right)^{-1}\left(\Sigma_{\boldsymbol{h}_s}^{-1}\mathbb{I}_s\boldsymbol{y} + \Sigma_{\boldsymbol{h}_t}^{-1}\mathbb{I}_t\boldsymbol{x}\right). \tag{20}$$

Applying the matrix inversion lemma [15],

$$(A + BD^{-1}C)^{-1} = A^{-1} - A^{-1}B(D + CA^{-1}B)^{-1}CA^{-1}, \tag{21}$$

equation (20) can be rewritten as

$$\hat{\boldsymbol{z}} = (\Sigma_{\boldsymbol{h}_t} + \Sigma_{\boldsymbol{h}_s})^{-1}(\Sigma_{\boldsymbol{h}_t}\mathbb{I}_s\boldsymbol{y} + \Sigma_{\boldsymbol{h}_s}\mathbb{I}_t\boldsymbol{x}). \tag{22}$$

Equation (22) is the final full form of the proposed Bayesian fusion model using a MAP estimate and assuming a noninformative prior of $\boldsymbol{z}$. Variance matrices $\Sigma_{\boldsymbol{h}_t}$ and $\Sigma_{\boldsymbol{h}_s}$ are parameters to be estimated.

2.3. Model simplification

Though providing the full theoretical estimate of $\hat{\boldsymbol{z}}$, equation (22) is impractical to use as it is for several reasons. The full covariance matrices $\Sigma_{\boldsymbol{h}_t}$ and $\Sigma_{\boldsymbol{h}_s}$, representing all the sources of correlations in space and time, cannot be estimated from only the measurements $\boldsymbol{x}$ and $\boldsymbol{y}$. This is because the unknown $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$ are only accessible at the measured positions in space and time. Also, the covariance matrices of size $NP \times NP$ are very large, making them very difficult to estimate accurately and to invert. Additional assumptions on the shape of $\Sigma_{\boldsymbol{h}_t}$ and $\Sigma_{\boldsymbol{h}_s}$ are necessary.

A common and simple approach is to assume diagonal covariance matrices. This implies the independence between all elements of $\boldsymbol{h}_t$ and $\boldsymbol{h}_s$. The simplified version of equation (22) becomes a point-wise formula:

$$\hat{\boldsymbol{z}}(i) = \frac{\boldsymbol{\sigma}_{\boldsymbol{h}_s}^2(i)}{\boldsymbol{\sigma}_{\boldsymbol{h}_s}^2(i) + \boldsymbol{\sigma}_{\boldsymbol{h}_t}^2(i)}\mathbb{I}_t\boldsymbol{x}(i) + \frac{\boldsymbol{\sigma}_{\boldsymbol{h}_t}^2(i)}{\boldsymbol{\sigma}_{\boldsymbol{h}_s}^2(i) + \boldsymbol{\sigma}_{\boldsymbol{h}_t}^2(i)}\mathbb{I}_s\boldsymbol{y}(i) \tag{23}$$

where $i(t, s)$ is the index of each point in time (t) and space (s). The variances $\boldsymbol{\sigma}_{\boldsymbol{h}_s}^2$ and $\boldsymbol{\sigma}_{\boldsymbol{h}_t}^2$ are functions of each position in space and time. Their estimation is detailed in the next section. Then equation (23) will be used to reconstruct HTHS data. As a weighted average, it proposes a compromise estimate from the measurements. With a symmetrical form in space and time, the model uses information from both measurements to correct large-scale reconstruction and recover certain information at smaller scales.

2.4. Estimation of statistical parameters

Let Z, X, Y, H_t, H_s, $\Gamma_{\boldsymbol{h}_t}$ and $\Gamma_{\boldsymbol{h}_s}$ be the (time, space) matrix forms of $\boldsymbol{z}$, $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{h}_t$, $\boldsymbol{h}_s$, $\boldsymbol{\sigma}_{\boldsymbol{h}_t}^2$ and $\boldsymbol{\sigma}_{\boldsymbol{h}_s}^2$ respectively. Z, H_t, H_s, $\Gamma_{\boldsymbol{h}_t}$ and $\Gamma_{\boldsymbol{h}_s}$ are of size $P \times N$, while X and Y are of size $Q \times N$ and $P \times M$. $\Gamma_{\boldsymbol{h}_t}$ and $\Gamma_{\boldsymbol{h}_s}$ are matrices of empirical variances, which are functions of time and space (t, s).

Variance matrices are estimated from H_t and H_s, which are available at the measurement positions only. We use

$$\mathbb{S}_t H_s = X - \mathbb{I}_s\mathbb{S}_s X \tag{24}$$

J. Stat. Mech. (2015) 519267

$$\mathbb{S}_s H_t = Y - \mathbb{I}_t \mathbb{S}_t Y \tag{25}$$

where $\mathbb{S}_t$ subsamples in time from time steps P to Q, and $\mathbb{S}_s$ subsamples in space from points N to M. These Q instants and M positions are the same as for LTHS and HTLS measurements. Since the flow is approximately stationary and spatial interpolation is independent of time, $\Gamma_{\boldsymbol{h}_s}(t, s)$ becomes $\Gamma_{\boldsymbol{h}_s}(s)$, a function of spatial locations only. These variances are estimated by averaging over all time steps:

$$\Gamma_{\boldsymbol{h}_s}(s) = \frac{1}{Q} \sum_{t=1}^{Q} (X(t, s) - \mathbb{I}_s \mathbb{S}_s X(t, s))^2. \tag{26}$$

The variance in $\Gamma_{\boldsymbol{h}_t}$ is a function of distance τ to the previous LTHS time step only, where $\tau/\delta t = 0, 1, 2, ..., P/Q$, and δt is the time lag between two consecutive HTHS time steps. $\Gamma_{\boldsymbol{h}_t}$ becomes a function of space and τ only, i.e. $\Gamma_{\boldsymbol{h}_t}(\tau, s)$. It is estimated by averaging over Q blocks (of P/Q snapshots) bounded by two consecutive LTHS instants:

$$\Gamma_{\boldsymbol{h}_t}(\tau, s) = \frac{1}{Q} \sum_{t_s} (Y(t_s, s) - \mathbb{I}_t \mathbb{S}_t Y(t_s, s))^2 \tag{27}$$

where $t_s/\delta t = \tau/\delta t, \tau/\delta t + P/Q, \tau/\delta t + 2P/Q, ..., \tau/\delta t + (Q-1)P/Q$. Since the flow is approximately homogeneous in the spanwise direction, $\Gamma_{\boldsymbol{h}_s}(s)$ and $\Gamma_{\boldsymbol{h}_t}(\tau, s)$ are also averaged over all blocks defined by the four neighboring HTLS measurements, see figure 2. The variances are then functions of only vertical positions and relative distances to the four closest HTLS sensors. These estimated variances are rearranged into a vector form $\boldsymbol{\sigma}^2_{\boldsymbol{h}_t}(i)$ and $\boldsymbol{\sigma}^2_{\boldsymbol{h}_s}(i)$ to complete the fusion model using the simplified formula in equation (23).

3. Numerical experiments

Section 3.1 describes the DNS database used to test the model. Section 3.2 discusses other reconstruction methods for comparison. Section 3.3 presents results of the fusion model in various cases.

3.1. DNS database

DNS database of a turbulent wall-bounded flow is used to test the model. This simulation uses the numerical procedure described in [33]. The flow is at a Reynolds number $\mathrm{Re}_\tau = 550$ based on the friction velocity. Cartesian coordinates of the simulation in space are (x, y, z) for streamwise, vertical and spanwise directions respectively. The domain size $L_x \times L_y \times L_z$ normalized by half the channel height H is $2\pi \times 2 \times \pi$. Fully resolved fluctuating streamwise velocities in a plane normal to the flow direction are considered as HTHS data. These data include $P = 10\,000$ snapshots at a spatial resolution of $N = 288 \times 257$ and sampling frequency of 40 Hz. Sparse LTHS and HTLS measurements are subsampled from HTHS data to learn the fusion model. HTHS is used

J. Stat. Mech. (2015) 519267

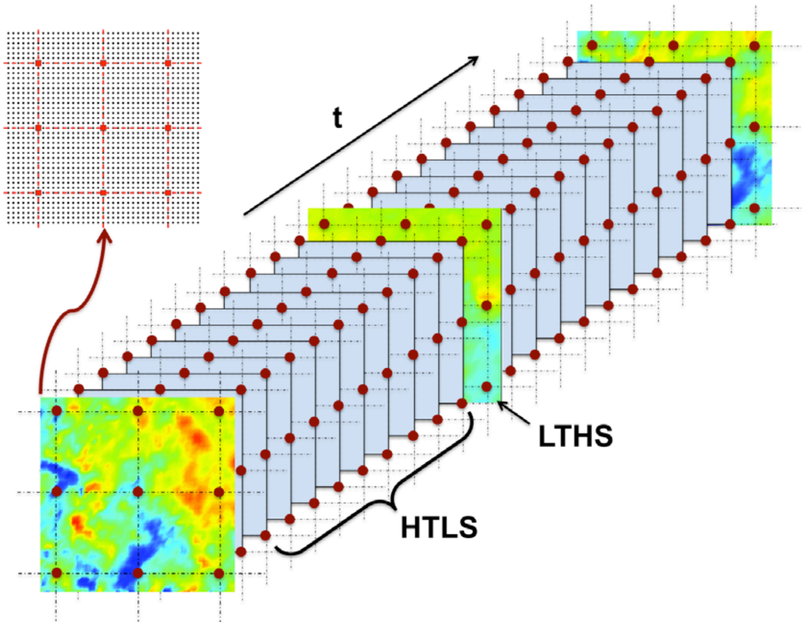

Figure 1. Sketch of the inverse problem, with the two sources of measurements: the LTHS (color images) and a coarse grid of HTLS (red dots among the black ones of LTHS). The inverse problem of HTHS data reconstruction is to fill in the space-time data-cube.

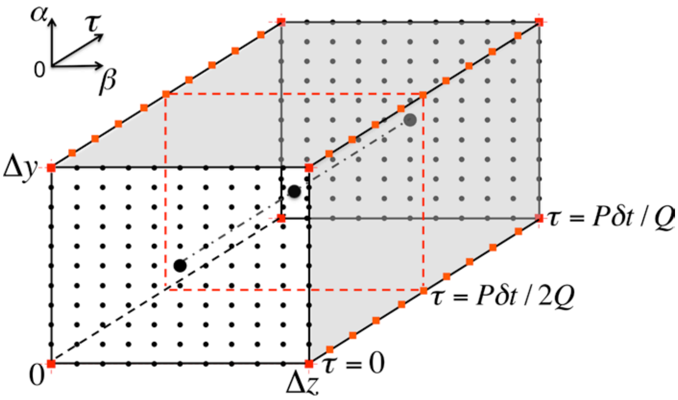

Figure 2. Sketch of an element block with local coordinates (α, β, τ). LTHS time steps are at $\tau/\delta t = 0$ and $\tau/\delta t = P/Q$. HTLS measurements are represented by red dots and LTHS measurements by black ones.

as the ground truth to estimate reconstruction errors. The extension to spanwise and vertical velocity components follows the same procedure.

Various cases are investigated. The subsampling ratios $\sqrt{N/M}$ applied in each direction of space are 5, 10 and 20. These ratios correspond to a number M of HTLS sensors of 51×57, 26×29 and 13×15 respectively. Each ratio has a spacing between two successive HTLS points in spanwise and vertical directions of Δz and Δy. Subsampling ratios P/Q in time are 4 ($Q = 2500$), 10 ($Q = 1000$) and 20 ($Q = 500$). Each ratio, both in space and time, corresponds to a certain amount of energy loss. This is essentially the energy of small scales separated from large scales by a low-pass filter $\mathbb{L}$. Here $\mathbb{L}$ is the fifth-order least-squares spline filter, either temporal 1D ($\mathbb{L}_t$) or spatial 2D ($\mathbb{L}_s$), using measurements as knots. This spline filter has the advantages of a sharp cutoff response and finite support. The energy loss is defined by comparing the filtered field $\mathbb{L}\boldsymbol{z}$ and the original field $\boldsymbol{z}$:

$$\Delta\kappa = \frac{\sqrt{\sum_{j\in\mathbb{J}} \boldsymbol{z}_j^2} - \sqrt{\sum_{j\in\mathbb{J}} [\mathbb{L}\boldsymbol{z}]_j^2}}{\sqrt{\sum_{j\in\mathbb{J}} \boldsymbol{z}_j^2}} \tag{28}$$

where $\mathbb{J}$ is the considered set of points. Table 2 gathers the energy loss in time ($\Delta\kappa_t$) and in space ($\Delta\kappa_s$) estimated with $\mathbb{L}_t$ and $\mathbb{L}_s$ respectively. The set $\mathbb{J}$ contains all points at $y/H = 1$.

3.2. Other methods for comparison

Other reconstruction methods are used for comparison with the present model.

3.2.1. Cubic spline interpolation. Interpolation techniques reconstruct HTHS velocities from either LTHS or HTLS measurements independently, i.e. $\boldsymbol{x} \longmapsto \hat{\boldsymbol{z}} = \mathbb{I}_t\boldsymbol{x}$ or $\boldsymbol{y} \longmapsto \hat{\boldsymbol{z}} = \mathbb{I}_s\boldsymbol{y}$. The cubic spline interpolations [26], either 1D in time or 2D in space, are used. These interpolations are by Matlab built-in functions, which follow the algorithm in [34].

3.2.2. Linear stochastic estimation. LSE estimates $\hat{\boldsymbol{z}}$ as a linear combination of measurements. Coefficients are estimated from the measurements by solving a system of linear equations to minimize the mean square errors of reconstructed fields. References [3, 5] describe the physical interpretations of this procedure. This section derives the model differently [31, 35] but in accordance with turbulence literature.

Matrix forms X , Y and Z described in section 2.4 are used to build the LSE model. Let $Y_s = \mathbb{S}_t Y$ of size $Q \times M$ denote a part of Y subsampled at the same instants as X . The LSE model finds the optimal matrix B of size $N \times M$ that minimizes the residual sum of squared errors:

$$B = \underset{B}{\operatorname{argmin}} \, \| Y_s B - X \|_2^2 \tag{29}$$

Let us set the gradient of this residual sum to zero:

$$\frac{\partial \| Y_s B - X \|_2^2}{\partial B} = Y_s^T (Y_s B - X) = 0, \tag{30}$$

then the optimal B is obtained as

$$B = (Y_s^T Y_s)^{-1} Y_s^T X. \tag{31}$$

Equation (31) requires the inversion of $(Y_s^T Y_s)$, which can be singular, leading to a high variance model with large coefficients. A small change of predictors Y can then lead to a very different reconstruction of Z, causing the model's instability. Tikhonov regularization [36], well known in machine learning problems as L2 penalty or ridge regression

J. Stat. Mech. (2015) 519267

Table 2. NRMSEs of all scales reconstruction errors for various cases.

	Ratios		Spacings		Energy loss		$\bar{\epsilon}$				$\epsilon_{\max}$			
Case	$\sqrt{N/M}$	P/Q	$\Delta z/H$	Δt (s)	$\Delta\kappa_s$	$\Delta\kappa_t$	$\mathbb{I}_s\boldsymbol{y}$	$\mathbb{I}_t\boldsymbol{x}$	LSE	Fusion	$\mathbb{I}_s\boldsymbol{y}$	$\mathbb{I}_t\boldsymbol{x}$	LSE	Fusion
1	5	10	0.05	0.25	0.29	4.70	0.14	0.32	0.25	**0.12**	**0.16**	0.56	0.33	**0.16**
2	5	20	0.05	0.50	0.29	13.12	0.14	0.54	0.38	**0.13**	**0.16**	0.89	0.49	**0.16**
3	10	4	0.11	0.10	5.03	0.50	0.36	**0.11**	0.30	**0.11**	0.47	**0.17**	0.37	0.18
4	20	4	0.22	0.10	9.99	0.50	0.68	**0.11**	0.57	**0.11**	0.86	**0.17**	0.68	**0.17**
5	5	4	0.05	0.10	0.29	0.50	0.14	0.11	0.13	**0.08**	0.16	0.18	0.15	**0.13**
6	10	10	0.11	0.25	5.03	4.68	0.36	0.32	0.34	**0.25**	0.47	0.55	0.49	**0.43**
7	20	20	0.22	0.50	9.99	13.12	0.68	0.54	0.64	**0.46**	0.85	0.85	0.78	**0.73**

Note: The subsampling ratios of HTLS measurements are $\sqrt{N/M}$ and equal in both spatial directions. The ratios of LTHS measurements in time are P/Q . The equivalent spacing in the spanwise direction is normalized by half the channel height as $\Delta z/H$ and the spacing in time is Δt. The normalized energy losses in space $\Delta\kappa_s$ and in time $\Delta\kappa_t$ are defined in equation (28). $\bar{\epsilon}$ and $\epsilon_{\max}$ are the mean and maximum NRMSE defined in equation (36). $\bar{\epsilon}$ is averaged over all space-time positions in the outer region $y/H \in [0.25, 1.75]$, while $\epsilon_{\max}$ is computed for one of the most difficult positions in space and time (the most remote from all nearby measurements). The smallest errors in each cases are boldfaced.

J. Stat. Mech. (2015) 519267

[31, 35], can be used as a remedy. It aims to solve this ill-posed problem by imposing an L2 penalty term on the residual sum of errors. The optimization problem (29) becomes

$$B = \underset{B}{\operatorname{argmin}} \| Y_s B - X\|_2^2 + \lambda \|B\|_2^2 . \tag{32}$$

Setting the gradient of the cost function (for $\lambda > 0$) to zero,

$$\frac{\partial(\| Y_s B - X\|_2^2 + \lambda \|B\|_2^2)}{\partial B} = Y_s^T(Y_s B - X) + \lambda B = 0 \tag{33}$$

the closed form of B is

$$B = (Y_s^T Y_s + \lambda I)^{-1} Y_s^T X. \tag{34}$$

The regularization parameter λ can be optimized by ten-fold cross-validation [37]. The fully resolved field of Z is then estimated using these coefficients:

$$Z = YB. \tag{35}$$

Matrix B encodes the predictor of Z knowing Y learnt from the joint observation of X and Y. A completely analogous procedure can be used that switches the roles of X and Y.

3.3. Results

3.3.1. Impact of subsampling ratios. The fusion model uses equation (23) to reconstruct fully resolved velocities $\hat{\boldsymbol{z}}$ in various cases. Reconstructed fields are compared with the original DNS via the normalized root mean square error (NRMSE):

$$\text{NRMSE} = \left(\frac{\sum_{j \in \mathbb{J}} (\hat{\boldsymbol{z}}_j - \boldsymbol{z}_j)^2}{\sum_{j \in \mathbb{J}} \boldsymbol{z}_j^2} \right)^{1/2} \tag{36}$$

where $\mathbb{J}$ is the considered set of points used to estimate the error. The field $\boldsymbol{z}$ is more or less difficult to estimate depending on the considered instant and position with respect to available measurements. To qualify, two types of NRMSE, the mean NRMSE $\bar{\epsilon}$ and the maximum NRMSE $\epsilon_{\max}$, are estimated. $\bar{\epsilon}$ is estimated over $\mathbb{J}$ including all space-time positions in the outer region of $y/H \in [0.25, 1.75]$, where the flow is approximately homogeneous. It represents how far the reconstructed field departs from ground truth in order to evaluate the reconstruction accuracy. $\epsilon_{\max}$ is estimated using all blocks (in time and in spanwise directions) bounded by HTLS sensors at $y/H = 0.94$ and $y/H = 1.06$, see figure 2. The set $\mathbb{J}$ includes centers at local coordinates $(\Delta y/2, \Delta z/2, P\delta t/2Q)$ of all blocks. $\bar{\epsilon}$ and $\epsilon_{\max}$ of $\mathbb{I}_s\boldsymbol{y}$, $\mathbb{I}_t\boldsymbol{x}$ and LSE reconstruction are also estimated for comparison.

Table 2 describes seven cases with their settings and reconstruction errors. In cases 1 and 2, the energy losses due to subsampling in time are much higher than those in space, and vice versa in cases 3 and 4. The model gives similar errors to the best interpolation, with smaller $\bar{\epsilon}$ and comparable $\epsilon_{\max}$. In cases 5 to 7, the losses are due to both the subsamplings in space and time in a balanced manner. The proposed model reduces $\bar{\epsilon}$ by 15% to 30% and $\epsilon_{\max}$ by 10% to 20% compared to the best of other methods.

J. Stat. Mech. (2015) 519267

Table 3. NRMSEs of large and small scales.

	$\bar{\epsilon}$				$\epsilon_{\max}$			
Case	$\mathbb{I}_s\boldsymbol{y}$	$\mathbb{I}_t\boldsymbol{x}$	LSE	Fusion	$\mathbb{I}_s\boldsymbol{y}$	$\mathbb{I}_t\boldsymbol{x}$	LSE	Fusion
Reconstruction at large scales								
5	0.08	0.09	0.10	**0.05**	**0.06**	0.15	0.11	0.07
6	0.24	0.25	0.24	**0.15**	0.22	0.45	0.28	**0.21**
7	0.56	0.36	0.51	**0.30**	0.60	0.66	0.60	**0.46**
Reconstruction at small scales								
5	0.98	0.56	0.73	**0.55**	0.98	0.86	0.89	**0.81**
6	0.98	0.81	0.90	**0.70**	0.97	1.15	1.07	**0.92**
7	0.99	0.92	0.95	**0.78**	0.93	1.08	0.90	**0.86**

Note: Notations are explained in table 2.

Improvements are expected from the weighted average in equation (23). The present model uses variances $\boldsymbol{\sigma}_s^2(i)$ and $\boldsymbol{\sigma}_t^2(i)$ as parameters of the flow's physics, and $\mathbb{I}_t\boldsymbol{x}$ and $\mathbb{I}_s\boldsymbol{y}$ as the specific flow information. It imposes the reconstruction to be consistent with measurements at nearby positions and proposes compromise estimates elsewhere. Simple interpolations use either HTLS or LTHS measurements only, losing information from the other source. LSE learns its coefficients from both measurements but inherits the limitations of the conditional averaging.

3.3.2. Reconstruction at large and small scales. In cases 1 to 4, the fusion model performs as the best interpolation with small improvements. This is expected since one measurement of HTLS or LTHS is much better resolved than the other. Cases 5 to 7 are the most interesting since energy losses due to subsampling in space and time are comparable. The model brings complementary information from both measurements and improves the reconstruction.

We study reconstructions of large and small scales in detail for these three cases. Spatial 2D filters $\mathbb{L}_s$ (see section 3.1) are used to separate large scales from small scales. These filters take HTLS points as knots to have a cutoff close to the Nyquist frequency. The large scales reconstructed by all methods are compared to the reference $\mathbb{L}_s\boldsymbol{z}$. Small scales are estimated using $\mathbf{I} - \mathbb{L}_s$ where $\mathbf{I}$ is the identity matrix. Table 3 shows NRMSEs estimated using equation (36) but normalized by the RMS of either $\mathbb{L}_s\boldsymbol{z}$ or $(\mathbf{I} - \mathbb{L}_s)\boldsymbol{z}$.

The fusion model recovers part of the small scales from complementary measurements. It gives the lowest $\bar{\epsilon}$ and $\epsilon_{\max}$ of small-scale reconstruction in all cases. It also reconstructs large scales better than other methods. For large scales, $\epsilon_{\max}$ remains the same in case 5 of small subsampling ratios and improves significantly in cases 6 and 7 of high ratios, with $\epsilon_{\max}$ reduced by 5% and 25% respectively, and $\bar{\epsilon}$ by 20% to 40% compared to the best of other methods.

3.3.3. Model performance analysis. We focus on case 6 for a model performance analysis. This case has about 5% energy losses due to both time and space subsamplings, which are critical to highlight interests of the present approach. The model reduces $\bar{\epsilon}$

J. Stat. Mech. (2015) 519267

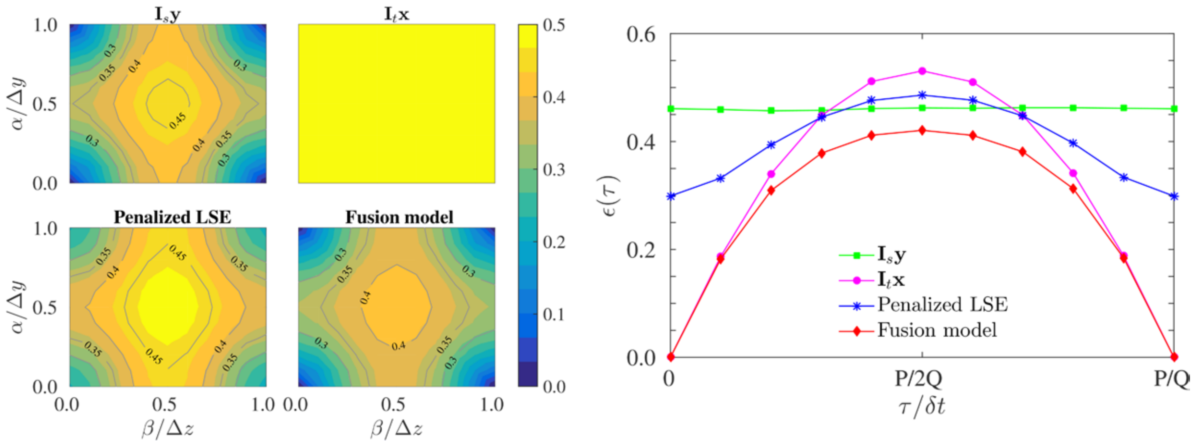

Figure 3. NRMSEs between reference streamwise velocities and those reconstructed by all methods as: (a) functions of spatial coordinates in an element block at the most difficult instant, i.e. at $(\alpha, \beta, P\delta t/2Q)$; (b) functions of time separations from the previous LTHS instant at the most difficult spatial location, i.e. at $(\Delta y/2, \Delta z/2, \tau)$.

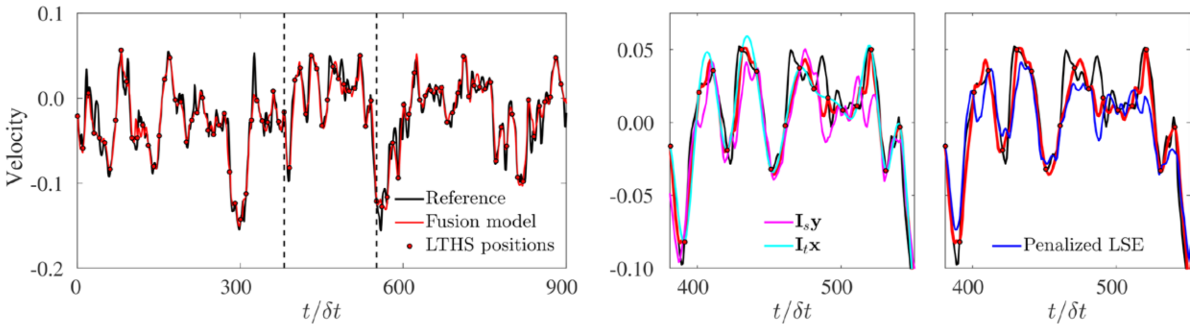

Figure 4. A time evolution of fluctuating streamwise velocity at $y/H = 1$ and $z/H = 0$, the centers of all such (α, β) planes in figure 2.

and $\epsilon_{\max}$ by 25% and 35% respectively for reconstruction at all scales, and by 10% and 5% for large-scale reconstruction.

To analyze reconstructions in space, figure 3(a) shows spatial NRMSE maps from all methods as functions of local coordinates (α, β). For each (α, β), NRMSE is estimated using equation (36), where $\mathbb{J}$ includes points at $(\alpha, \beta, P\delta t/2Q)$ of all blocks used to estimate $\epsilon_{\max}$ (see section 3.3.1). For all methods, NRMSEs are small close to the four HTLS positions in the corners and increase when approaching the center. Time interpolation behaves differently since its errors are independent of spatial coordinates. The fusion model yields the smallest errors at all positions. It improves significantly near the center compared to spatial interpolation, the best of other methods.

To analyze reconstructions in time, figure 3(b) shows the NRMSE curves from all methods as functions of distances τ from the previous LTHS time step. For each τ, NRMSE is estimated using $\mathbb{J}$ including points at local coordinates $(\Delta y/2, \Delta z/2, \tau)$ of all blocks used to estimate $\epsilon_{\max}$. NRMSEs are small close to the LTHS measurements

doi:10.1088/1742-5468/2015/00/000000

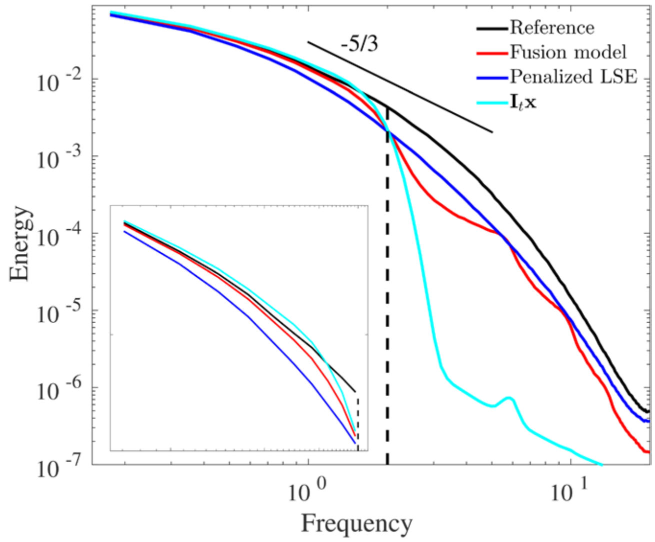

Figure 5. Spectra of the fluctuating velocity in figure 4.

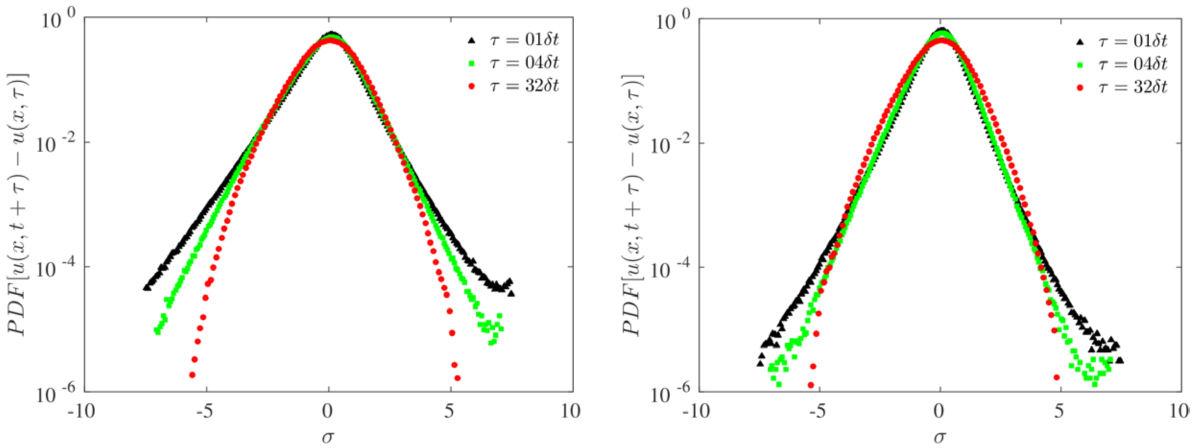

Figure 6. Probability distribution functions of velocity increments in (left) the original DNS amd (right) the reconstructed field.

($\tau/\delta t = 0$ and $\tau/\delta t = P/Q$) and increase when moving toward the middle ($\tau/\delta t = P/2Q$). Spatial interpolations are different with NRMSEs independent of time. The fusion model yields the minimum errors at all time steps. Even in the middle of two LTHS instants, the maximum fusion error remains significantly lower than that of all other methods.

Figure 4 shows a time evolution of the point at $y/H = 1$ and $z/H = 0$ ($\alpha = \Delta y/2$ and $\beta = \Delta z/2$ in local coordinates), the most remote from its neighboring HTLS sensors. A good agreement between fused and reference velocity is still obtained. A zoom-in period is shown also for detailed comparisons with other methods. While time interpolation captures only low frequencies, spatial interpolation generates high frequencies but they are weakly correlated with the truth. The fusion model proposes a good compromise to improve both large- and small-scale reconstruction. It also captures detailed peaks much better than LSE, since LSE smooths these small scales out by minimizing the mean square errors.

J. Stat. Mech. (2015) 519267

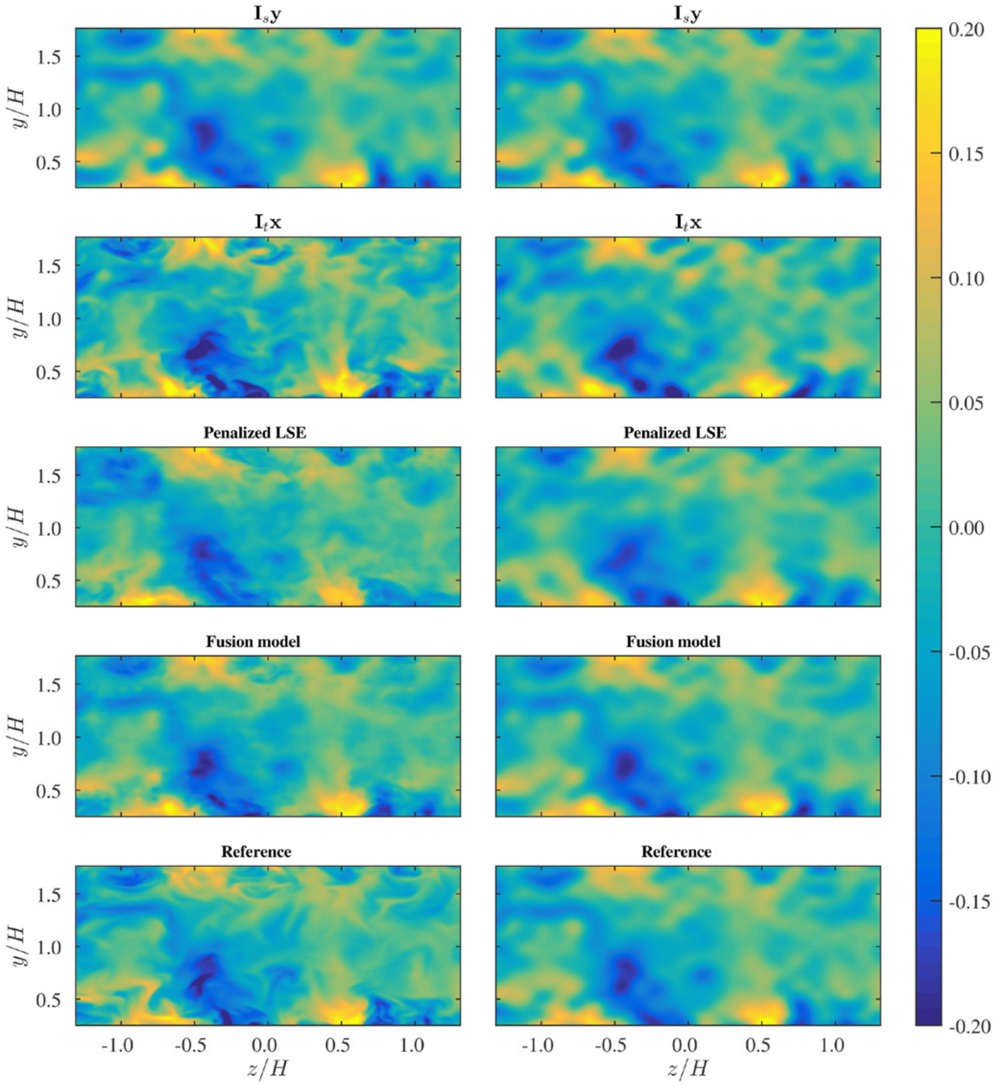

Figure 7. A sample snapshot of fluctuating streamwise velocity at one of the most difficult instants to estimate (in the middle of two LTHS time steps): reconstruction of all scales (left) and large scales only (right). The figure is best viewed on screen.

Figure 5 compares temporal spectra of the above evolutions. Time interpolation fails to estimate the signal at frequencies above a certain cutoff. LSE keeps both large and small scales, but the loss of large-scale energy is critical. This loss is highlighted in the enlarged picture of low spectral frequencies. The present model improves the estimation at both low and high frequencies.

Figure 6 shows estimates of the probability density function of time increments $u(x, t+\tau) - u(x, t)$ for the original DNS field as well as for the reconstructed field. As expected, the original field displays intermittent non-Gaussian distributions. More importantly, the reconstructed field, while less intermittent, still clearly exhibits non-Gaussian increments at small scales. Note that the reconstruction error is essentially due to the difficulty of accurately reconstructing these small scales. It is expected that any reconstruction method will lead to fields that are less intermittent than the original one.

J. Stat. Mech. (2015) 519267

Figure 7 compares snapshots reconstructed by different methods. This snapshot is at the instant most remote from its two neighboring LTHS time steps. The model reconstructs the velocity field correctly with more flow details than spatial interpolation. It also recovers large scales better than LSE and time interpolation methods.

4. Conclusions

This work proposes a Bayesian fusion model using a MAP estimate to reconstruct high-resolution velocities of a turbulent channel flow from low-resolution measurements in space and time. It searches for the most probable field given available measurements. This approach yields a simple but efficient weighted average formula in equation (23). Weighting coefficients are learnt from measurements and encode the physics of the flow. The informed fusion of information from available measurements improves the interpolation of large scales and recovers details at small scales.

Numerical experiments using a DNS database of a turbulent wall-bounded flow at a moderate Reynolds number illustrate the efficiency and robustness of the proposed method. Low-resolution measurements are extracted to learn model parameters, while original data are used as the ground truth to estimate reconstruction errors. The model is tested in various cases with different subsampling ratios. Results are compared to more standard methods such as cubic spline interpolation and penalized LSE. Bayesian fusion always produces the most accurate reconstruction. The best results are obtained when missing spatial and temporal information are of the same order of magnitude. In these cases, it provides a better large-scale reconstruction while a certain amount of small-scale detail is also recovered. The search for an even more accurate fusion and super-resolution method is the subject of ongoing work.

AQ2

References

[1] Coudert S *et al* 2011 *Exp. Fluids* **50** 1–12
[2] Adrian R J 1977 On the role of conditional averages in turbulence theory *Turbul. Liquids* **1** 323–32
[3] Adrian R J 1979 *Phys. Fluids* **22** 2065–70
[4] Guezennec Y G 1989 *Phys. Fluids* A **1** 1054–60
[5] Adrian R J 1992 Stochastic estimation of conditional structure: a review *Appl. Sci. Res.* **53** 291–303
[6] Ewing D and Citriniti J H 1999 Examination of a lse/pod complementary technique using single and multi-time information in the axisymmetric shear layer *IUTAM Symp. on Simulation and Identification of Organized Structures in Flows, Fluid Mechanics and Its Applications* vol 52 (Berlin: Springer) pp 375–84
[7] Mokhasi P, Rempfer D and Kandala S 2009 *Physica* D **238** 290–308
[8] Durgesh V and Naughton J W 2010 *Exp. Fluids* **49** 571–83
[9] Nguyen T D, Wells J C, Mokhasi P and Rempfer D 2010 *Meas. Sci. Technol.* **21** 115406
[10] Meyer R D and Tan G 2014 *Energy Build.* **73** 59–68
[11] Bonnet J P, Cole D R, Delville J, Glauser M N and Ukeiley L S 1994 *Exp. Fluids* **17** 307–14
[12] Melnick M B and Thurow B S 2012 Experimental investigation of a turbulent boundary layer using simultaneous 3-d flow visualization and 2d piv *50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition* pp 7–52
[13] Tu J H, Griffin J, Hart A, Rowley C W, Cattafesta L N and Ukeiley L S 2013 *Exp. Fluids* **54** 1–20
[14] Moser R D 1990 Statistical analysis of near-wall structures in turbulent channel flow *Technical Report* NASA Ames Research Center
[15] Kay S M 1993 *Fundamentals of Statistical Signal Processing: Estimation Theory* (Upper Saddle River, NJ: Prentice-Hall)

J. Stat. Mech. (2015) 519267

[16] Hardie R C, Eismann M T and Wilson G L 2004 *IEEE Trans. Image Process.* **13** 1174–84
[17] Trees H L V 1967 *Detection, Estimation, and Modulation Theory, Optimum Array Processing* vol 1 (New York: Wiley)
[18] Sage A P and Melsa J L 1971 Estimation theory with applications to communications and control *Technical Report* DTIC Document
[19] Levitan E and Herman G T 1987 *IEEE Trans. Med. Imaging* **6** 185–92
[20] Ma B, Lakshmanan S and Hero A O 2000 *IEEE Trans. Intell. Transp. Syst.* **1** 135–47
[21] Challa S and Koks D 2004 *Sadhana—Acad. Proc. Eng. Sci.* **29** 145–74
[22] Koks D and Challa S 2003 An introduction to Bayesian and Dempster–Shafer data fusion *Technical Report* DSTO Systems Sciences Laboratory
[23] Durrant-Whyte H and Henderson T C 2008 Multisensor data fusion *Springer Handbook of Robotics* (Berlin: Springer) pp 585–610
[24] Zhang Y, Backer S D and Scheunders P 2009 *IEEE Trans. Geosci. Remote Sens.* **47** 3834–43
[25] Joshi M and Jalobeanu A 2010 *IEEE Trans. Geosci. Remote Sens.* **48** 1245–55
[26] Unser M 1999 *IEEE Signal Process. Mag.* **16** 22–38
[27] Thevenaz P, Blu T and Unser M 2000 *IEEE Trans. Med. Imaging* **19** 739–58
[28] Reinsch C H 1967 *Numer. Math.* **10** 177–183
[29] Cook E R and Peters K 1981 The smoothing spline: a new approach to standardizing forest interior tree-ring width series for dendroclimatic studies *Tree-Ring Bulletin* vol 41 (Arizona: Tree-Ring Society) pp 45–53
[30] Kendall M G, Stuart A and Ord J K(ed) 1987 *Kendall's Advanced Theory of Statistics* (New York: Oxford University Press)
[31] Bishop C M 2006 *Pattern Recognition and Machine Learning* (*Information Science and Statistics*) (New York: Springer)
[32] Gelman A, Carlin J B, Stern H S and Rubin D B 2013 *Bayesian Data Analysis* 3rd edn (London: Chapman and Hall)
[33] Marquillie M, Laval J P and Dolganov R 2008 *J. Turbul.* **9** 1–23
[34] Boor C D 1978 *A Practical Guide to Splines* (*Mathematics of Computation* vol 27) (Berlin: Springer)
[35] Hastie T, Tibshirani R and Friedman J H 2009 *The Elements of Statistical Learning* 2nd edn (New York: Springer)
[36] Cordier L, Majd B A E and Favier J 2010 *Int. J. Numer. Methods Fluids* **63** 269–96
[37] Geisser S 1993 *Predictive Inference* (London: Chapman and Hall)

QUERIES

Page 2
AQ1
Please specify the corresponding author and provide his/her email address.
Page 18
AQ2
Please check the details for any journal references that do not have a link as they may contain some incorrect information.