On combining linear stochastic estimation and proper orthogonal decomposition for flow reconstruction

Bérengère Podvin, Sylvain Nguimatsia, Jean-Marc Foucaut, Christophe Cuvier, Yann Fraigneau

Noname manuscript No.
(will be inserted by the editor)

On Combining Linear Stochastic Estimation and Proper Orthogonal Decomposition for Flow Reconstruction

Bérengère Podvin · Sylvain Nguimatsia · Jean-Marc Foucaut · Christophe Cuvier · Yann Fraigneau

Received: date / Accepted: date

Abstract We present an estimation method combining Proper Orthogonal Decomposition (POD) and Linear Stochastic Estimation (LSE). The method is based on a direct mapping of the POD amplitudes from the measurement space to the state space. The method is tested in the turbulent boundary layer for a numerical simulation as well as for experimental data. The goal is to recover the full velocity field on a fine grid from coarse measurements of a single (longitudinal) velocity component. A significant fraction of the turbulent kinetic energy for each component is captured by the estimation. A scale-by-scale analysis shows that lower-order modes corresponding to large scales are recovered accurately. Although exact reproduction is not possible at small scales, examination of the spatial and temporal content of the estimated field shows a good statistical agreement with the real field at all scales.

B. Podvin
LIMSI rue Von Neumann, Université Paris-Sud, Orsay
Tel.: +33- 1 69 85 80 69
Fax: +33- 1 69 85 80 88 E-mail: podvin@limsi.fr

S. Nguimatsia
LIMSI rue Von Neumann, Université Paris-Sud, Orsay

J.M Foucaut
Ecole Centrale Lille; F-59000, France
Université Lille FRE3723, LML Laboratoire de Mécanique de Lille, Lille, France

C Cuvier
Ecole Centrale Lille; F-59000, France
Université Lille FRE3723, LML Laboratoire de Mécanique de Lille, Lille, France

Y Fraigneau
LIMSI rue Von Neumann, Université Paris-Sud, Orsay

1 Introduction

Turbulence characterizes a wide variety of flows including industrial aerodynamics , geophysical phenomena and biomedical flows. Having access to exhaustive spatio-temporal flow information is in general impossible in a realistic context, and can still be difficult for academic configurations, despite regular increases in computational power and progress in experimental techniques. Solving the estimation question, i.e developing methods to reconstruct the missing data from limited measurements, is therefore an important issue in many situations. One possible application could be to generate data that mimics efficiently the features of turbulent flows, and could be used as a model for the simulation of complex physical processes. Another application of interest is the implementation of control strategies where it may be helpful to estimate the flow instead of resorting to intrusive measurement techniques.

Perhaps the most popular estimation technique is Linear stochastic estimation (LSE), introduced by Adrian [1]. Adrian used LSE to educe structures of turbulent flows by approximating the conditional averages of fields Y in terms of measured data events X. **We will call the variables that we wish to estimate the** *state* **variables (Y in our example)**. In the case where the state variables have Gaussian distributions, LSE can be shown to constitute an exact solution to the estimation problem [1]. Several variants of the technique have been implemented over the years. A extension that includes quadratic terms has been proposed by Brereton [2]. A formulation of the technique in the spectral space has been developed by Tinney [3].

However linear stochastic estimation problems are often ill-conditioned, and trying to invert the problem directly typically tends to amplify singular eigenvalues of small amplitude. This gives rise to "overfitting" : trying to to reproduce the present data as well as possible leads to a choice of coefficients that actually worsens the prediction. The concept of regularization may help avoid this phenomenon. The idea is to modify the objective to be minimized, typically by adding a constraint on the norm of the solution. The new objective is no longer to find the coefficients A that minimize $\|Y - AX\|^2$, but those that minimize the combination of this objective and an additional constraint

$$\|Y - AX\|^2 + \Lambda \|A\|_p^2\|$$

. The weight Λ of this new constraint, which is not directly prescribed by the physical problem to be solved, is a free parameter, while $\|_\|p$ represents a norm. When $p = 2$, the regularization is said to be of Tikhonov type [?], while $p = 1$ is the basis of the LASSO algorithm [4]. The choice of the norm and of Λ affects the solution, and there is no guarantee that it will lead to a substantially better solution [?].

Dekou et al. [5] used Tikhonov regularization in conjunction with standard LSE, *following Bonnet et al.*. **They recovered very little energy in the cross-stream components (only 10% in the wall-normal component, see figure ? of their paper)**.

Moreover, the implementation of LSE may require regularization, as was the case for Dekou [5], who used LSE in conjunction with a regularization technique to recover the three components of the velocity field on a fine grid from coarse measurements of the longitudinal velocity component. A significant contribution to the estimation problem was

made by Bonnet et al. [6], who combined LSE with Proper Orthogonal Decomposition (POD), a statistical technique introduced in fluid mechanics by Lumley [7]. The idea is to apply LSE to the flow field, then to project the LSE-estimated flow fields onto POD eigenfunctions, and to use the resulting amplitudes to reconstruct the flow. This approach has been implemented in diverse applications such as [8], [9].

In the present paper, we present an alternative way to combine POD and LSE for the estimation problem. **Unlike Bonnet et al., who seek their estimate as a direct combination of their measurements in physical space (classic LSE implementation),** *then* **reconstruct a solution from the projection of that estimate onto the POD eigenfunction basis, POD is not used as a** *a posteriori* **filter, but as an** *a priori* **decomposition space.** The central idea is to apply LSE directly in the space of POD coefficients, so that the POD amplitudes of the measurements are directly mapped to those of the field to be estimated in a one-step process. The method requires a training set, consisting of measurements and state fields obtained at the same times. The computation of the POD bases for the measurements and the states can be carried out either with the direct method or the method of snapshots. An advantage of our approach is that unlike classic estimation methods, no regularization technique is required. We first describe the method and discuss its characteristics. The method is then implemented in a turbulent boundary layer, following the study of Dekou [5]. Tests are carried out for a numerical simulation as well as for experimental data.

2 Description of the method

2.1 Proper Orthogonal Decomposition

We consider a velocity field $\underline{u}(\underline{x},t)$ defined over a physical domain D. This field constitutes the state to be estimated. We assume that a linear measurement function of this field is available, which we denote $P\underline{u}(\underline{x},t)$ where P is a linear operator acting on space variables (e.g P could be a projection operator, or it could be the combination of a projection and a translation). Applying Proper Orthogonal Decomposition to the velocity field means that there exists a denumerable set of eigenfunctions $\{\underline{\phi}^n(\underline{x})\}$ and hierarchically ordered eigenvalues λ^n, $\lambda^1 > \lambda^2 > \ldots \lambda^n > \ldots$, which are solutions of a Fredholm equation

$$\int < \underline{u}(\underline{x},t)\underline{u}(\underline{x}',t) > .\underline{\phi}^n(\underline{x}')d\mu(\underline{x}') = \lambda^n\underline{\phi}^n(\underline{x}) \tag{1}$$

where $<>$ represents a temporal average, $< \underline{u}(\underline{x},t)\underline{u}(\underline{x}',t) >$ is the spatial autocorrelation tensor at zero time lag and $d\mu$ is the integration measure. By construction, we have

$$\underline{u}(\underline{x},t) = \sum_{n=1}^{\infty} a^n(t)\underline{\phi}^n(\underline{x}), \tag{2}$$

where

$$a^n(t) = \int \underline{u}(\underline{x},t).\underline{\phi}^n(\underline{x})d\mu(\underline{x}). \tag{3}$$

If the eigenfunctions are normalized, one can show that

$$< a^n(t)a^m(t) >= \delta_{nm}\lambda^n. \tag{4}$$

If we consider the quantity $\underline{v} = P\underline{u}$ defined on a domain D' and associated with an integration measure μ', there also exists an infinite set $\{\underline{\psi}^n\}$ of denumerable eigenfunctions such that

$$P\underline{u}(\underline{x},t) = \sum_{n=1}^{\infty} b^n(t)\underline{\psi}^n(\underline{x}), \tag{5}$$

where

$$b^n(t) = \int_{D'} P\underline{u}(\underline{x},t).\underline{\psi}^n(\underline{x})d\mu'(\underline{x}). \tag{6}$$

Substituting equation (2) in equation (6) and using the linearity of P, one has

$$b^n(t) = \sum_{m=1}^{\infty} a^m(t) \int_{D'} P\underline{\phi}^m(\underline{x}).\underline{\psi}^n(\underline{x})d\mu'(\underline{x}). \tag{7}$$

The linear relationship between the POD coefficients $a^n(t)$ and the POD coefficients of the measurements $b^n(t)$ depends on the projection of the POD eigenfunctions $\underline{\phi}$ onto the measured eigenfunctions $\underline{\psi}$. Formal inversion of equation (7) provides a way to estimate the field $\underline{u}$ from the measurements v. We note that so far no approximation has been introduced in the derivation.

2.2 Method of Snapshots

In the method of snapshots [10], the spatial autocorrelation tensor $< \underline{u}(\underline{x},t)\underline{u}(\underline{x}',t) >$ is replaced in the Fredholm equation with its numerical evaluation from a set of N realizations:

$$< \underline{u}(\underline{x},t)\underline{u}(\underline{x}',t) > \leftarrow \frac{1}{N}\sum_{n=1}^{N} \underline{u}(\underline{x},t^n)\underline{u}(\underline{x}',t^n) \tag{8}$$

It is important to realize, as has been pointed by George [11], that although the term on the right-hand-side converges to the term on the left-hand-side when N goes to infinity, the solutions of the corresponding Fredholm equations may be different at finite N.

We are looking for a finite set of eigenfunctions $\underline{\phi}^n$ that represents in an optimal way the N-dimensional set of realizations $\underline{u}(\underline{x},t^n)$. These eigenfunctions $\underline{\phi}^n$ are expressed as a linear combination of the realizations $\underline{u}(\underline{x},t^m)$ (note that the problem remains continuous in space):

$$\underline{\phi}^n(\underline{x}) = A_{nm}\underline{u}(\underline{x},t^m) \tag{9}$$

where $A_{nm} = a^n(t_m)$, and

$$\underline{u}(\underline{x},t_m) = \sum_{n=1}^{N} a^n(t_m)\underline{\phi}^n(\underline{x}). \tag{10}$$

If we denote D the diagonal matrix containing the square root of the POD eigenvalues $D_{nn} = \sqrt{\lambda^n}$, the matrix AD^{-1} is a unitary matrix i.e $(AD^{-1})(AD^{-1})^T = I$ where I is the identity matrix and T represents the transposition operator.

The representativity or adequacy of the POD eigenfunction basis derived in this fashion can be measured by the error e between a field $\underline{u}(\underline{x},t*)$ which does not belong to the initial set of snapshots and its projection onto the set of eigenfunctions $\underline{\phi}^n$:

$$e(t^*) = \int (\underline{u}(\underline{x},t^*) - \sum_{n=1}^{N} a^n(t^*)\underline{\phi}^n(\underline{x}))^2 d\mu(\underline{x}) \tag{11}$$

where

$$a^n(t^*) = \int \underline{u}(\underline{x}, t*).\underline{\phi}^n(\underline{x}) d\mu(\underline{x}). \tag{12}$$

By construction, the error e is zero if the field belongs to the initial set of snapshots. It is not zero for a field that does not belong to that set.

2.3 Reconstruction

2.3.1 Principle

The idea of our reconstruction method is to carry out estimation directly in the space of POD coefficients. We use the same set of N snapshots to compute two POD bases associated with realizations $\underline{u}$ and measurements $\underline{v} = P\underline{u}$. For each snapshot, we have

$$\underline{u}(\underline{x}, t_i) = \sum_{n=1}^{N} a^n(t_i)\underline{\phi}^n(\underline{x}) \tag{13}$$

and

$$\underline{v}(\underline{x}, t_i) = \sum_{n=1}^{N} b^n(t_i)\underline{\psi}^n(\underline{x}). \tag{14}$$

We note that the expression is exact for the N snapshots if the N modes are used for the reconstruction.

If we use matrix notations, and denote U the snapshot matrix where the i-th column is the i-th snapshot, $\underline{\Phi}$ the eigenfunction matrix where the i-th column is the i-th POD eigenfunction, and A the temporal coefficient matrix defined in equation (9), one has

$$U = \underline{\Phi} A. \tag{15}$$

Similarly, for the measurements $V(\underline{v})$, the associated eigenfunctions $\underline{\Psi}(\underline{\psi}^n)$ and corresponding temporal coefficients B (b), one has

$$V = \underline{\Psi} B. \tag{16}$$

A and B can be made unitary if they are renormalized with the eigenvalue matrices D_U and D_V which contain respectively the square root of the POD eigenvalues associated with $\underline{u}$ and $\underline{v}$:

$$\tilde{A} = AD_U^{-1} = D_U^{-1}A$$

and

$$\tilde{B} = BD_V^{-1} = D_V^{-1}B.$$

There is an exact linear mapping from $\tilde{A}$ to $\tilde{B}$, given by

$$\tilde{A} = \tilde{M}\tilde{B} \tag{17}$$

where

$$\tilde{M} = \tilde{A}\tilde{B}^T = D_U^{-1}A(D_V^{-1}B)^T = D_U^{-1}AB^TD_V^{-1} \tag{18}$$

We note that since matrices are unitary, inversion is simply transposition, which limits numerical errors. The corresponding mapping M from to A to B such that $A = MB$ is therefore given by

$$M = D_U \tilde{M} D_V^{-1}. \tag{19}$$

2.3.2 Remarks

- The snapshots that are used to determine the POD basis constitute the training set.
- An important remark is that $\tilde{M}$ is unitary, and therefore the rescaled norm of the state estimate $\|D_U^{-1}A\|$ is equal to the rescaled norm $\|D_V^{-1}B\|$. This provides some guarantee on the energetic content of the estimation relative to that of the measurements.
- The mapping can be defined even if the snapshot method is not used. In that case, A and/or B are obtained by projecting the snapshots constituting the training set onto the POD eigenfunctions obtained by the direct method.
- Equation (19) shows that large values are expected in the far-right columns of M, since the measurement POD spectrum D_V^2 is expected to decay faster than the full state spectrum D_U^2 as the dimension of the measurements is typically smaller than the state. This means that small errors in the high-order POD **coefficients** of the measurements may be amplified. To limit this effect, we will in practice truncate the matrix M i.e set to zero the contribution of higher-order measurement modes.
- The connection with linear stochastic estimation is direct since the goal of linear stochastic estimation is to estimate an unknown quantity q' given the knowledge of another quantity q as a linear function of that quantity q ([1]):

$$E[q'|q] = \mathcal{A}q \tag{20}$$

  Comparing equations (21), (19) and (20), one can see that this is equivalent to performing the linear stochastic estimation directly in POD space.

2.3.3 Algorithm

Given instantaneous measurements $\underline{v}(\underline{x}, t^*)$, and POD bases $\underline{\phi}^n$ and $\underline{\psi}^n$, a reconstruction of the field $\underline{u}^{est}$ can be given by

$$\underline{u}^{est}(\underline{x}, t^*) = \sum_{n=1}^{N} a_{est}^n(t*)\underline{\phi}^n(\underline{x}) \tag{21}$$

where

$$a_{est}^n(t*) = M_{nm} b^m(t*) \tag{22}$$

and

$$b^n(t*) = \int_{D'} \underline{v}(\underline{x}, t*).\underline{\psi}^n(\underline{x}) d\mu'(\underline{x}). \tag{23}$$

The algorithm for linear stochastic estimation is therefore as follows:

1. construct for a same set of snapshots the POD bases $\underline{\Phi}^n$ and $\underline{\Psi}^n$, the eigenvalue matrices D_u and D_v, the coefficients matrices A and B corresponding respectively to the state field $\underline{u}$ and the measurements $\underline{v}$, and compute the matrix M.
2. for each measurement $\underline{v}(\underline{x}, t*)$, compute the coefficients $b^n(t*)$, and obtain an estimate a^n_{est}(t*) for the POD coefficients of $\underline{u}$
3. reconstruct the field $\underline{u}_{est}(\underline{x}, t*)$ using $a^n_{est}(t*)$ and $\underline{\Phi}^n(\underline{x})$.

By construction, if the field is one of the snapshots used to construct the POD database and if all POD modes are included, the estimated field should coincide with the projected field and in the case of the method of snapshots, the full field. However, if a smaller number of modes is retained for M (see remark above), there may be some differences even if the snapshot belongs to the database.

2.3.4 Evaluation of the procedure

To evaluate the method, it will be of interest to consider three types of fields:

- the full field $\underline{u}(\underline{x}, t)$
- the projected field $\underline{u}_{projected}(\underline{x}, t) = \sum_n a^n(t)\underline{\phi}^n(\underline{x})$ where $a^n(t)$ is the true POD **coefficient** obtained by projection of the full field onto the basis of eigenfunctions,
- the estimated field $\underline{u}_{estimated}(\underline{x}, t) = \sum_n a^n_{est}(t)\underline{\phi}^n(\underline{x})$ where $a^n_{est}(t)$ is the estimated POD **coefficient**.

The difference between the full field and the projected field characterizes the adequacy of the truncated POD basis. The difference between the projected and the estimated field characterizes the error committed on the amplitudes and therefore that of the estimation method.

3 Application of the procedure to the turbulent boundary layer

The method is tested for the reconstruction of wall turbulence, following the study of Dekou [5]. The measurements ($\underline{v}$) consist of the streamwise velocity component on a coarse grid in a cross-section plane. The state to estimate ($\underline{u}$) consists of the full (three-component) velocity field on the plane with a fine resolution. One part of the flow realizations is used to construct the POD bases - training set. Another part of the realizations constitutes the validation set - on which the method is evaluated. Two cases, corresponding to numerical simulation and experiment, are considered. In all cases we checked that application of the method to a snapshot that is used to construct the basis yields an estimated field which is nearly identical to the true field and the projected field (see also discussion below).

3.1 Numerical Test Case

3.1.1 Setting

The first case consists of the numerical simulation of a turbulent channel flow at $Re_\tau = 960$ where Re_τ is the Reynolds number based on the friction velocity and channel half-height. The simulation has been described in [12]. The size of the numerical domain is $(L_x, L_y, L_z) = (2\pi, 2, \pi)h$ where x, y, z refer to the streamwise, wall-normal and spanwise directions respectively. Periodic boundary conditions are used in the horizontal directions x and z. It is resolved with 512 grid points in each direction with a regular spacing in the horizontal directions **and a hyperbolic tangent stretching function for the mesh in the wall-normal direction**. **The resolution is about 6 (12) wall units in the spanwise (streamwise) direction and the first point in the wall-normal direction is at** $y_+ = 0.3$. The domain for measurements and the estimate consists of a single cross-section plane at $x = 0$ (an arbitrary location, since boundary conditions are periodic in the streamwise direction). The measurement grid sizes in the wall-normal and the spanwise direction are respectively five and ten times lower than the simulation resolution. Our goal is to reconstruct the three-dimensional field on the full plane grid. 1200 realizations spanning a total time of $3h/u_\tau$ are used to build the POD basis, corresponding to several dozens of eddy turnover times.

Results shown below were obtained for a validation set of **360** fields. We checked that results were robust with respect to the definition of the validation set. **The validation fields are obtained by integration of the numerical simulation at times beyond those corresponding to the training set.** The configuration with the two different grid sizes is illustrated in figure 1 and the grid specifics are given in table 1.

3.1.2 Results

The method of snapshots is used for both the state and the measurements. 800 modes out of 1200, capturing more than 96% of the total fluctuating kinetic energy, were retained for the estimation matrix. Figure 2 compares the relative fraction of energy associated with both POD decompositions. As expected, one can see that the three-dimensional signal on the fine grid converges much less rapidly than its coarse one-dimensional restriction, indicating that a substantial amount of information will not be directly accessible through the measurements.

Figure 3 compares the turbulent intensities of each velocity component associated with the full field, the projected field and the estimated field over the half-channel. The top row shows results for the training set (i.e the snapshots used to compute the POD basis). By construction the estimated field should be exactly equal to the projected field and to the full field if all modes (1200) were included. Since only 800 modes are kept, the projected field lacks some energy in the cross-stream components and there is a very slight discrepancy between the spanwise components of the projected and estimated fields. However the agreement is very close.

Results on the bottom row of figure 3 correspond to the validation set. We can see that the discrepancy between the full field and the projected field is larger, which indicates that the validation fields cannot be entirely described by the POD basis. However the statistics of the estimated field remain in relatively good agreement with those of

the projected field. The estimation is actually slightly more energetic than the projection. We note that there is no constraint on the estimation, so the energy of the estimated POD modes can be either larger or smaller than those of the true POD modes. Despite small differences, a good agreement between the projection and the estimation is observed.

This is confirmed by the direct comparison of the true and estimated POD coefficients for the first six modes in figure 4. A mode-by-mode analysis, shown in figure 5 shows that the most energetic, lower-order modes are relatively well estimated as their energy level is correct and they are well correlated with the true POD modes. However, for higher-order modes $n \geq 100$, the correlation coefficient drops and the energy level tends to be overestimated.

Figure 6 compares the one-dimensional spectrum in the spanwise direction of the different velocity components for the estimated and the full fields at a height $y+ = 140$, which corresponds to the top portion of the wall layer [13]. Averaging was performed in time. The agreement is quite satisfactory. This confirms that the estimated velocity fields - and in particular the cross-stream components, about which no information is directly available - are characterized by an appropriate spectral content.

3.2 Experimental Test Case

3.2.1 Setting

The second test case consists of experimental measurements obtained at LML, in the context of the WALLTURB experiment [14]. The goal of the WALLTURB project was to analyze and to model wall turbulence. In the paper we present results corrresponding to measurements obtained in a cross-section of a turbulent boundary layer generated by a flow with an incoming speed of 5 m/s. We also tested the procedure for another database obtained at U=10m/s. Results were found to be very similar and only the case $U = 5$ m/s is presented in the text.

The boundary layer height at the location of the measurements **is defined as the height at which the mean velocity reaches 99% of its free-stream value and is** 0.3 m. The Reynolds number based on the boundary layer height and the friction velocity is $R_\tau = 3760$. The data consists of hot-wire (HW) measurements of the streamwise velocity field acquired at a frequency of 30 **kHz**, partially synchronized with 4Hz PIV measurements of the 3D velocity field (see figure 7). Both HW and PIV measurements are taken on a cross-section plane, but the PIV plane is separated from the hot-wire plane by a distance of 1 cm, which represents about 100 wall units in the streamwise direction. The hot wire rake consists of 143 sensors (13 x 11 grid) with an uneven spacing (shown in figure 7). The PIV resolution field, also shown in figure 7, is 2 mm, which corresponds to a resolution increase factor of more than 10 in each direction. 64 uncorrelated blocks of data each spanning a duration of 6 s were used for the present study. For the upstream speed of 5m/s, each block contains 16 PIV velocity fields and 180 000 HW data points. To build the synchronised POD database for the hot-wire and the PIV measurements the hot-wire measurements were averaged over the duration of each PIV field acquisition. We used the signals corresponding to the 15 first PIV fields of each

block, while the last PIV field of each block was retained for the validation set. The results were not found to be dependent on that particular definition of the training and the validation sets.

In contrast with the numerical simulation results, the number of POD modes associated with the measurements is limited by the grid size (13 x 11), and not by the number of snapshots available. This means that the direct method was used to compute the POD basis associated with the hot-wire measurements. The POD amplitudes were then computed by projection of the field onto the hot-wire (HW) POD basis. Figure 8 a) illustrates that the information contained in the measurements is indeed limited, while only 74% of the fluctuating kinetic energy of the 3D PIV field is captured by 140 POD modes.

We checked that the projected field was identical to the full field when the full field was part of the POD database. Since only 140 modes are used in the estimation, the estimated field cannot match exactly the full field, even if it is part of the POD database. Results for the validation set, consisting of 64 uncorrelated snapshots, are shown in the remainder of the paper, for a truncation of 140 modes. First, an estimation of the error was made for each mode. Figure 8 b) shows that the energy levels of the estimated amplitudes are slightly lower than that of the real amplitudes, but exhibit the same asymptotic trend.

Figure 9 compares the turbulent intensities for each velocity component of the full field, the field projected on the truncation, and the field estimated from the hot-wire measurements. Again, one can see that the projected field contains less energy than the full field, which is primarily an effect of the limited truncation, but could also be due in part to the representativity of the POD basis. The estimated field is still less energetic, but table 2 shows that both projected and estimated fields nevertheless capture a substantial fraction of the full field for all velocity components. The wall-normal component is the least well recovered, which reflects the fact that it is dominated by small scales. We note that the estimated energy is much larger than what was obtained with classic LSE [5], in particular for the cross-stream components.

Figure 10 compares at a given instant in time the instantaneous measurement (top row) with the corresponding full (PIV), projected and estimated fields (respectively in the second, third and fourth rows). We note that the estimation represents a considerable improvement for the streamwise velocity component (left column, compare rows 1 and 4 with row 2). For the wall-normal velocity (middle column), we can see that both estimated and projected fields fail to capture the energetic small-scale component, but correctly recover the large-scale motion, and in particular the ejection at the top of the layer. The spanwise component of the estimation is also close to the projection and captures adequately the large-scale structure of the flow.

4 Conclusion

We have presented a method to recover information about turbulent flows from partial measurements. The key idea of the method is to apply Linear Stochastic Estimation in POD space, i.e to determine the POD amplitudes of the field to be estimated as a combination of the POD amplitudes of the measurements. The POD eigenfunctions for the measurements and the **variables that we wish to determine** can be derived either with the classical method or

the method of snapshots. A training set consisting of joint POD amplitudes for the states and the measurements is used to determine the mapping between the two POD decompositions. The mapping can be expressed as a simple matrix product, which does not require matrix inversion or the implementation of regularization techniques. The method allows a scale-by-scale analysis of the estimation error, and provides an assessment of the representativity of the training set.

The method was tested for the estimation of the full velocity field in a cross-section of a turbulent boundary layer using coarse, one-component velocity measurements. Evaluation of the method was carried out in a numerical context, as well as for experimental data. In both cases a large fraction of the energetic content of the different velocity components was recovered. The large-scale structure of the flow was shown to be accurately estimated. Although accurate, fine-scale reproduction of the real flow was not possible, The temporal and spatial statistics of the estimated fields were found to be in good agreement with their real counterparts. **The combined LSE/POD procedure therefore constitutes an improvement over the standard implementation of linear stochastic estimation, and could therefore be of use for a variety of flows to recover missing data due to insufficient sampling in space and/or time.**

References

1. R.J. Adrian and P. Moin. Stochastic estimation of organized turbulent structure : homogeneous shear flow. *J. Fluid Mech.*, 190:531–559, 1988.
2. G.J. Brereton. Stochastic estimation as a statistical tool for approximating turbulent conditional averages. *Physics of FLuids A*, 4(3):2046–2059, 1992.
3. C.E. Tinney, F. Coiffet, J. Delville, A.M. Hall, P. Jordan, and M.N. Glauser. On spectral linear stochastic estimation. *Experiments in Fluids*, 41:763–775, 2007.
4. R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.
5. R. Dekou, J.M. Foucaut, S. Roux, and J. Delville. Large-scale organization of a near wall turbulent flow. *International Journal of Heat and Fluid Flow*, 61:12–20, 2016.
6. J.P. Bonnet, D.R. Cole, J. Delville, M.N. Glauser, and L.S. Ukeiley. Stochastic estimation and proper orthogonal decomposition: Complementary techniques for identifying structure. *Exp. Fluids*, 17:307–314, 1994.
7. J.L. Lumley. The structure of inhomogeneous turbulent flows. In A.M Iaglom and V.I Tatarski, editors, *Atmospheric Turbulence and Radio Wave Propagation*, pages 221–227. Nauka, Moscow, 1967.
8. J. Taylor and M. Glauser. Toward practical flow sensing and control via pod and lse-based low-dimensional tools. *ASME 2002 Fluids Engineering Division Summer Meeting, Montreal, Quebec, 14-18 July*, 2002.
9. V. Durghesh and J. W. Nauton. Multi-time delay lse-pod complementary approach applied to unsteady high-reynolds-nuber near-wake flow. *Experiments in FLuids*, 49(3):571–583, 2010.
10. L. Sirovich. Turbulence and the dynamics of coherent structures part i: Coherent structures. *Quart. Appl. Math.*, 45(3):561–571, 1987.
11. W.K. George. A 50-year retrospective and the future. In L. Danaila A. Pollard, L. Castillo and M. Glauser, editors, *Whither Turbulence and Big Data in the 21st century?*, Berlin, Germany, 2017. Springer.
12. B. Podvin and Y. Fraigneau. A few thoughts on proper orthogonal decomposition in turbulence. *Physics of FLuids*, 29:020709, 2017.
13. M. Stanislas, L. Perret, and J.M. Foucaut. Vortical structures in the turbulent boundary layer: a possible route to a universal representation. *J. Fluid Mech.*, 602:327–382, 2008.
14. *Progress in Wall Turbulence: Understanding and Modelling*, Ercoftac Series, Berlin, Germany, 2009. Springer.

Grid parameters	N_z	N_y	$\Delta y_{min}+$	$\Delta y_{max}+$
State (numerical simulation)	512	512	0.28	9
Measurements	51	102	1.5	55

Table 1 Grid parameters for the numerical simulation.

$rms/rms_{fullfield}$	u	v	w
Projected field	0.87	0.67	0.84
Estimated field	0.77	0.50	0.75

Table 2 Ratio of the total turbulent intensity rms of the fields integrated over the entire boundary layer relative to the full field.

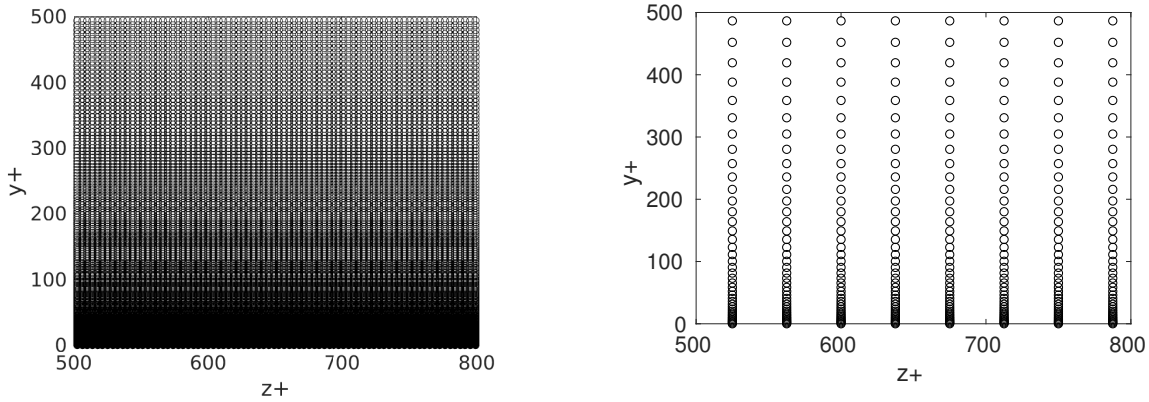

Fig. 1 Numerical grid configuration (detail): left) fine grid for state estimation right) coarse grid for measurements.

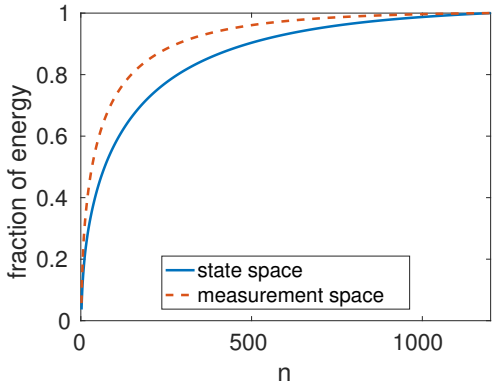

Fig. 2 Relative fraction of fluctuating energy captured by the POD modes $\frac{\sum_2^n \lambda^n}{\sum_2^N \lambda^N}$ corresponding to the measurement and the state spaces.

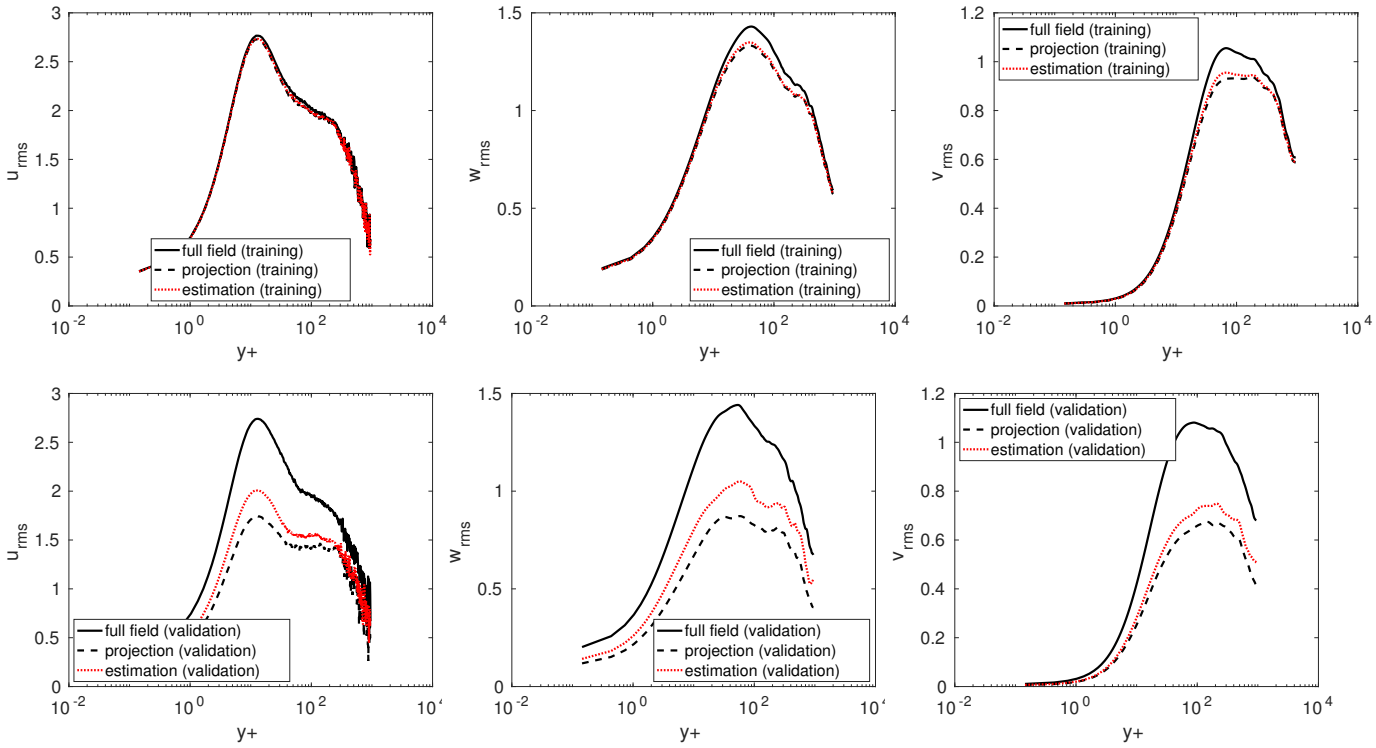

Fig. 3 Turbulent velocity rms for the full field, the projected field , the estimated field - 800 modes are retained for the projection and the estimation; from left to right: streamwise velocity, wall-normal velocity and spanwise velocity; Top row: training set; Bottom row: validation set.

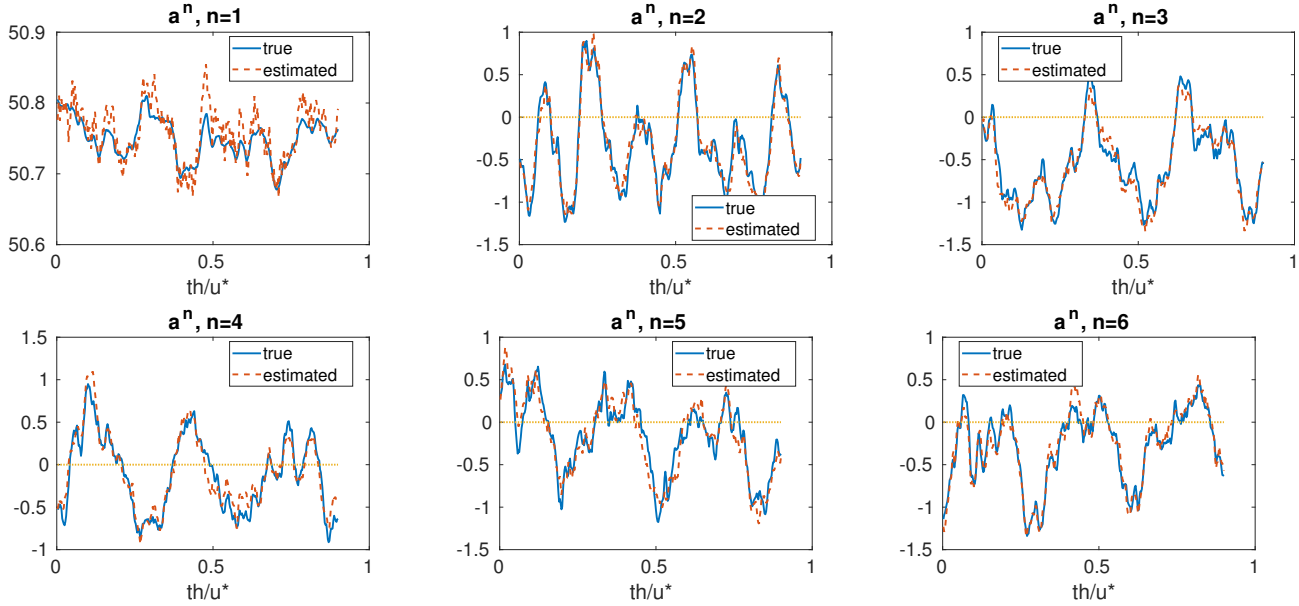

Fig. 4 Comparison for the validation set of the true a^n and estimated POD coefficients a^n_{est} for the first six modes. 800 modes are retained for the estimation; top row: from left to right: $n = 1, 2, 3$; bottom row: from left to right: $n = 4, 5, 6$

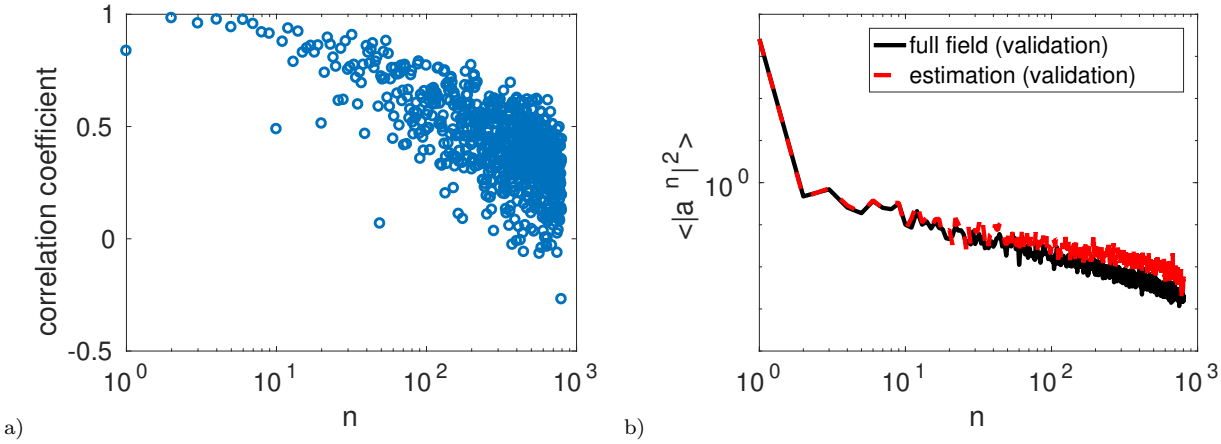

Fig. 5 a) Correlation coefficient between the real and the estimated POD coefficient; b) Energy level $< |a^n|^2 >$ for the real and estimated POD coefficients. Results are shown for the validation set only.

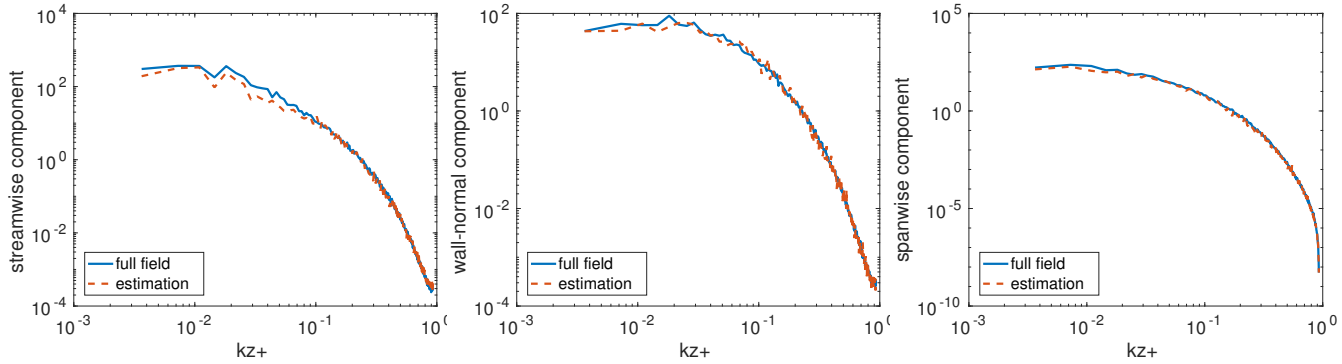

Fig. 6 Comparison of Fourier spectra in the spanwise direction for the full and the estimated fields at $y+ = 140$. Averages are taken in time. From left to right: streamwise (u), wall-normal (v) and spanwise (w) velocity component.

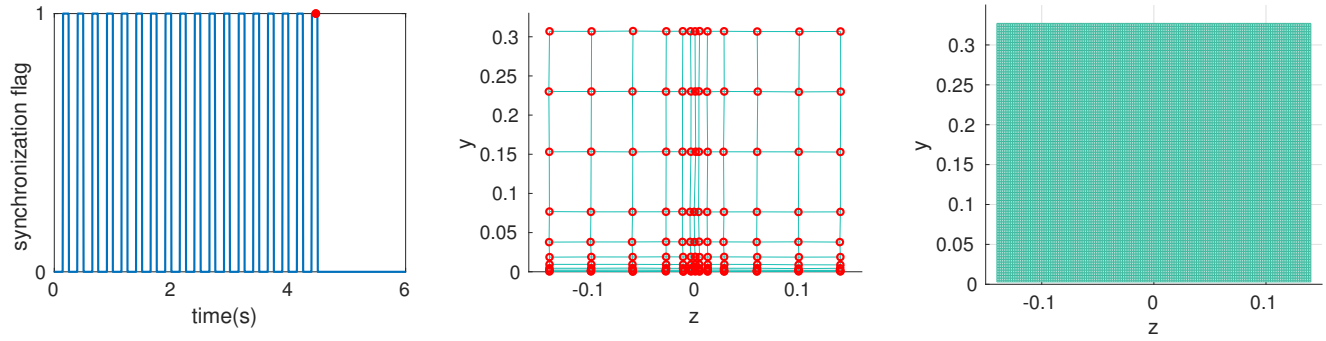

Fig. 7 Experimental set-up: left) synchronization flag indicating when the hot-wire measurements coincide with the PIV acquisition times over one block of data; the circle indicates the PIV field that was used for the validation set (the hot-wire data was averaged over the PIV acquisition duration). middle) coarse grid corresponding to the hot-wire (HW) measurements; right) fine grid corresponding to the PIV measurements for state estimation. Units are given in m. The PIV and the HW planes are separated by a positive streamwise distance of 0.01 m.

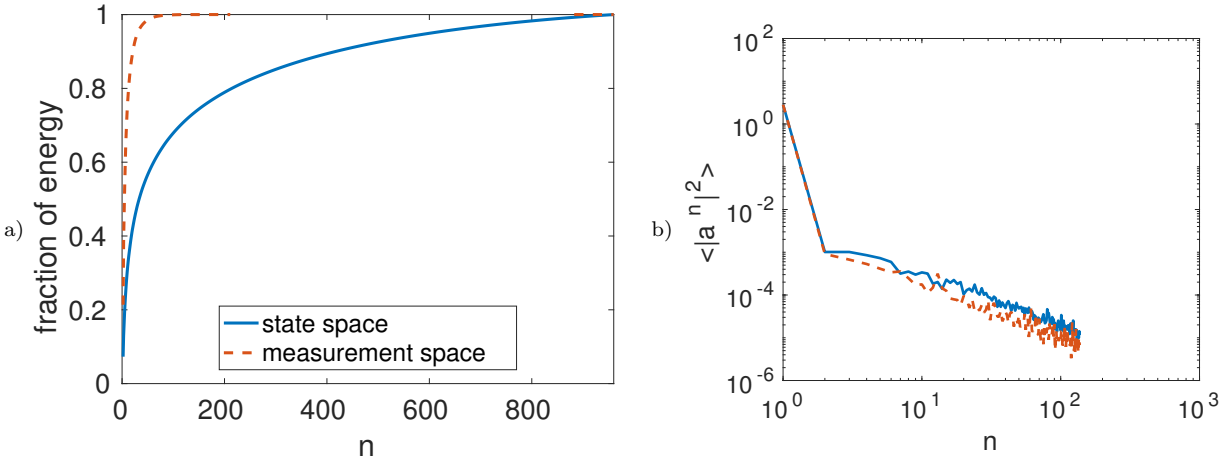

Fig. 8 a) Relative fraction of fluctuating energy for the state (PIV) and measurement (HW) space defined as in figure (2) b) Energy level for the true and estimated POD coefficients $< |a^n|^2 >$.

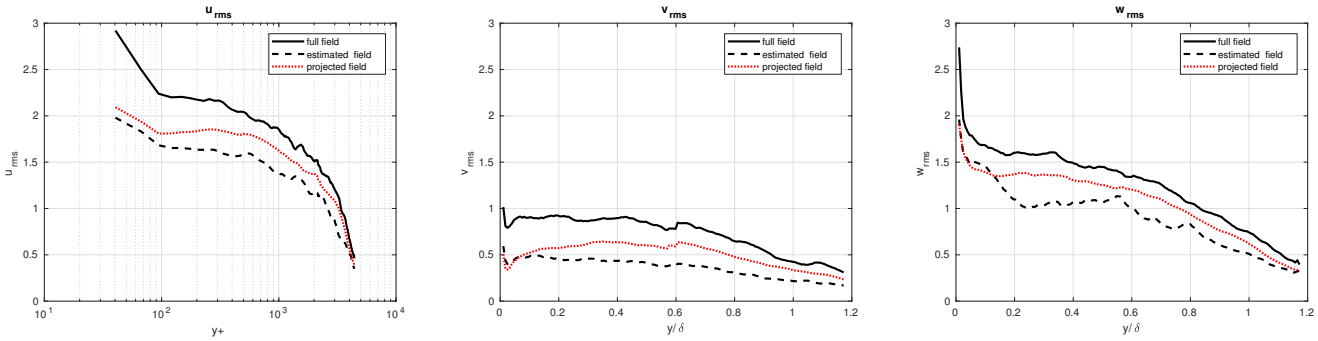

Fig. 9 Turbulent velocity r.m.s for the full field, the projected field , the estimated field - 140 modes are retained for the projection and the estimation; from left to right: streamwise velocity, wall-normal velocity and spanwise velocity; Results are shown only for the validation set.

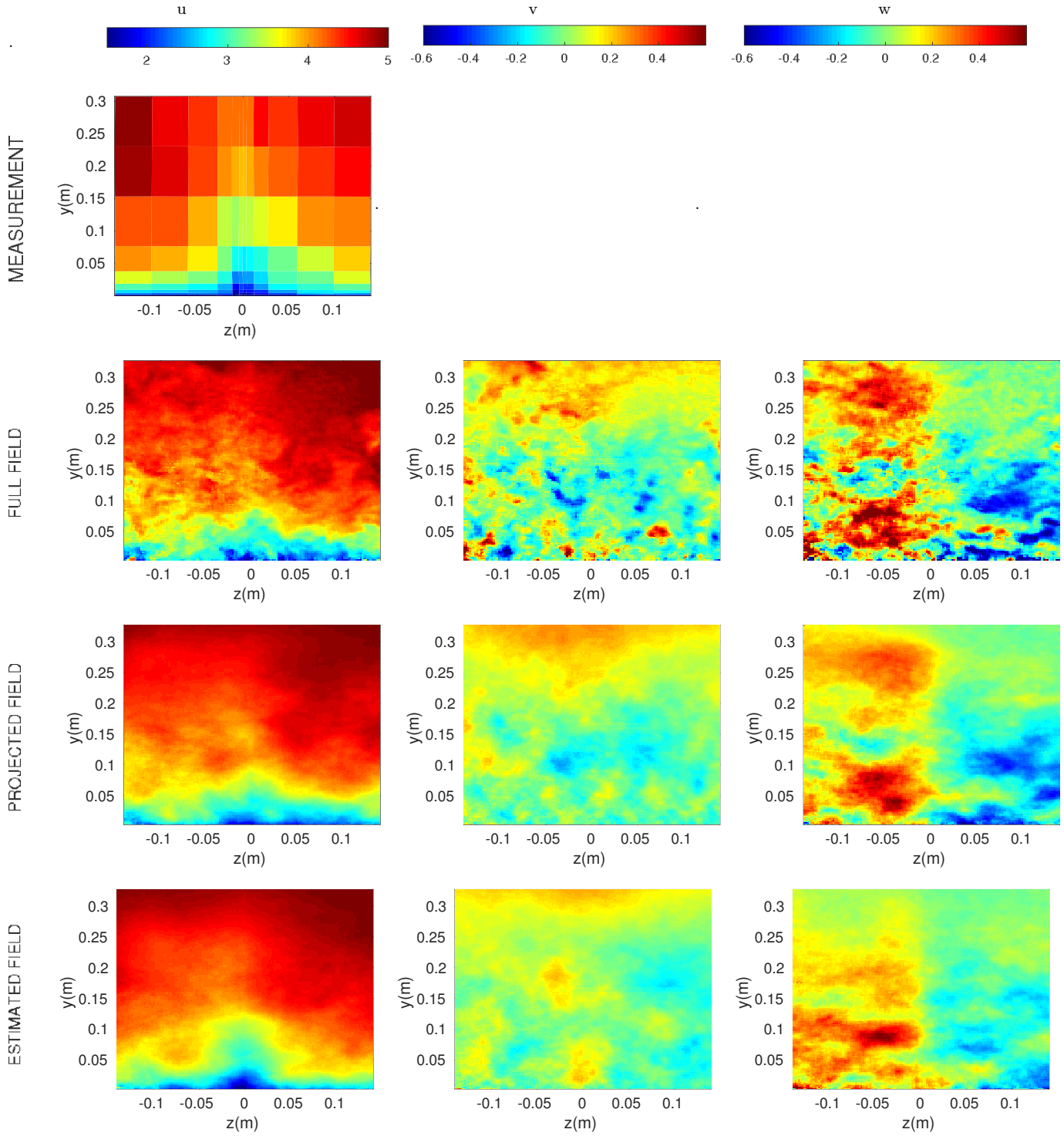

Fig. 10 Instantaneous field at a given instant. From left to right: streamwise, wall-normal and spanwise velocity component. The top row represents the field corresponding to the hot-wire measurements of the longitudinal component. The second row represents the corresponding PIV field at this instant. The third row represents the projected field and the fourth row the estimated field. Both projection and estimation are computed with 140 modes. Velocity units are given in m/s.