



Large-eddy simulation-based reconstruction of turbulence in a neutral boundary layer using spectral-tensor regularization

Ahmed Alreweny^{1,†}, Stefan Vandewalle² and Johan Meyers¹

¹Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300A, B3001 Leuven, Belgium

²Department of Computer Science, KU Leuven, Celestijnenlaan 200A, B3001 Leuven, Belgium

(Received 29 June 2023; revised 6 December 2023; accepted 19 January 2024)

We propose an efficient method to reconstruct the turbulent flow field in a neutrally stratified atmospheric boundary layer using large-eddy simulation (LES) and a series of lidar measurements. The reconstruction is formulated as a strong four-dimensional variational data assimilation problem, which involves optimizing two competing terms that contribute in the objective functional. The first term is a likelihood term, while the second contains the initial background distribution of turbulent velocity fluctuations and works as a regularization term. However, computing and storing the full background covariance tensor in turbulent flows is time consuming and resource intensive. In the current work, we investigate the possibility of replacing the complex background tensor by simple analytical approximations based on spectral tensors such as the Hunt–Graham–Wilson (HGW) model (*Boundary-Layer Meteorol.*, vol. 85, 1997, pp. 35–52) or the Mann model (*J. Fluid Mech.*, vol. 273, 1994, pp. 141–168). Afterwards, the problem is solved using a quasi-Newton algorithm and preconditioned to enhance the convergence rate. We test the method using virtual lidar measurements collected on a fine reference LES. Results show a super-linear convergence rate of the optimization algorithm to a local minimum and very good agreement between virtual lidar measurements and reconstruction in the scanning region. Furthermore, we demonstrate that incorporating the Saffman energy spectrum ($E(k) \sim k^2$ where E is the energy spectrum and k is the magnitude of the wavenumber vector) at low wavenumbers into the Mann spectral tensor yields a longer streamwise correlation length, resulting in reduced reconstruction error when compared with the Batchelor spectrum ($E(k) \sim k^4$). Finally, we observe that using the HGW model or Mann model with a Saffman spectrum yields similar results.

Key words: turbulent boundary layers, atmospheric flows, homogeneous turbulence

† Email address for correspondence: ahmed.alreweny@kuleuven.be

1. Introduction

Knowledge of the turbulent wind field in the atmospheric boundary layer (ABL) has great value in many fields, including wind energy (Iungo, Wu & Porté-Agel 2013), weather forecasting (Sun, Flicker & Lilly 1991), pollution dispersion (Nguyen & Soulhac 2021) and wind loading of structures (Guo *et al.* 2022), among others. Over the past decade, pulsed light detection and ranging (lidar) has demonstrated the ability to provide instantaneous measurements of the turbulent wind field over a large area that can extend to several kilometres (Peña *et al.* 2013). Despite this significant advance in measurement techniques, the full three-dimensional reconstruction of the turbulent flow in the ABL remains underdetermined, meaning that the degrees of freedom fixed by the measurements are much lower than the total degrees of freedom in a chosen domain. One simple yet crude approach to reconstructing the velocity field from raw measurements is to apply the static flow assumption, which involves disregarding the temporal evolution of the flow among other spatial assumptions (e.g. the dimensions of the velocity vector). Therefore, the sparse measurements are patched together in space and interpolated in order to provide a smooth (time-averaged) field (Aitken *et al.* 2014). Alternatively, the sparse measurements in space and time can be combined with the knowledge of the evolution model (i.e. the Navier–Stokes equations) to reconstruct the three-dimensional turbulent flow field. This process is referred to as state estimation or data assimilation (Le Dimet, Navon & Ștefănescu 2017).

Many data assimilation methods are based on Bayes' theorem, whereby the posterior probability density function of the state is estimated by updating the background distribution with fresh observations. The four-dimensional variational data assimilation (4D-Var) approach is a maximum *a posteriori* (MAP) version of the Bayesian framework which has been widely used in the literature for weather forecasting (Bannister 2017; Gustafsson *et al.* 2018). The weak formulation of the methodology involves solving an optimization problem that minimizes three quantities: the mismatch between real and synthetic measurements, the deviation from the background distribution, and the model error. However, this formulation requires knowledge of the full spatio-temporal correlation function of the model error, which is very high-dimensional and tedious to identify (Lorenç 1986). Moreover, it requires optimizing over the space–time direction, which results in a huge size for the control vector. Alternatively, a strong formulation of the problem is usually considered, in which a deterministic model along the time direction is used. As a result, the spatial background distribution is required only at the beginning of the assimilation window. Nevertheless, numerical computation of the latter correlation in the context of turbulent flow in the ABL remains prohibitively expensive.

Reconstructing the ABL flow by combining large-eddy simulations (LES) with 4D-Var and lidar measurements was first explored by Lin, Chai & Sun (2001). In this work, turbulent flow structures were retrieved in a convective boundary layer. This study was followed by a series of papers in which the same methodology was used to reconstruct the ABL flow based on a measurement campaign using two lidars (Chai & Lin 2003; Chai, Lin & Newsom 2004; Newsom & Banta 2004; Newsom *et al.* 2005; Xia *et al.* 2008). However, in these studies, the spatial correlation was ignored, and the problem was regularized by a Laplacian-based norm. Moreover, the continuity equation was not strictly imposed and was replaced by a penalization term in the cost function. In a more recent work by Bauweraerts & Meyers (2020), LES-based reconstruction of turbulent flow in a neutrally stable ABL was investigated using a more formal 4D-Var approach. To this end, a database of the spatial correlation tensor was constructed offline based on ensemble averaging over long prior LES. Afterwards, a strongly constrained 4D-Var problem was formulated in a

Karhunen–Loève (KL) basis, which is a divergence-free basis by construction, leading to a better conditioned problem. However, numerical computation of the two-point covariance tensor requires a huge amount of resources, computation time and disk storage, which leads to an impractical assimilation algorithm. Moreover, the covariance database is case specific, meaning that it must be recomputed for every possible atmospheric condition.

Over the past few decades, several attempts to model the full two-point second-order statistics of turbulence in a boundary layer were made. Hunt & Graham (1978) studied the statistics of turbulent structures in the free-stream flow near a moving flat surface. Later on, Wilson (1997) exploited these ideas to derive a simple isotropic analytical approximation (referred to as the Hunt–Graham–Wilson or HGW model) for the two-point correlation tensor in the atmospheric convective boundary layer (CBL). In another approach, Mann (1994) developed a three-dimensional spectral model for a homogeneous and neutrally stable ABL turbulence. Currently, the Mann model is widely employed for synthetic turbulence generation (Mann 1998) as well as for assessing structural loading on wind turbines (Sabale & Gopal 2019; Chen *et al.* 2022).

In the current work, our focus is on testing the feasibility of substituting the numerical database of the initial two-point background correlation tensor in the 4D-Var problem of Bauweraerts & Meyers (2020) with simple analytical approximations such as the HGW and Mann models. To this end, we study the convergence and accuracy of the reconstructed field. Furthermore, we also investigate the effect of increasing the assimilation time on the accuracy of the retrieved solution. The assimilation problem is formulated in a solenoidal basis to enforce continuity and preconditioned to enhance the convergence rate. The main focus of our work is on reconstructing the turbulence fluctuation field. Therefore, it is assumed that the mean vertical profile as well as the friction velocity u_* and surface roughness are known from the reference simulation. In a practical setting, these parameterizations may need to be estimated as well, e.g. using a hierarchical Bayesian method. However, this is beyond the scope of the current work.

The paper is structured as follows. In § 2, the variational data assimilation problem is derived. The analytical models of the two-point correlation tensor are introduced in § 3. Section 4 discusses the optimization methodology used in the current study. In § 5, the set-up of the case study is introduced; results are discussed in § 6. Finally, conclusions and suggestions for future research are presented in § 7.

2. Variational data assimilation

2.1. State space model

In this section, the 4D-Var problem in the context of turbulent flow reconstruction is formulated. The problem aims to reconstruct the turbulent flow state using a time series of lidar measurements and an LES model. As shown by Stuart (2010), the continuous 4D-Var problem needs to be formally derived in a functional space, requiring us to deal with probability measures with infinite dimensions. In order to avoid this complexity, Lorenc (1986) suggested to represent the model in a finite discrete basis before deriving the 4D-Var problem. For instance, Bauweraerts & Meyers (2020) formulated their 4D-Var problem in a truncated KL basis, which led to a mathematically simpler problem formulation. In the current study, we exploit the spatial discretization of the governing equation (see Appendix A) to derive the strong 4D-Var problem directly in a discrete basis. This is further elaborated below.

Consider a velocity vector $\mathbf{U}_0 = [U_0^1, U_0^2, U_0^3]^T \in \mathbb{R}^N$ defined at time $t = 0$ on a domain of size $L_1 \times L_2 \times H$, with $N_1 \times N_2 \times N_3$ uniformly distributed grid points. Our current

discretization uses a staggered grid for the vertical velocity component (see [Appendix A](#) for details), so that $N = N_1 N_2 (3N_3 - 1)$. Given the initial condition \mathbf{U}_0 , the LES model can be used to progress the solution in time. We write in short

$$\mathbf{U}(t) = \mathcal{M}_t(\mathbf{U}_0), \tag{2.1}$$

where \mathcal{M}_t is the solution operator representing the time integration of the LES model described in [Appendix A](#) and $\mathbf{U}(t)$ is the (modelled) state at time t . Furthermore, the initial condition \mathbf{U}_0 needs to be solenoidal. This leads to a constraint in the 4D-Var problem, i.e. the discretized divergence of the initial velocity \mathbf{U}_0 must be zero (see [Appendix C](#)). To avoid a constrained optimization problem, we simply eliminate the continuity constraint. To this end, we construct a projection and reconstruction operator such that $\boldsymbol{\Theta}_0 = \mathcal{P}(\mathbf{U}_0) \in \mathbb{R}^M$ with $M = N_1 N_2 (2N_3 - 1)$, $\mathbf{U}_0 = \mathcal{R}(\boldsymbol{\Theta}_0)$ with $\boldsymbol{\Theta}_0 = \mathcal{P}(\mathcal{R}(\boldsymbol{\Theta}_0))$ and $\text{div}(\mathcal{R}(\boldsymbol{\Theta}_0)) = 0$. More details are provided in [Appendix C](#).

2.2. Methodology

In this section, the strong 4D-Var problem is introduced. Consider a series of $N_s + 1$ observations $\mathbf{Y} \triangleq [\mathbf{y}_0, \dots, \mathbf{y}_{N_s}]$ sampled every T_s , with $\mathbf{y}_k \in \mathbb{R}^{N_m}$ a vector of N_m measurements at time t_k , where $t_k = t_0 + kT_s$. The latter vector can be modelled as

$$\mathbf{y}_k = \mathbf{h}_k(\mathcal{M}_k(\mathbf{U}_0) + \boldsymbol{\epsilon}_k) + \mathbf{v}_k, \tag{2.2}$$

where \mathbf{h} is the observation function introduced in § 5 for the lidar measurements and \mathbf{v} and $\boldsymbol{\epsilon}$ are the measurement and model errors, respectively. Using the strong assumption that $\mathbf{h}_k(\mathcal{M}_k(\mathbf{U}_0) + \boldsymbol{\epsilon}_k) = \mathbf{h}_k(\mathcal{M}_k(\mathbf{U}_0)) + \mathbf{h}_k(\boldsymbol{\epsilon}_k)$, which is valid for linear observation functions, and assuming that $\boldsymbol{\xi}_k = \mathbf{h}_k(\boldsymbol{\epsilon}_k) + \mathbf{v}_k$ is independent and normally distributed with variance $\gamma^2 \mathbf{I}$, we can express Bayes' rule as $p(\boldsymbol{\Theta}_0 | \mathbf{Y}) \sim p(\mathbf{Y} | \boldsymbol{\Theta}_0)p(\boldsymbol{\Theta}_0)$, with

$$p(\mathbf{Y} | \boldsymbol{\Theta}_0) \propto \prod_{k=1}^{N_s} \exp(-\|\mathbf{y}_k - \mathbf{h}_k(\mathcal{M}_k(\mathcal{R}(\boldsymbol{\Theta}_0)))\|^2 / 2\gamma^2) \tag{2.3}$$

and $p(\boldsymbol{\Theta}_0) \propto \exp(-\boldsymbol{\Theta}_0^T \mathbf{B}^{-1} \boldsymbol{\Theta}_0 / 2)$ with $\mathbf{B} = \mathcal{P} \mathbf{R} \mathcal{P}^T$, where \mathbf{R} is the two-point correlation tensor defined in § 3. In order to obtain the best estimate of the state $\boldsymbol{\Theta}_0$, the MAP estimate $\arg \max_{\boldsymbol{\Theta}_0} p(\mathbf{Y} | \boldsymbol{\Theta}_0)p(\boldsymbol{\Theta}_0)$ is often used. This can be obtained by taking the logarithm as $-\log[p(\boldsymbol{\Theta}_0 | \mathbf{Y})]$, leading to the following unconstrained optimization problem over the initial state $\boldsymbol{\Theta}_0$:

$$\underset{\boldsymbol{\Theta}_0}{\text{minimise}} \quad \mathcal{J}(\boldsymbol{\Theta}_0) = \underbrace{\frac{1}{2} \boldsymbol{\Theta}_0^T \mathbf{B}^{-1} \boldsymbol{\Theta}_0}_{\mathcal{J}_B} + \underbrace{\frac{1}{2\gamma^2} \sum_{n=1}^{N_s} \|\mathbf{y}_n - \mathbf{h}_n(\mathcal{M}_t(\mathcal{R}(\boldsymbol{\Theta}_0)))\|^2}_{\mathcal{J}_L}. \tag{2.4}$$

In the current study, our focus is on the initial background tensor \mathbf{B} and the regularization term \mathcal{J}_B . The latter term plays a crucial role in ensuring a well-posed assimilation problem and promoting fast convergence. Furthermore, it holds particular significance in regularizing the solution in areas where no measurements are available. At one extreme, the most straightforward approach for the covariance tensor is to use the identity matrix, $\mathbf{R} = \mathbf{I}$ (or $\mathbf{B} = \mathcal{P} \mathcal{P}^T$), which is equivalent to the standard ridge regression. However, this approach leads to slow convergence and relatively high reconstruction errors, as will be further discussed. At the other extreme, an ideal approach would

involve numerically constructing the true background tensor, which was demonstrated by Bauweraerts & Meyers (2020) to be prohibitively expensive.

In an attempt to achieve a trade-off between accuracy and simplicity, we study the feasibility of replacing the true background tensor with simple yet realistic analytical approximations. To this end, we use two different models for the two-point covariance tensor \mathbf{R} , which are discussed in detail in the following section.

3. Two-point correlation tensor

3.1. Overview

For a given mean velocity field $\bar{\mathbf{u}}(\mathbf{x})$, the fluctuations $\mathbf{u}'(\mathbf{x}) = \mathbf{u}(\mathbf{x}) - \bar{\mathbf{u}}(\mathbf{x})$ can be defined. The two-point velocity-correlation tensor is defined as

$$R_{ij}(\mathbf{x}, \check{\mathbf{x}}) = \langle u'_i(\mathbf{x})u'_j(\check{\mathbf{x}}) \rangle, \quad (3.1)$$

where $\langle \cdot \rangle$ indicates ensemble averaging. For homogeneous turbulence, the correlation tensor depends only on the separation $\mathbf{r} = (\check{x}_1 - x_1, \check{x}_2 - x_2, \check{x}_3 - x_3)$ between the two points. Hence, the isotropic-homogeneous spectral density tensor $\hat{\Phi}^{iso}(\mathbf{k})$, with wavenumber vector $\mathbf{k} = (k_1, k_2, k_3)$, can be obtained by applying the Fourier transform

$$\hat{\Phi}^{iso}(\mathbf{k}) = \frac{1}{8\pi^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{ij}(\mathbf{r}) \exp(-i(k_1 r_1 + k_2 r_2 + k_3 r_3)) dr_1 dr_2 dr_3. \quad (3.2)$$

For incompressible flows, the latter tensor can be expressed in terms of the energy spectrum $E(k)$ (Pope 2000) as

$$\hat{\Phi}_{ij}^{iso}(\mathbf{k}) = \frac{E(k)}{4\pi k^4} (\delta_{ij} k^2 - k_i k_j), \quad (3.3)$$

with $k = (k_1^2 + k_2^2 + k_3^2)^{1/2}$. An expression for the energy spectrum $E(k)$ can be analytically derived (e.g. Pope 2000) based on the asymptotic behaviour of the turbulent energy for large and small scales. In this paper, we investigate two variants of the energy spectrum $E(k)$ based on the asymptotic behaviour as $k \rightarrow 0$. The spectrum can be expressed as

$$E(k) = a\sigma_{iso}^2 \ell \frac{(k\ell)^p}{(1 + (k\ell)^2)^{5/6+p/2}}, \quad (3.4)$$

where ℓ is a length scale and σ_{iso}^2 is the isotropic variance. The power p determines the slope of $E(k)$ as $k \rightarrow 0$. When $p = 4$, the energy spectrum exhibits a behaviour of $E(k) \sim k^4$ for large scales, commonly known as Batchelor turbulence (Davidson 2010). Conversely, $p = 2$ leads to a behaviour of $E(k) \sim k^2$, which is also known as Saffman turbulence (Saffman 1967; Pope 2000). The scaling factor a is chosen such that the integration of $E(k)$ over all wavenumbers yields the total turbulent energy E_{tot} for both Batchelor and Saffman cases. This results in values of 1.45 and 1.1886 for the Batchelor and Saffman cases, respectively.

Although assuming homogeneous isotropic turbulence leads to great mathematical simplifications, wall-bounded flows in reality are known to behave otherwise. Nevertheless, in a neutral pressure-driven ABL over smooth terrain, turbulence is close to homogeneous in the horizontal plane (Holmes, Lumley & Berkooz 1996), and therefore homogeneity may be assumed in the x_1 and x_2 directions but not in the x_3 direction.

Consequently, the two-point correlation functions become dependent on the absolute vertical location of each point rather than the distance between them, and the Fourier transform over the vertical distance in (3.2) may not be applied. Instead, the two-point correlations are expressed in terms of the vertically inhomogeneous spectral tensor \hat{R} , which can be constructed from the function

$$\hat{R}_{ij}(k_1, k_2; x_3, \check{x}_3) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{ij}(r_1, r_2; x_3, \check{x}_3) \exp(-i(k_1 r_1 + k_2 r_2)) dr_1 dr_2. \tag{3.5}$$

Owing to our pseudo-spectral numerical discretization (see Appendix A), we focus on obtaining the correlation function directly as $\hat{R}_{ij}(k_1, k_2; x_3, \check{x}_3)$ instead of $R_{ij}(\mathbf{x}, \check{\mathbf{x}})$ in the remainder of this section.

3.2. The HGW model

Hunt & Graham (1978) studied the effect of introducing a moving rigid surface suddenly at time $t = 0$ on the isotropic homogeneous tensor $\Phi^{iso}(\mathbf{k})$. To model the blockage by the wall, the velocity field $\mathbf{u}(\mathbf{x}, t)$ was written as a sum of the homogeneous flow field $\mathbf{u}^H(\mathbf{x}, t)$ from the homogeneous tensor and a blockage field $\mathbf{u}^B(\mathbf{x}, t)$ obeying $u_3^H(x_1, x_2, 0, t) + u_3^B(x_1, x_2, 0, t) = 0$. When the shear-free case is considered (i.e. zero mean shear), \mathbf{u}^B is simply an irrotational velocity field that does not affect the initial vorticity field (Lee & Hunt 1989). For further details, we refer the reader to the original paper. Wilson (1997) used these ideas to derive an analytical model for the vertically inhomogeneous spectral tensor $\hat{R}(k_1, k_2; x_3, \check{x}_3)$ in the atmospheric CBL. The model was named the HGW model and is given as

$$\begin{aligned} \hat{R}_{ij}^{HGW}(k_1, k_2; x_3, \check{x}_3) &= \hat{R}_{ij}^{iso}(k_1, k_2; \check{x}_3 - x_3) \\ &+ e^{-\kappa \check{x}_3} m_j(k_1, k_2) \hat{R}_{3i}^{iso*}(k_1, k_2; x_3) \\ &+ e^{-\kappa x_3} m_i^*(k_1, k_2) \hat{R}_{3j}^{iso}(k_1, k_2; \check{x}_3) \\ &+ e^{-\kappa(x_3 + \check{x}_3)} m_i^*(k_1, k_2) m_j(k_1, k_2) \hat{R}_{33}^{iso}(k_1, k_2; 0), \end{aligned} \tag{3.6}$$

where $\kappa = (k_1^2 + k_2^2)^{1/2}$ and $(m_1, m_2, m_3) = (ik_1/\kappa, ik_2/\kappa, -1)$, with $\hat{R}_{ij}^{iso}(k_1, k_2, r)$ obtained by computing the inverse Fourier transform (IFT) of the three-dimensional spectral tensor $\hat{\Phi}_{ij}^{iso}(\mathbf{k})$ in the direction of k_3 , which can be analytically computed for Batchelor turbulence (Wilson 1997) (see Appendix D for more details). Although the model in (3.6) accounts for the vertical inhomogeneity, it does not capture the anisotropic behaviour imposed by the mean flow. The implications of this are further discussed in § 6.

3.3. The Mann model

Many turbulent tensor models in the literature are inherently isotropic. However, anisotropy effects are often included by introducing stretching factors in different directions, leading to many extra coefficients to be tuned (in addition to the isotropic tensor parameters) (Kristensen *et al.* 1989). Alternatively, Mann (1994) introduced a turbulence model based on rapid distortion theory, which accounts for anisotropic effects via a single additional parameter Γ . In this model, the spectral tensor is assumed to be initially isotropic at time $t = 0$, as given in (3.3). Afterwards, the isotropic tensor is distorted by

a constant shear profile for a certain period of time that is equal to the eddy lifetime. The resulting model is summarized as

$$\hat{\Phi}_{11}^{Mann}(\mathbf{k}) = E(k_0)(k_0^2 - k_1^2 - 2k_1k_{30}\zeta_1 + (k_1^2 + k_2^2)\zeta_1^2)/(4\pi k_0^4), \quad (3.7a)$$

$$\hat{\Phi}_{22}^{Mann}(\mathbf{k}) = E(k_0)(k_0^2 - k_2^2 - 2k_2k_{30}\zeta_2 + (k_1^2 + k_2^2)\zeta_2^2)/(4\pi k_0^4), \quad (3.7b)$$

$$\hat{\Phi}_{33}^{Mann}(\mathbf{k}) = E(k_0)(k_1^2 + k_2^2)/(4\pi k_0^4), \quad (3.7c)$$

$$\hat{\Phi}_{12}^{Mann}(\mathbf{k}) = E(k_0)(-k_1k_2 - k_1k_{30}\zeta_2 - k_2k_{30}\zeta_1 + (k_1^2 + k_2^2)\zeta_1\zeta_2)/(4\pi k_0^4), \quad (3.7d)$$

$$\hat{\Phi}_{13}^{Mann}(\mathbf{k}) = E(k_0)(-k_1k_{30} + (k_1^2 + k_2^2)\zeta_1)/(4\pi k_0^2k^2), \quad (3.7e)$$

$$\hat{\Phi}_{23}^{Mann}(\mathbf{k}) = E(k_0)(-k_2k_{30} + (k_1^2 + k_2^2)\zeta_2)/(4\pi k_0^2k^2), \quad (3.7f)$$

where k_0 is the magnitude of the initial wavenumber vector $\mathbf{k}_0 = (k_1, k_2, k_{30})$ with $k_{30} = k_3 + \beta k_1$, and where β is the non-dimensional eddy lifetime. Mann proposed an eddy lifetime model that has the correct asymptotic behaviour when compared with observations in wall-bounded flows. The non-dimensional eddy lifetime β was expressed in terms of the hypergeometric function ${}_2F_1$. However, Wilson (1998) reformulated this expression in terms of the incomplete beta function B :

$$\beta = \frac{\sqrt{3}\Gamma}{k\ell} \left[B_{1/(1+k^2\ell^2)} \left(\frac{1}{3}, \frac{5}{2} \right) \right]^{-1/2}, \quad (3.8)$$

which is much easier to compute using available numerical libraries.

The quantities ζ_1 and ζ_2 are given as

$$\zeta_1 = C_1 - \frac{k_2}{k_1}C_2, \quad \zeta_2 = \frac{k_2}{k_1}C_1 + C_2, \quad (3.9a,b)$$

with

$$C_1 = \frac{\beta k_1^2(k_0^2 - 2k_{30}^2 + \beta k_1k_{30})}{k^2(k_1^2 + k_2^2)} \quad (3.10)$$

and

$$C_2 = \frac{k_2k_0^2}{(k_1^2 + k_2^2)^{3/2}} \arctan \left[\frac{\beta k_1(k_1^2 + k_2^2)^{1/2}}{k_0^2 - \beta k_{30}k_1} \right]. \quad (3.11)$$

Unlike the HGW model, the Mann model is anisotropic, but it remains homogeneous in all directions. To account for the vertical inhomogeneity introduced by the wall, Mann (1994) proposed a modification to his original model based on a similar approach to that described in § 3.2. However, because of the non-zero mean shear in the model, the same procedure as in § 3.2 leads to a highly complex mathematical model that requires numerical integrations over the k_3 direction, with a computational cost of $\mathcal{O}(N_3^3)$ per horizontal wavenumber pair (k_1, k_2) . This would be a bottleneck in our assimilation algorithm. To avoid this complexity, we directly obtain $\hat{R}_{ij}^{Mann}(k_1, k_2, r_3)$ by applying the inverse fast Fourier transform to the model in (3.7). In contrast to the analytical IFT conducted over an infinite domain in the HGW model, the numerical IFT here yields a periodic correlation function in the vertical direction. This periodicity needs to be eliminated to avoid any unwanted effects on the assimilation results. To achieve this, the IFT is applied on an extended domain in the vertical direction, which is at least twice as

large as the reconstruction domain height. The additional part is then discarded. A similar approach was adopted in Mann (1998) in the context of synthetic turbulence generation in a finite domain. While it is acknowledged that employing the homogeneous version of the Mann model, as opposed to the vertically inhomogeneous variant, may yield less precise results in close proximity to the wall, it should be noted that this discrepancy diminishes as one moves away from the wall. In this case, the two versions of the model are anticipated to exhibit similar behaviour, as indicated by Mann (1994).

Before utilizing the Mann model in our assimilation algorithm, one remaining issue needs to be addressed. Specifically, the model in (3.7) is undefined when $k_1 = 0$ due to the singularity of ζ_1 and ζ_2 in (3.9a,b). To avoid this issue, the limit values $\lim_{k_1 \rightarrow 0} \zeta_1 = -\beta$ and $\lim_{k_1 \rightarrow 0} \zeta_2 = 0$ are used (Gilling 2009).

4. Optimization methodology

To solve the unconstrained optimization problem (2.4), the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints (L-BFGS-B) implemented by Byrd *et al.* (1995) is used to determine the search direction d^k . Since the L-BFGS-B is a quasi-Newton algorithm, the Hessian matrix is approximated based on a set of correction pairs of the differences between previous gradients and search directions. In the current study, five correction pairs are used. Once a search direction is approximated, a step size α^k is selected according to the Moré–Thuente line search algorithm (Moré & Thuente 1994). The typical values of the constants $c_1 = 10^{-4}$ and $c_2 = 0.9$ were used for the Wolfe conditions (Nocedal & Wright 2006). Since the gradient of the cost function $\partial \mathcal{J} / \partial \hat{\Theta}_0$ is a key ingredient of the quasi-Newton algorithms, it is crucial to have an efficient algorithm to compute it. In this study the continuous adjoint method is used (see Appendix B). After deriving the continuous adjoint equation, it is discretized and solved similarly to the forward model. Afterwards, the gradient is obtained as

$$\frac{\partial \mathcal{J}}{\partial \hat{\Theta}_0} = \hat{B}^{-1} \hat{\Theta}_0 - \hat{\mathcal{R}}^* \hat{Z}_0, \tag{4.1}$$

where \hat{Z}_0 corresponds to the discretized adjoint velocity at the beginning of the assimilation window and $\hat{\mathcal{R}}^*$ is the Hermitian transpose of $\hat{\mathcal{R}}$.

In the remainder of this section, we propose a technique to enhance the convergence rate of the optimization algorithm described above. Consider the problem in (2.4); the Hessian matrix can be represented as a sum of the Hessians of both terms, $\mathcal{H} = (\hat{B}^{-1} + \hat{C}^{-1})^{-1}$, where \hat{C}^{-1} is the unknown Hessian matrix of the likelihood term, which contains implicitly second derivatives of the state equations, and is prohibitively expensive to explicitly formulate. During the optimization procedure, the L-BFGS-B attempts to estimate \mathcal{H} using subsequent gradient information. The new iteration $\hat{\Theta}_{0_{k+1}}$ can then be written as

$$\begin{aligned} \hat{\Theta}_{0_{k+1}} &= \hat{\Theta}_{0_k} + \alpha_k d_k, \\ &= \hat{\Theta}_{0_k} + \alpha_k \tilde{\mathcal{H}}_k (\partial \mathcal{J} / \partial \hat{\Theta}_0)_k, \end{aligned} \tag{4.2}$$

where k is the iteration index and $\tilde{\mathcal{H}} \approx (\hat{B}^{-1} + \hat{C}^{-1})^{-1}$ is the L-BFGS-B estimate of the Hessian. Since \hat{B} is known from § 3, it can be used to precondition the problem such that

The LES-based reconstruction of turbulence

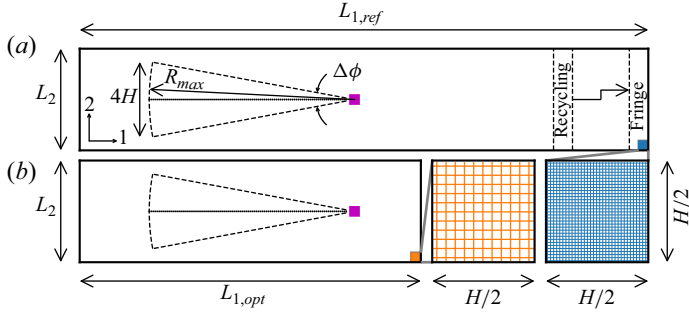


Figure 1. Schematic of the reference (a) and reconstruction (b) domains with the lidar mount (purple) and the scanning region (inside dashed lines). The figure is reprinted from Bauweraerts & Meyers (2020).

the L-BFGS-B does not need to predict its complex structure. To this end, we write

$$\hat{\Theta}_{0_{k+1}} = \hat{\Theta}_{0_k} + \alpha_k \tilde{\mathcal{H}}'_k \hat{\mathbf{B}} (\partial \mathcal{J} / \partial \hat{\Theta}_0)'_k, \quad (4.3)$$

where $(\partial \mathcal{J} / \partial \hat{\Theta}_0)'_k = \hat{\mathbf{B}}_k (\partial \mathcal{J} / \partial \hat{\Theta}_0)_k$ is the preconditioned gradient. Using the latter gradient rather than the original one in the L-BFGS-B leads to the estimation of $\tilde{\mathcal{H}}' \approx (\hat{\mathbf{B}}^{-1} + \hat{\mathbf{C}}^{-1})^{-1} \hat{\mathbf{B}}^{-1} = (\mathbf{I} + \hat{\mathbf{B}} \hat{\mathbf{C}}^{-1})^{-1}$, which has potentially a much simpler structure and leads in our test cases to much faster convergence.

5. Case set-up

5.1. Set-up of lidar and virtual measurements

In this study, the measurement series $\mathbf{Y} \triangleq [y_0, \dots, y_{N_s}]$ is obtained virtually on a fine reference simulation (see § 5.2) using the implementation of Bauweraerts & Meyers (2020) for the Lockheed Martin WindTracer lidar (Krishnamurthy *et al.* 2013). A full description of the implementation and set-up can be found in the original papers. However, for the sake of completeness, a brief overview is provided in this section.

In the current work, we collect our measurements in plan-position-indicator (PPI) scanning mode with a zero elevation angle and constant azimuthal sweeping covering the horizontal sector shown in figure 1. The measurements are collected for a period of $T_m = 0.1H/u_*$ from a single lidar mounted at $\mathbf{x}_m = [0, 0, 0.1H]$, with a range gate of $\Delta r = 105$ m and total number of gates of $N_r = 100$. Therefore, at each sampling time $t_n = t_0 + nT_s$, with $T_s = 5 \times 10^{-4}H/u_* = 1$ s, a vector of wind-speed readings $\mathbf{y}_n = [y_{n,1}, \dots, y_{n,N_r}]$ is measured using $y_{n,i} = h_{n,i}(\mathbf{u}(x, t))$ and stored at equally spaced locations along the lidar beam. The observation function $h_{n,i}$ is given as (Bauweraerts & Meyers 2020)

$$h_{n,i}(\mathbf{u}(x, t)) \triangleq \frac{1}{T_s} \int_{t_{n-1}}^{t_n} \int_{\Omega} \mathbf{u}(x, t) \cdot \mathbf{e}_l(t) \mathcal{G}(\mathbf{Q}(t)(x - \mathbf{x}_i(t))) \, dx \, dt, \quad (5.1)$$

where \mathcal{G} is the lidar filter kernel, \mathbf{e}_l is the lidar beam direction, \mathbf{Q} is a rotation matrix from the reference to the lidar coordinate system and \mathbf{x}_i is the measurement location given as $\mathbf{x}_i(t) = \mathbf{x}_m + (r_0 + \Delta r(i - 1))\mathbf{e}_l(t)$, with the lidar range gate $\Delta r = 105$ m and $i = 1, \dots, N_r$. The lidar beam moves in a horizontal sweeping pattern with a constant azimuthal angular velocity of $|\partial \phi / \partial t| = 2\Delta \phi / T_m = 0.00371 \text{ rad s}^{-1}$, where $\Delta \phi$ is the azimuthal range covered by the lidar. Figure 1 summarizes this picture.

i	1	2	3
σ_{ii}^2/u_*^2	4.77	2.68	1.46

Table 1. Values of velocity variances as given in Kaimal *et al.* (1972).

5.2. Set-up of reference simulation

In the current study, we assume a pressure-driven boundary layer with a height of $H = 1$ km, a friction velocity of $u_* = 0.5 \text{ m s}^{-1}$ and a roughness length of $z_0 = 0.1$ m. After a spin-up period of $50H/u_*$ on a reference domain $30H \times 5.4H \times H$ with a fine grid of size $2000 \times 360 \times 200$, a series of virtual measurements are collected in the PPI scanning mode over a period of $T_m = 200$ s and also 400 s. After obtaining the measurement series \mathbf{Y} , the reconstruction problem can be started. The reconstruction is done on a shorter domain with a coarser mesh than the reference simulation. In this study, the reconstruction domain has dimensions of $18H \times 5.4H \times H$ with a grid size of $360 \times 108 \times 60$ (see figure 1). Before solving the optimization problem (2.4), the background correlation tensor $\hat{\mathbf{R}}$ is constructed from the analytical models in § 3 on the reconstruction grid and stored to be used as a prior for the MAP problem in (2.4). Computing and storing the analytical tensor takes only a few seconds using the same computational nodes allocated for the optimization problem (see § 5.4). Therefore, they are usually deleted after solving the optimization problem and recomputed again when needed.

5.3. Set-up of correlation tensor

5.3.1. The HGW model

As discussed above, the HGW correlation tensor in (3.6) is a function of the length scale ℓ and the isotropic variance σ_{iso}^2 . Meteorological data from the ABL (Kaimal *et al.* 1972) show that the variances of the different velocity components σ_u^2 , σ_v^2 and σ_w^2 in the surface layer (i.e. $x_3 \leq 0.1H$) are not equal to each other (see table 1). However, because of the isotropic assumption in the HGW model, the model cannot pick up a preferred direction, meaning that only a single choice of the variance can be imposed. In this study, we choose the isotropic variance to be equal to the streamwise variance of Kaimal *et al.* (1972), leading to $\sigma_{iso}^2 = 4.77u_*^2$. On the other hand, the length scale ℓ is chosen to match the inertial subrange asymptote (Peltier *et al.* 1996), which leads to the following expression (Wilson 1997):

$$\ell = 0.8743 \frac{\sigma_{iso}^3}{\epsilon}, \tag{5.2}$$

where ϵ is the turbulent kinetic energy dissipation rate. Using the similarity theory in the surface layer of a neutrally stable ABL, the dissipation rate can be approximated as (Stull 1988)

$$\epsilon = 1.24 \frac{u_*^3}{\bar{\kappa}z}, \tag{5.3}$$

where $\bar{\kappa} = 0.41$ and z is the vertical distance, which is chosen to be equal to the lidar height $z = 0.1H$. The resulting values of the parameters are summarized in table 3. It is worth mentioning that we noticed that the reconstruction results are not very sensitive to small changes in the model parameters.

i	1	2	3
σ_{ii}^2/u_*^2	4.75	1.95	1.15

Table 2. Values of velocity variances as calculated from the reference LES at $x_3 = 0.1H$.

Model	ℓ	σ_{iso}^2	Γ
HGW (Batchelor)	$0.301H$	$4.77u_*^2$	—
Mann (Batchelor)	$0.045H$	$1.37u_*^2$	4.2
Mann (Saffman)	$0.046H$	$1.36u_*^2$	3.0

Table 3. Parameters of the spectral tensors in § 3.

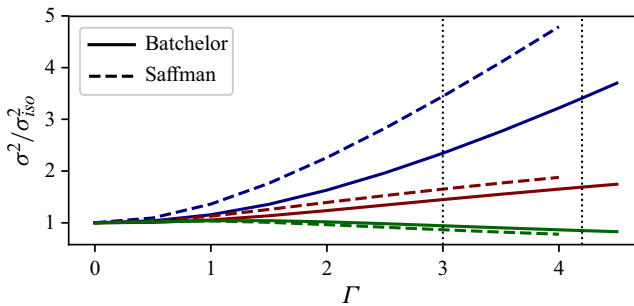


Figure 2. The normalized velocity variance of the Mann model as Γ changes using Batchelor and Saffman energy spectra. The blue colour represents $\sigma_u^2/\sigma_{iso}^2$, while the red and green represent $\sigma_v^2/\sigma_{iso}^2$ and $\sigma_w^2/\sigma_{iso}^2$, respectively. The vertical dotted lines correspond to the selected Γ values following the procedure in § 5.3.2 for both Batchelor and Saffman cases.

As briefly mentioned in § 3, the analytical IFT for the spectral tensor $\Phi^{iso}(\mathbf{k})$ used in the HGW model is only available for the Batchelor turbulence case (see Appendix D), which is the main advantage of this model. Therefore, we limit our analysis for the HGW to the Batchelor case.

5.3.2. The Mann model

In addition to the two parameters discussed above (σ_{iso}, ℓ), the Mann model has an additional parameter Γ which reflects the amount of distortion applied to the initial isotropic tensor. If $\Gamma = 0$, no shear is imposed and the isotropic tensor in (3.3) is retrieved. As Γ is increased, σ_u^2 and σ_v^2 increase while σ_w^2 and σ_{uw} decrease. This behaviour is illustrated in figure 4 of Mann (1994). To choose the model parameters, Mann used a least-squares fitting algorithm to match the one-dimensional spectra produced by the model to the ones obtained experimentally. Here, we use a simpler procedure, similar to the one reported by Wilson (1998), that is also more in line with the typical data that would be available from atmospheric measurements. Firstly, using figure 2, Γ is chosen to satisfy the ratio σ_u^2/σ_w^2 obtained from the reference LES at $x_3 = 0.1H$ (see table 2). Afterwards, Γ and σ_u^2/u_*^2 are used to obtain σ_{iso}^2 from figure 2. Finally, the length scale ℓ is obtained similarly to the previous section. Table 3 summarizes the parameters obtained following this procedure for both Batchelor and Saffman energy spectra.

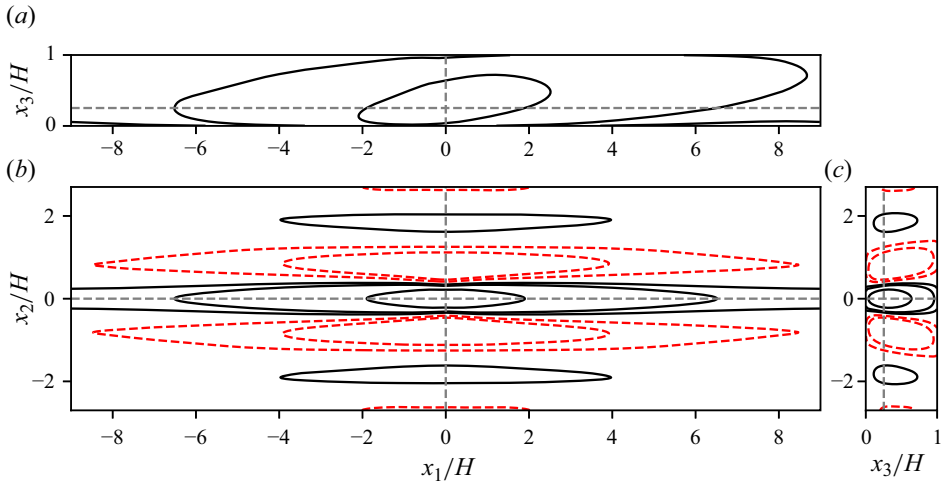


Figure 3. Cross-section of the two-point correlation $C_{11}(x, \tilde{x})$ with reference point $x = [0, 0, H/4]$ at $x_3 = H/4$ (b), $x_2 = 0$ (a) and $x_1 = 0$ (c). The contour lines are drawn for $C_{11} = [-0.1, -0.05, 0.05, 0.1, 0.3]$: solid black, $C > 0$; dashed red, $C < 0$. The figure is reprinted from Bauweraerts & Meyers (2020).

As will be described in § 5.4, the absolute value of σ_{iso}^2 in table 3 becomes somehow redundant in our 4D-Var problem due to the Pareto front technique used to fix γ^2 . Nonetheless, we opted to propose a way to fix it as part of the parameter selection process in the current section.

5.3.3. Two-point correlation tensor

In this section, we compute the diagonal component of the normalized two-point correlation function in the streamwise direction, given as $C_{11}(x, \tilde{x}) = R_{11}(x, \tilde{x})/\sigma_u^2(x, \tilde{x})$, for both the Mann and the HGW models presented in § 3, and compare them with those obtained by Bauweraerts & Meyers (2020) in figure 3.

Figures 4 and 6 present C_{11} for the HGW and Mann models, respectively, for a reference point located at $x = [0, 0, H/4]$, which is a bit higher than the lidar height in § 5.1. However, we opted to show the correlation function at this height to facilitate comparisons with figure 3. Moreover, the difference between the two heights is negligible (not shown). The figures provide a visual comparison of the correlation functions between the two models. For the HGW model with Batchelor spectrum (figure 4), the correlation consists of three small positive central lobes with a spanwise width close to what is observed in Bauweraerts & Meyers (2020). However, the HGW model underpredicts the correlation length by an order of magnitude when compared with the value of $10H$ reported by Fang & Porté-Agel (2015) around the reference point.

On the other hand, the Mann model in figures 5 and 6 exhibits a more complex correlation function. In the horizontal plane, the function displays three closed lobes elongated in the x_1 direction, with a size significantly larger than in the HGW case. However, it can be clearly seen that utilizing the Saffman spectrum (figure 6) leads to the longest correlations among the proposed models. In the spanwise direction, the correlation function decreases until it reaches a saturation point at a small negative value. In contrast, Bauweraerts & Meyers (2020) observed long alternating positive and negative lobes extending throughout the entire domain in the streamwise direction, with a correlation

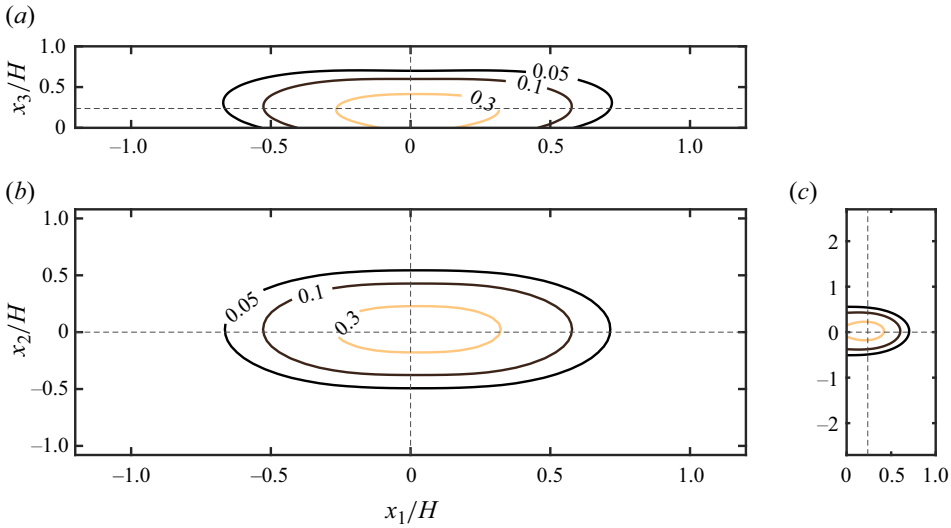


Figure 4. Cross-section of the HGW two-point correlation $C_{11}(x, \check{x})$ with reference point $x = [0, 0, H/4]$ at $x_3 = H/4$ (b), $x_2 = 0$ (a) and $x_1 = 0$ (c).

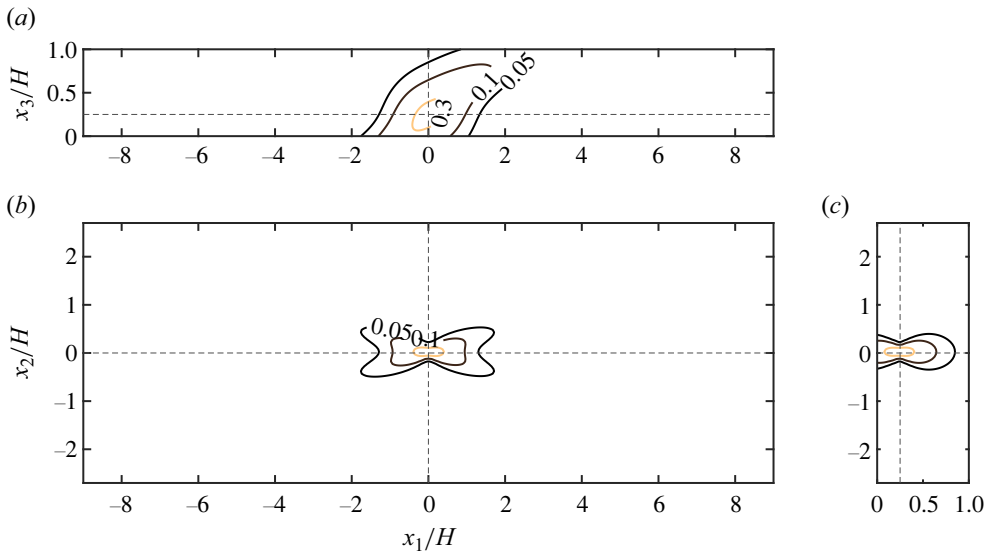


Figure 5. Cross-section of the Mann (with Batchelor spectrum) two-point correlation $C_{11}(x, \check{x})$ with reference point $x = [0, 0, H/4]$ at $x_3 = H/4$ (b), $x_2 = 0$ (a) and $x_1 = 0$ (c).

width much less than the one observed here. These long lobes were attributed to the large-scale coherent structures in a neutral ABL as observed in Alcayaga *et al.* (2020).

In the x_2 – x_3 direction, we also observe central positive lobes. In the x_1 – x_3 plane, the Mann model produces positive lobes that are inclined in the direction of the mean flow with an inclination angle of around 19° , which is slightly higher than what is observed in other studies (Marusic & Heuer 2007; Sillero, Jiménez & Moser 2014), where a value of around 10° was reported. We observed that the inclination is very sensitive to the choice of Γ . However, in the current work, we prioritize a practicable way of selecting parameters,

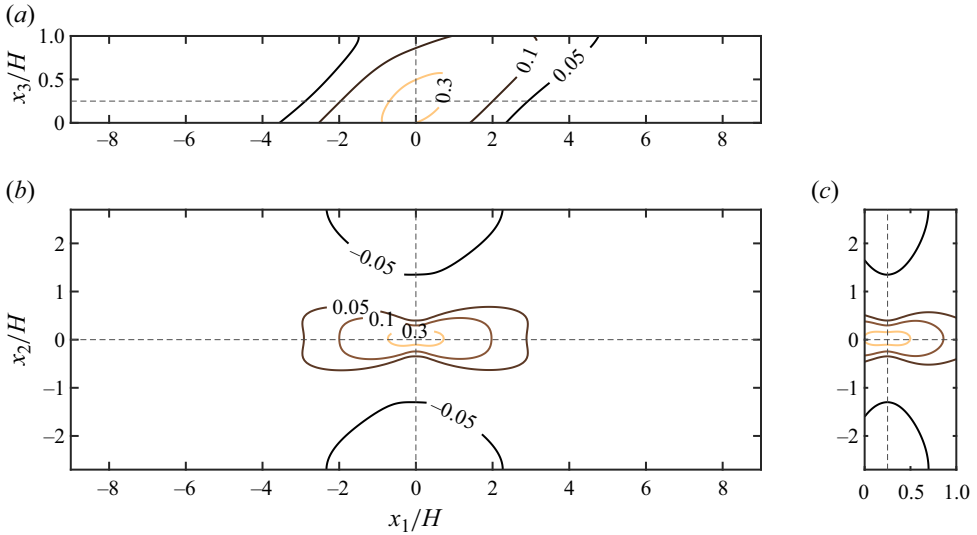


Figure 6. Cross-section of the Mann (with Saffman spectrum) two-point correlation $C_{11}(\mathbf{x}, \check{\mathbf{x}})$ with reference point $\mathbf{x} = [0, 0, H/4]$ at $x_3 = H/4$ (b), $x_2 = 0$ (a) and $x_1 = 0$ (c).

as outlined in § 5.3.2, rather than further fitting on the inclination angle, which is not directly identifiable as an input to the prior in our 4D-Var formulation.

5.4. Set-up of the assimilation problem

After determining the correlation tensor in the previous section, one remaining component to be chosen before solving the optimization problem (2.4) is the variance γ^2 . Looking at the optimization problem, the objective function can be rewritten in a more explicit way as $\mathcal{J}(\Theta_0) = \Theta_0^T \mathbf{B}'^{-1} \Theta_0 / 2\sigma_{iso}^2 + \sum_{n=1}^{N_s} \|\mathbf{y}_n - \mathbf{h}_n(\mathcal{M}_t(\mathcal{R}(\Theta_0)))\|^2 / 2\gamma^2$, where \mathbf{B}' is the normalized correlation tensor. As can be noticed, the relative weight of each term with respect to the other is determined by the ratio $\gamma^2 / \sigma_{iso}^2$, instead of the individual values of γ^2 and σ_{iso}^2 . However, since σ_{iso}^2 was already fixed in § 5.3 as part of the parameter selection procedure, γ^2 now serves as the only weighting parameter between the two terms in the cost function, and the absolute value of σ_{iso}^2 becomes somehow redundant. For a fixed σ_{iso}^2 , as γ^2 is increased, less trust is given to the likelihood term, leading to a smoother solution. Decreasing γ^2 gives more significance to the likelihood term, and the obtained solution becomes more irregular. In order to choose a suitable value for γ^2 , a trade-off between solution complexity and smoothness needs to be achieved. Similarly to Bauweraerts & Meyers (2020), this is done through Pareto front analysis, which requires solving the optimization problem multiple times per model for different values of γ^2 . Figure 7 shows that $\gamma^2 = 10^{-3}$ achieves a good trade-off for both models, so it will be used in the rest of this study. According to our experience, the reconstruction accuracy appears to be much more sensitive to big changes in γ^2 than the convergence behaviour.

After choosing γ^2 , the optimization problem (2.4) is solved for an optimization time horizon equivalent to the lidar scanning period T_m , with an initial condition $\hat{\Theta}_0 = 0$. The spatial mean flow profile of the initial field as well as the surface roughness and u_* are assumed to be known from the reference simulation. Starting from the initial condition,

The LES-based reconstruction of turbulence

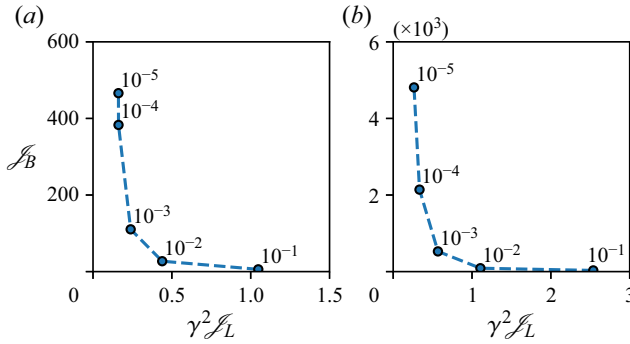


Figure 7. Pareto front for the optimization problem (2.4) using the HGW (a) and Mann regularization with Batchelor spectrum (b). The figure is obtained by solving the optimization problem many times using different values of γ^2 . The points are labelled with the corresponding γ^2 value.

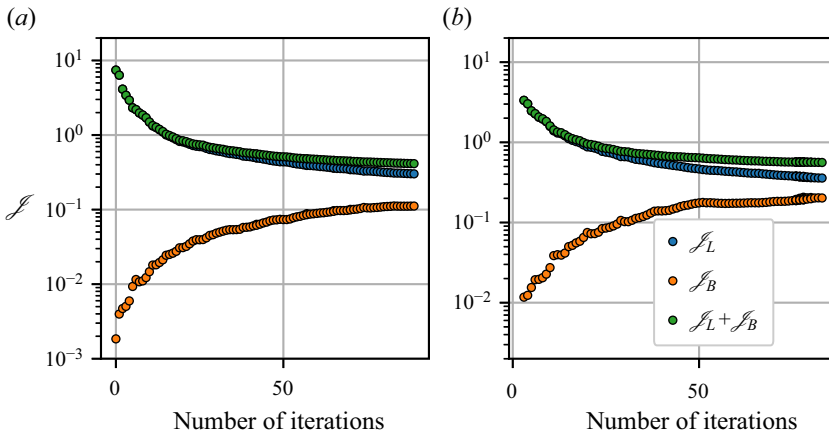


Figure 8. Cost function convergence (100 iterations) using the HGW (a) and Mann regularization with Batchelor spectrum (b).

the optimization algorithm converges to a local minimum where $\nabla \mathcal{J} = 0$. In this study, the optimization algorithms stop upon achieving a relative value of the gradient norm $\|\nabla \mathcal{J}\|/\|\nabla \mathcal{J}_0\| = 6 \times 10^{-3}$, where $\nabla \mathcal{J}_0$ is the initial gradient obtained using an initial guess for the initial condition $\hat{\Theta}_0 = 0$. Figures 8 and 9 show the convergence history of the HGW and Mann models with the Batchelor spectrum. Each iteration corresponds to an L-BFGS-B iteration, which involves a forward and adjoint simulation as well as a line search procedure (if the initial step α_k does not satisfy the Wolfe conditions). However, we only needed to perform the additional line search procedure a few times, so the total number of iterations mainly represents the number of forward simulations (or half of the forward and adjoint simulations). Figure 9 shows that we achieve a super-linear convergence rate when either of the analytical models is used.

As a point of comparison, we also investigated ridge regression. We found that it leads to slow convergence (a slope of -1.1), and only provides reasonable results inside the scanning area, as further discussed in Appendix E. We will use results from the ridge regression as a further point of comparison in the reconstruction error analysis discussed in § 6.2.

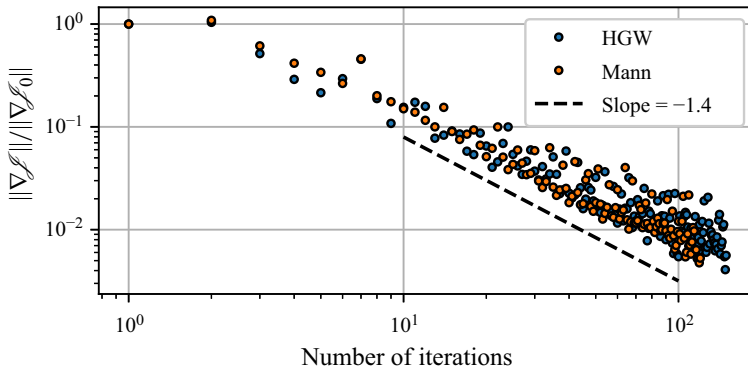


Figure 9. Convergence of the relative gradient using the HGW regularization (blue) and Mann regularization with Batchelor spectrum (orange).

The optimization in this study was performed on the wICE supercomputer of the Flemish supercomputer centre using four computational nodes, each of which has two Intel Xeon Platinum 8360Y 2.4 GHz CPUs with 256 GB RAM. The total wall time of one assimilation problem with $T_m = 0.1H/u_*$ (provided 100 iterations) is around 8 hours. It must be highlighted that this cost includes the preprocessing cost (involving the computation of the background tensor), which is negligible in our assimilation algorithm compared with the previous study of Bauweraerts & Meyers (2020).

6. Results

In this section, the reconstructed turbulent velocity field is visualized, and the results are compared with the reference field from which the virtual lidar measurements were collected.

6.1. Single assimilation window

We first discuss the results of a single assimilation window with an assimilation horizon of $T_m = 0.1H/u_*$ using the analytical models described in §3. Figure 10 presents a comparison between the reconstructed velocity fluctuations and the reference values obtained using two different models: the HGW model with a Batchelor spectrum and the Mann model with a Saffman spectrum. Similar plots were also generated for the Mann model with a Batchelor spectrum, but those are not shown here as they are not significantly different from the Saffman case. The results are shown at the beginning ($t = t_i$) and the end ($t = t_f$) of the assimilation window for vertical and horizontal cross-sections at $x_2 = 0$ and $x_1 = 0.1H$, respectively. Inside the scanning area, both models show good reconstruction accuracy in figure 10(g–l). Additionally, some fluctuations were also captured in small regions upstream and downstream of the scanning area due to the horizontal transport by the mean flow (Bauweraerts & Meyers 2020). Away from the scanning region, the likelihood term in the optimization problem (2.4) vanishes due to the absence of any measurements, and the only remaining contribution in the MAP problem is due to the initial background distribution provided by the analytical models.

For the vertical section in figure 10(a–f), the lidar measurements are available only at a height of $x_3 = 0.1H$. The mean velocity in the vertical direction is zero, so that no mean advection in that direction occurs. Therefore, the small structures are reconstructed only in small local regions around the lidar beam. However, due to the decaying correlations in the

The LES-based reconstruction of turbulence

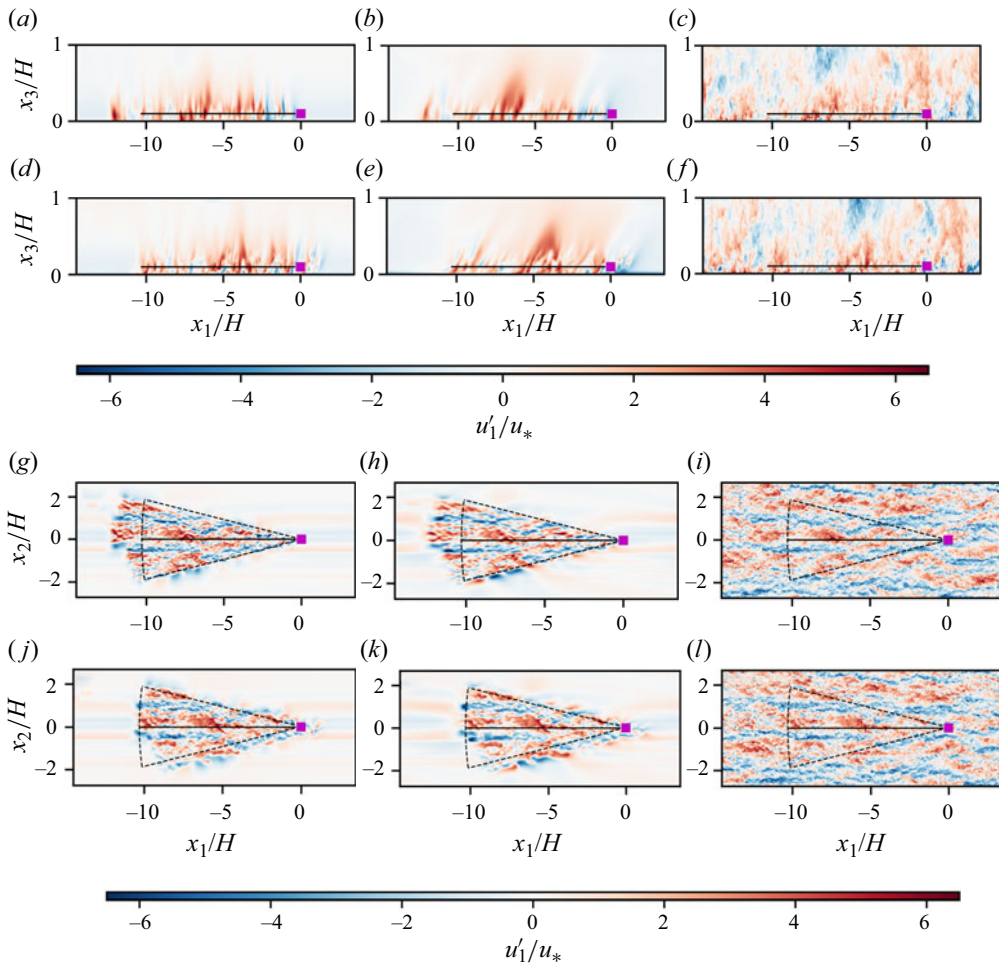


Figure 10. The streamwise velocity fluctuation component at the $x_2 = 0$ (a–f) and $x_3 = 0.1H$ (g–l) planes. The right column is the reference velocity and the left two columns are the reconstructed velocities using the HGW model and the Mann model (with Saffman spectrum), respectively. Panels (a–c) and (g–i) show the velocity at $t = 0$, while (d–f) and (j–l) show the velocity at $t = 0.1H/u_*$. The dashed lines represent the boundaries of the scanning area, and the solid line represents the lidar beam.

vertical direction provided by the analytical models, large structures are reconstructed up to a height of approximately $x_3 = 0.7H$. These observations agree with the results obtained in figures 4–6. Owing to the isotropic assumption in the HGW model, no inclination in the reconstructed fluctuations in the vertical direction is observed, whereas in the Mann model the inclination in the direction of the mean flow can be clearly seen.

To more closely examine the accuracy of the velocity reconstruction in the scanning area, we compare the reference and reconstructed streamwise velocity components for different sections at the middle of the assimilation window ($t = 0.05H/u_*$) in figure 11. At the lidar height ($x_3 = 0.1H$), both models exhibit very good agreement between the reference and reconstructed velocity components. However, as the height increases, the reconstruction accuracy gradually decreases, as expected from the earlier discussion. On the other hand, the reconstruction closer to the wall is accurate only for the large and not

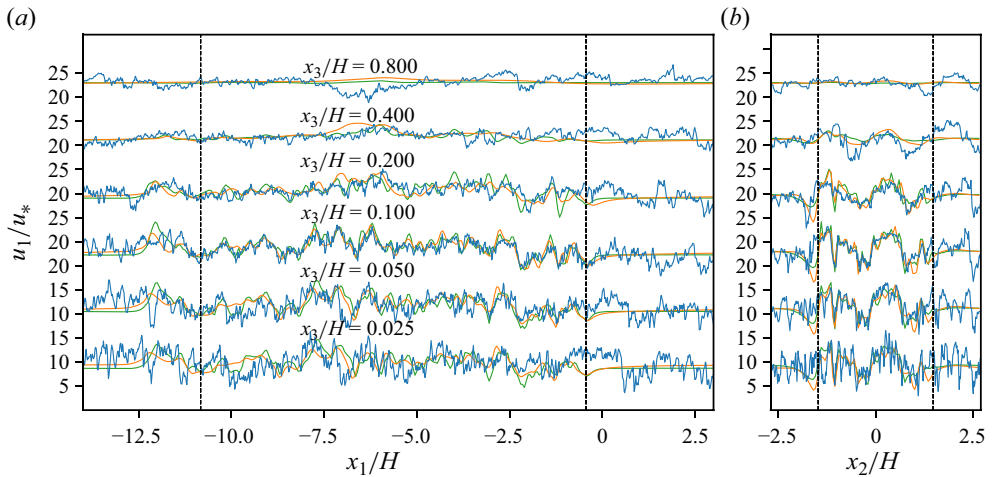


Figure 11. Comparison between the reconstructed velocity using the HGW model (green), the reconstructed velocity using the Mann model with Saffman spectrum (orange) and the reference velocity (blue) for the streamwise component at different lines in the vertical direction. The measurements are taken at $t = 0.05H/u_*$.

the small scales, which is expected from figures 4–6. This observation is seen to be true for both models, as seen in figure 11.

6.2. Error statistics

In this section, the statistics of the error in the reconstructed streamwise velocity component are studied. To this end, we define two metrics based on the instantaneous reconstruction error for the streamwise velocity component $\epsilon(\mathbf{x}, t) = I_c^f(u_1(\mathbf{x}, t)) - u_{1,ref}(\mathbf{x}, t)$, where u_1 is the reconstructed velocity, $u_{1,ref}$ is the reference velocity and I_c^f represents interpolation from the coarse reconstruction to the fine reference grid.

In order to facilitate a direct comparison with the results of the reconstruction method based on proper orthogonal decomposition (POD) presented in Bauweraerts & Meyers (2020), we define, in a first step, the spatially averaged error $\langle (\epsilon(\mathbf{x}, t))^2 \rangle_\Gamma$, where Γ is the horizontal area swept by the lidar beam after truncating one convective distance $U_\infty T$ to avoid boundary effects. In figure 12, we present a comparison of the normalized spatially averaged errors of the HGW, Mann (Saffman) and simple ridge regression models and the POD-based reconstruction errors reported in Bauweraerts & Meyers (2020). The horizontal dashed line represents the reconstruction using the mean flow only, with (given the normalization by $\langle u_{1,ref}^2 \rangle_\Gamma$) a corresponding value of 1. Examining figure 12(a) reveals that the ridge regression regularization results in an acceptable accuracy only within the scanning region, corroborating the findings detailed in Appendix E. In contrast, using either the HGW or Mann regularization yields a reconstruction accuracy closer to that achieved by the considerably more resource-intensive POD-based approach. In figure 12(b), errors are depicted over time at the lidar height $x_3 = 0.1H$. Notably, both the HGW and the Mann models show highly satisfactory reconstruction accuracy, closely rivalling that of the POD approach. The ridge regression reconstruction exhibits larger errors at the start of the assimilation window that decrease gradually with time. In alignment with observations from Bauweraerts & Meyers (2020), optimal reconstruction

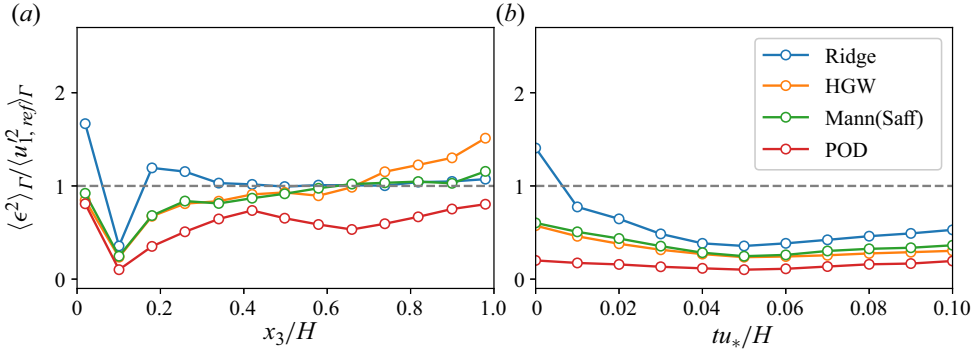


Figure 12. The normalized spatially averaged error inside the scanning area for the ridge, HGW and Mann (Saffman) models and the POD-based regularization in Bauweraerts & Meyers (2020). Panel (a) shows the errors at $t = 0.05u_*/H$, while (b) shows the results at $x_3 = 0.1H$. The horizontal dashed line represents estimation with the mean flow, which leads to an expected error of $\langle \epsilon^2 \rangle_\Gamma = \langle u_{1,ref}^2 \rangle_\Gamma$.

results are consistently achieved at the midpoint of the assimilation window for all models considered. Further insights into this phenomenon will be explored in § 6.3.

We now further focus on a detailed error comparison by introducing a second metric, namely the mean-squared error (MSE) of the reconstructed velocity, which can be defined as

$$MSE(\mathbf{x}) = \langle \langle \epsilon^2(\mathbf{x}, t) \rangle_T \rangle_e, \tag{6.1}$$

where $\langle \cdot \rangle_e$ and $\langle \cdot \rangle_T$ denote the ensemble average and time average, respectively. Furthermore, the error metric is normalized using the background variance $\langle \langle u_{1,ref}^2 \rangle_T \rangle_e$ similar to the above. The normalized mean-squared error (NMSE) can then be defined as $NMSE(\mathbf{x}) = MSE(\mathbf{x}) / \langle \langle u_{1,ref}^2 \rangle_T \rangle_e$. Figure 13 shows sections of the NMSE for the reconstructed streamwise velocity component considering 10 different assimilation windows (with a time horizon of $T = 200$ s) to compute the ensemble average for each model in § 3, while the time average is performed over each assimilation window. The NMSE is computed based on three different models: the HGW model (figure 13a,d), the Mann model with Batchelor spectrum (figure 13b,e) and the Mann model with Saffman spectrum (figure 13c,f). Within the scanning region in the horizontal plane, all models exhibit a high level of accuracy in reconstructing the flow. However, outside the scanning area, the accuracy drops significantly for the Mann model with Batchelor spectrum in the streamwise direction compared with the other two models. Nevertheless, using the Saffman instead of the Batchelor spectrum reduces these errors, as seen in the figure. This can be attributed to the fact that the Saffman spectrum leads to longer positive correlations in the longitudinal direction compared with the Batchelor spectrum, as discussed in § 5.3.3. We further looked at energy spectra over the domain width and over the different reconstruction windows (not shown here), observing that, indeed, the Saffman case has much more energy in the large scales. In the regions further away from the measurement area, the NMSE is close to 1, which indicates that the model is not able to predict any turbulent features, which is a logical result of the lack of measurements in this region.

In the vertical direction, figure 13(a–c) demonstrates a marginal improvement in the reconstruction accuracy close to the lidar height for the Mann model using either of the energy spectra compared with the HGW case, but these differences may fall within the averaging accuracy.

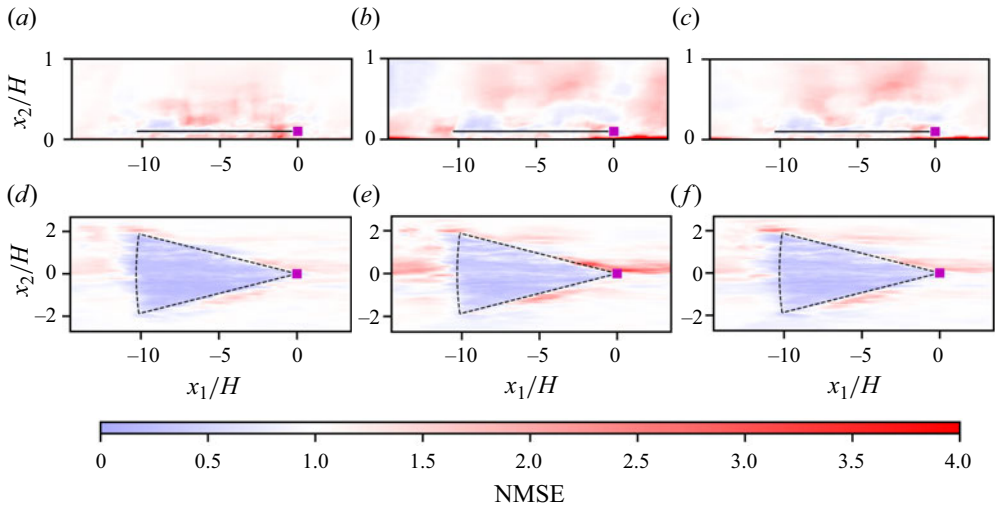


Figure 13. The NMSE of the reconstructed streamwise velocity component at $x_3 = 0.1H$ using the HGW regularization (a,d), the Mann regularization with Batchelor spectrum (b,e) and the Mann regularization with Saffman spectrum (c,f).

It is important to note that obtaining a high reconstruction accuracy far away from the scanning region is not the primary goal of the data assimilation problem. In fact, this region is usually filtered out in practice to end up with only the region of interest (i.e. the scanning area, possibly slightly extended to incorporate effects of advection in the reconstruction). While the correlation function plays a major role (filling the gaps) inside the region of interest by correlating the scalar speed measurements to their surroundings, the absence of such a correlation function leads to an undetermined system, even inside the measurement area. Moreover, the accuracy outside can be also enhanced by considering a more realistic mean flow, as will be discussed in § 7.

6.3. Effect of the time horizon (HGW model)

In this section, we investigate the influence of the time horizon on the accuracy of reconstruction. We replicate the analysis presented in § 6.1 for two distinct windows characterized by different assimilation horizons T_p using the HGW model. The first window spans a time interval of 200 s, commencing at $t_i = t_f - 200$ s. Meanwhile, the second window covers a wider period of 400 s, initiating at $t_i = t_f - 400$ s. Here, t_f represents a specific time instant. Increasing the time horizon further leads to divergence of the adjoint solution due to the chaotic nature of turbulence. Therefore, we limit our analysis to these two windows. Figure 14(f–h) show three snapshots of the reference field at $t = t_f - 400$ s, $t = t_f - 200$ s and $t = t_f$, respectively. The reconstructed field for the short and long assimilation windows is shown in figure 14(a,b) and 14(c–e), respectively. Moreover, the instantaneous reconstruction error is shown in figure 14(i–k) and 14(l,m) for the long and short assimilation windows, respectively.

At $t = t_f - 200$ s, it can be clearly seen that the reconstruction accuracy is significantly increased when the time instant is located at the middle of the assimilation window, as seen in figure 14(j), compared with the initial instant for the short window in figure 14(l). On the other hand, the accuracy at the final instant $t = t_f$ for both windows in figure 14(k,m) are

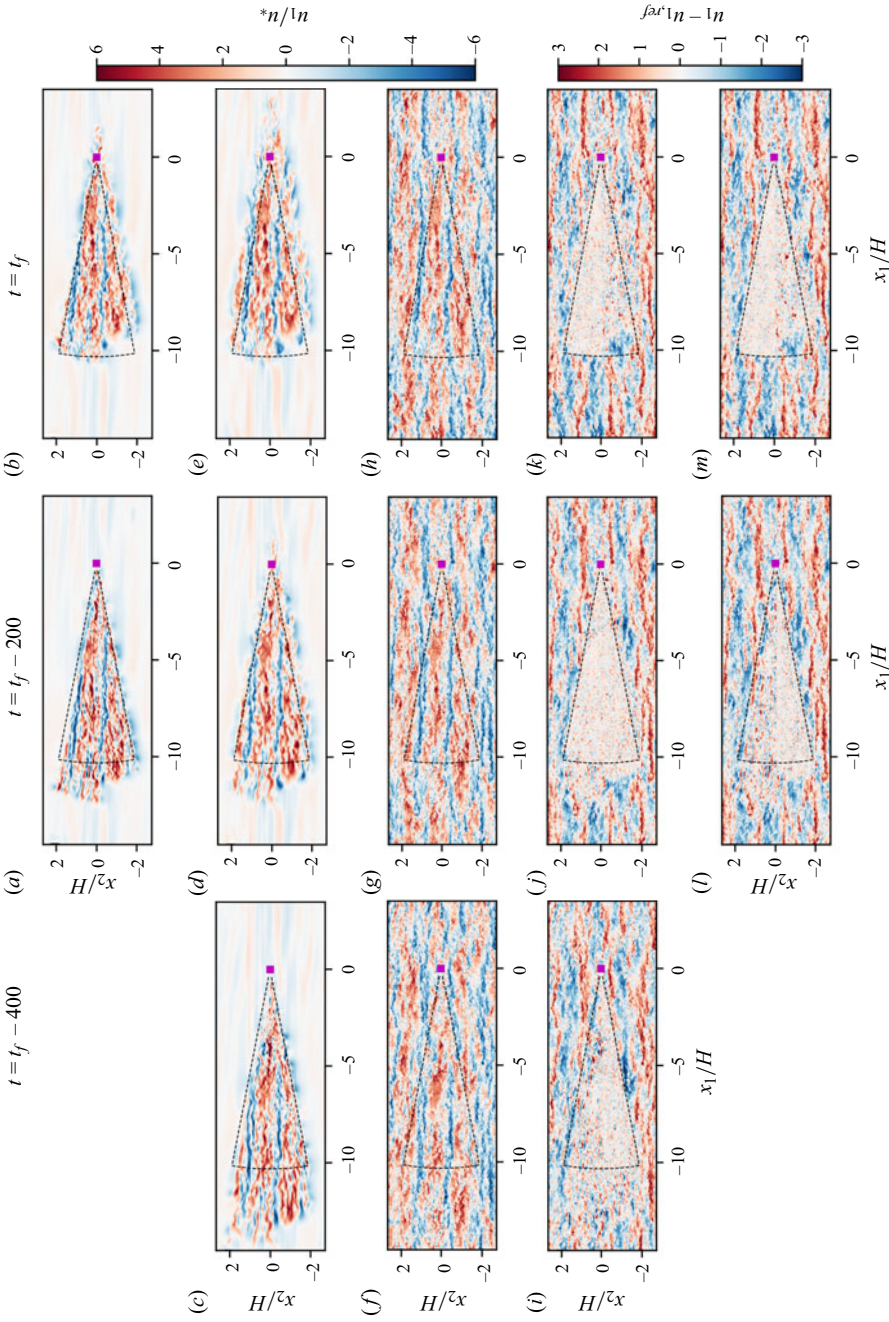


Figure 14. The reconstructed streamwise velocity field at $x_3 = 0.1H$ using the HGW model at $t = t_f$ and $t = t_f - 200$ s (*a, b*) for an assimilation window of $T_s = 200$ s, and at $t = t_f - 400$ s, $t = t_f - 200$ s and $t = t_f$ for an assimilation window of $T_s = 400$ s (*c–e*). The corresponding instantaneous reconstruction error is also shown in (*i–m*). In (*f–h*) snapshots of the reference field are shown.

seen to be nearly the same as each other, and lower when compared with the reconstruction of the long window at $t = t_f - 200$ s.

7. Conclusion and future perspectives

In this paper, we reconstructed the turbulent flow field in a neutrally stratified ABL using a strong 4D-Var assimilation approach combined with an LES model. Whilst this problem was tackled before by Bauweraerts & Meyers (2020), the focus in the current study was on providing a more practical, fast and cost-effective assimilation algorithm. To this end, we formulated our problem in a solenoidal space to enforce the continuity equation and regularized it using two fast analytical approximations for the background covariance tensor. The problem was then preconditioned using the analytical tensor to increase the convergence rate, which resulted in a super-linear convergence to a local minimum.

The error study revealed that regularizing the MAP problem with either the HGW or the Mann model (with Saffman spectrum) yields favourable reconstruction accuracy. However, it is worth noting that the HGW model has limitations in capturing the shear profile, leading to errors in the vertical direction close to the lidar plane when compared with the Mann model. This limitation may become particularly significant in scenarios where there is strong shear present, but in the current work the difference between the HGW and Mann (with Saffman) models was rather small. Furthermore, for the Mann model, the error study revealed the sensitivity of the reconstruction accuracy to the length of the correlation function in the streamwise direction. The Saffman energy spectrum, characterized by a slower decay as $k \rightarrow 0$, allows for the representation of longer structures when employed in the Mann model. In contrast, the Batchelor case exhibits a faster decay. The ability to represent longer structures with the Saffman spectrum can be advantageous in capturing larger-scale features in the flow.

In the current work, we also studied the effect of increasing the assimilation time horizon on the reconstruction accuracy. The study showed that in a strong 4D-Var method, the best reconstruction accuracy is obtained at the middle of the assimilation window rather than the terminal time, which can be attributed to the fact that the error history of the MAP estimate is not included in the case of the strong 4D-Var. Formulating a weak 4D-Var such that the maximum accuracy is obtained at the final time can be of particular significance, especially in applications such as optimal control.

In this study, we limited our analysis to the neutrally stratified ABL. Extending the analysis to other stability conditions should be possible. For instance, the modified Mann model in Chougule *et al.* (2018) can be used to regularize the temperature fluctuations alongside the velocity fluctuations in a stable ABL. Furthermore, it is possible to relax the assumption of a known mean flow by estimating both the initial mean profile and the fluctuation field. However, this estimation process requires obtaining the statistical properties of the mean flow, including its direction and magnitude, in addition to the turbulent part. One approach to achieve this is by, for example, using a hierarchical Bayesian analysis, in which the mean flow is first included (e.g. using weather models) and the turbulent field is then included at a lower level.

The focus of the current study was on assessing the possibility of directly applying the HGW and Mann models within a reconstruction algorithm. However, further development of an analytical model that is able to more precisely represent the significant structural characteristics such as the meandering behaviour can be valuable (Hutchins & Marusic 2007).

Finally, extending the time horizon in this analysis revealed promising results that can be further explored by further increasing the time horizon. While this is not possible using

the current single shooting approach due to the chaotic divergence of the LES model, novel techniques such as multiple shooting can be used for this purpose.

Acknowledgements. The first author thanks L. Fang for insightful discussions.

Funding. The authors acknowledge support from the KU Leuven research council, grant number C16/17/008, and from the project BeFORECAST, funded by the Energy Transition Funds of the Federal Public Service Economy of the Belgian Federal Government.

Declaration of interests. The authors report no conflict of interest.

Author ORCIDs.

📍 Ahmed Alreweny <https://orcid.org/0000-0002-4527-3702>;

📍 Stefan Vandewalle <https://orcid.org/0000-0002-8988-2374>;

📍 Johan Meyers <https://orcid.org/0000-0002-2828-4397>.

Appendix A. State space model

A.1. Governing equations

The neutrally stratified ABL flow considered in this manuscript is modelled through the incompressible Navier–Stokes equations and solved using a standard LES strategy. This model was described in detail in many previous articles (e.g. Munters, Meneveau & Meyers 2016; Bauweraerts & Meyers 2020). Nevertheless, it will be briefly discussed here for completeness. The filtered incompressible Navier–Stokes momentum and continuity equations are given as

$$\left. \begin{aligned} \frac{\partial \tilde{\mathbf{u}}}{\partial t} + (\tilde{\mathbf{u}} \cdot \nabla) \tilde{\mathbf{u}} &= -\frac{1}{\rho} \nabla(p_\infty + \tilde{p}) - \nabla \cdot \boldsymbol{\tau}_{sgs}, \\ \nabla \cdot \tilde{\mathbf{u}} &= 0, \end{aligned} \right\} \quad (\text{A1})$$

where $\tilde{\mathbf{u}}$ is the filtered velocity at the grid scale, ∇p_∞ is the mean pressure gradient given as $\nabla p_\infty = [-u_*^2 H^{-1}, 0, 0]$ and \tilde{p} is the filtered pressure fluctuation. The standard constant-coefficient Smagorinsky model (Smagorinsky 1963) is used to model the deviatoric part of the subgrid-scale stress tensor with $C_s = 0.14$, combined with wall damping (Mason & Thomson 1992) to avoid excessive dissipation of kinetic energy with $n = 1$. The trace of the stress tensor $\boldsymbol{\tau}_{sgs}$ is absorbed into the filtered pressure \tilde{p} as a common practice in incompressible LES.

In this paper, all simulations are done on a computational domain that has periodic boundary conditions in the horizontal plane. Therefore, the domain length is made sufficiently large to prevent any boundary effects on the solution in the domain of interest. In the vertical direction, an impermeable wall stress boundary condition is used following Bou-Zeid, Meneveau & Parlange (2005). At the top boundary, the vertical motion is blocked by an impermeable slip boundary condition. The governing equations are discretized using the Fourier pseudospectral method described in § A.2 and solved using our in-house LES code SP-Wind (Meyers & Sagaut 2007; Munters *et al.* 2016).

A.2. Discretization

Using the horizontal homogeneity assumption and the truncated discrete Fourier analysis, the components of the discrete velocity vector \mathbf{U} defined in § 2 can be expressed as

$$U_{mn}^j = \sum_{P_{min}=-N_1/2}^{P_{max}=(N_1/2-1)} \sum_{Q_{min}=-N_2/2}^{Q_{max}=(N_2/2-1)} \hat{U}_{pq}^j \exp\left(i2\pi \left(\frac{pn}{N_1} + \frac{qm}{N_2}\right)\right), \quad (\text{A2})$$

where U_{mn}^j has a length of N_3 when $j = 1, 2$ and $N_3 - 1$ when $j = 3$. Here, m and n are indices corresponding to horizontal spatial locations such that $x_1 = mL_1/N_1$ and $x_2 = nL_2/N_2$. Further, \hat{U}_{pq}^j is the complex-valued Fourier coefficient of the Fourier mode with streamwise wavenumber $k_1 = 2\pi p/L_1$ and spanwise wavenumber $k_2 = 2\pi q/L_2$. Rewriting the latter equation in matrix form for the three velocity components, including all grid points in the vertical direction, gives

$$\left. \begin{aligned} \begin{bmatrix} U_{mn}^1(t) \\ U_{mn}^2(t) \\ U_{mn}^3(t) \end{bmatrix} &= \begin{bmatrix} \mathcal{W}^1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{W}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{W}^3 \end{bmatrix} \begin{bmatrix} \hat{U}^1(t) \\ \hat{U}^2(t) \\ \hat{U}^3(t) \end{bmatrix}, \\ U_{mn}(t) &= \mathcal{W}\hat{U}(t), \end{aligned} \right\} \quad (\text{A3})$$

where $\hat{U}^j = [\hat{U}_{p_{min}q_{min}}^j, \dots, \hat{U}_{p_{max}q_{max}}^j]^T$ and \mathcal{W}^j is a block matrix of inverse discrete Fourier transform sub-matrices in the horizontal plane and has a size of $N_1N_2N_3 \times N_1N_2N_3$ when $j = 1, 2$ and $N_1N_2(N_3 - 1) \times N_1N_2(N_3 - 1)$ when $j = 3$. The second line of (A3) is a compact representation of the discretization.

Appendix B. Adjoint model

As already mentioned in § 4, the gradient in the current study is obtained using the continuous adjoint method derived in Goit & Meyers (2015) and Bauweraerts & Meyers (2020). The resulting adjoint equation can be written as

$$\left. \begin{aligned} -\frac{\partial \boldsymbol{\zeta}}{\partial t} &= \tilde{\mathbf{u}} \cdot \nabla \boldsymbol{\zeta} + \boldsymbol{\zeta} \cdot \nabla \tilde{\mathbf{u}} + \nabla \cdot \boldsymbol{\tau}_{SGS}^* + \frac{1}{\rho} \nabla \pi + \sum_{i=1}^{N_r} \mathbf{f}_i, \\ \nabla \cdot \boldsymbol{\zeta} &= 0, \end{aligned} \right\} \quad (\text{B1})$$

with

$$\mathbf{f}_i = \frac{1}{\gamma^2 T_s} \sum_{n=1}^{N_s} (y_{n,i} - h_{n,i}) \mathcal{G}(\mathcal{Q}(t)(\mathbf{x} - \mathbf{x}_i(t))) \mathbf{e}_l(t) H\left(\frac{T_s}{2} - |t - t_{n-1/2}|\right), \quad (\text{B2})$$

where $\boldsymbol{\zeta}$ and π are the adjoint velocity and pressure, respectively. Further, $\boldsymbol{\tau}_{SGS}^*$ is the adjoint subgrid-scale model. The adjoint equation is discretized and solved backward in time similarly to the forward model, starting from the terminal condition $\boldsymbol{\zeta}(\mathbf{x}, t_f) = 0$. The adjoint solution at the beginning of the assimilation window $\boldsymbol{\zeta}(\mathbf{x}, 0)$ is then used to compute the gradient using (4.1).

Appendix C. Projection and reconstruction operations

In this section, the projection and reconstruction operators in § 2 are derived in the Fourier space. Assuming incompressibility, the continuity equation in the horizontal wavenumber space can be written as

$$0 = ik_1 \hat{U}_{pq}^1 + ik_2 \hat{U}_{pq}^2 + D_z \hat{U}_{pq}^3, \quad (\text{C1})$$

The LES-based reconstruction of turbulence

where D_z is a discretization matrix in the vertical direction of size $N_3 \times (N_3 - 1)$. Similarly, the vertical component of the vorticity is given as

$$\hat{\Omega}_{pq}^3 = -ik_2 \hat{U}_{pq}^1 + ik_1 \hat{U}_{pq}^2. \tag{C2}$$

Writing (C2) for all wavenumbers (except the DC mode, i.e. $(k_1, k_2) = (0, 0)$) in matrix form, we get

$$\left. \begin{aligned} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{B} & \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{U}^1 \\ \hat{U}^2 \\ \hat{U}^3 \end{bmatrix} &= \begin{bmatrix} \hat{U}^3 \\ \hat{\Omega}^3 \end{bmatrix}, \\ \hat{\mathcal{P}} \hat{U} &= \hat{\Theta}, \end{aligned} \right\}, \tag{C3}$$

with

$$\hat{\Omega}^3 = \left[\hat{\Omega}_{p_{min}q_{min}}^3, \dots, \hat{\Omega}_{p_{max}q_{max}}^3 \right]^T \quad \forall (p, q) \neq (0, 0), \tag{C4}$$

where $\hat{\mathcal{P}}$ is a linear projection operator and

$$\mathbf{B} = \begin{bmatrix} -ik_{2,min} \mathbf{I}_{N_3 \times N_3} & \dots & \mathbf{0}_{N_3 \times N_3} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{N_3 \times N_3} & \dots & -ik_{2,max} \mathbf{I}_{N_3 \times N_3} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} ik_{1,min} \mathbf{I}_{N_3 \times N_3} & \dots & \mathbf{0}_{N_3 \times N_3} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{N_3 \times N_3} & \dots & ik_{1,max} \mathbf{I}_{N_3 \times N_3} \end{bmatrix}. \tag{C5a,b}$$

To retrieve the velocity field, the continuity equation (C1) is added to the system (C3) and then the system is inverted to obtain the inverse transformation (for $(k_1, k_2) \neq (0, 0)$)

$$\hat{U} = \hat{\mathcal{R}} \hat{\Theta}, \tag{C6}$$

with $\hat{\mathcal{R}}$ the reconstruction operator

$$\hat{\mathcal{R}} = \frac{1}{\kappa_i^2} \begin{bmatrix} \mathbf{E} & \mathbf{B} \\ \mathbf{F} & \mathbf{C} \\ \kappa_i^2 \mathbf{I} & \mathbf{0} \end{bmatrix}, \tag{C7}$$

where $\kappa_i = (k_{1,i}^2 + k_{2,i}^2)^{1/2}$ and

$$\mathbf{E} = \begin{bmatrix} -ik_{1,min} \mathbf{D}_z & \dots & \mathbf{0}_{N_3 \times N_3 - 1} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{N_3 \times N_3 - 1} & \dots & -ik_{1,max} \mathbf{D}_z \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} -ik_{2,min} \mathbf{D}_z & \dots & \mathbf{0}_{N_3 \times N_3 - 1} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{N_3 \times N_3 - 1} & \dots & -ik_{2,max} \mathbf{D}_z \end{bmatrix}. \tag{C8a,b}$$

Using the transformation variable $\hat{\Theta}$ instead of the velocity \hat{U} in the MAP problem allows us to eliminate the divergence-free constraint from the optimization problem, since the reconstructed velocity using (C6) is solenoidal by definition. Since the projection and reconstruction operators are not defined for the DC mode (which corresponds to the spatial mean in physical space), this mode is given explicitly in the optimization algorithm.

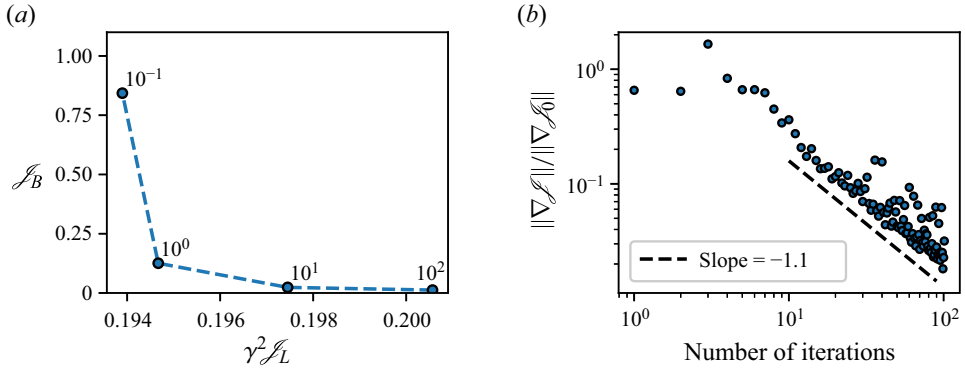


Figure 15. Pareto front (a) and the convergence of the relative gradient using the simple ridge regularization (b).

Appendix D. Analytical IFT of the homogeneous tensor

In this section, the analytical IFT of the tensor $\hat{\Phi}_{ij}^{iso}(\mathbf{k})$ in (3.3) is calculated for the Batchelor turbulence case in the direction of k_3 . These calculations are available in the original paper of Wilson (1997) and are repeated here only for completeness. The relevant parts of the spectrum are

$$\left. \begin{aligned}
 \hat{R}_{11}^{iso}(k_1, k_2; r_3) &= \frac{\sigma_{iso}^2 \ell^2 \zeta_h^{\nu+1}}{\pi 2^\nu \Gamma(\nu) (1 + \kappa^2 \ell^2)^{\nu+1}} \\
 &\times \left[\left(\nu + \frac{3}{2} \right) K_{\nu+1}(\zeta_h) - \frac{\zeta_h (1 + k_1^2 \ell^2)}{2(1 + \kappa^2 \ell^2)} K_{\nu+2}(\zeta_h) \right], \\
 \hat{R}_{33}^{iso}(k_1, k_2; r_3) &= \frac{\sigma_{iso}^2 \kappa^2 \ell^4 \zeta_h^{\nu+2}}{\pi 2^{\nu+1} \Gamma(\nu) (1 + \kappa^2 \ell^2)^{\nu+2}} K_{\nu+2}(\zeta_h), \\
 \hat{R}_{12}^{iso}(k_1, k_2; r_3) &= -\frac{\sigma_{iso}^2 k_1 k_2 \ell^4 \zeta_h^{\nu+2}}{\pi 2^{\nu+1} \Gamma(\nu) (1 + \kappa^2 \ell^2)^{\nu+2}} K_{\nu+2}(\zeta_h), \\
 \hat{R}_{13}^{iso}(k_1, k_2; r_3) &= -\frac{i \sigma_{iso}^2 k_1 \ell^3 \zeta_h^{\nu+2}}{\pi 2^{\nu+1} \Gamma(\nu) (1 + \kappa^2 \ell^2)^{\nu+3/2}} K_{\nu+1}(\zeta_h),
 \end{aligned} \right\} \quad (D1)$$

where $\nu = 1/3$, $\zeta_h = (r_3/\ell)\sqrt{1 + \kappa^2 \ell^2}$ and K_ν is the modified Bessel function of the second kind of order ν .

Appendix E. Ridge regression results

In this section, we study the reconstruction result, in analogy to § 6, for the ridge regression case (i.e. when $\mathbf{B} = \mathcal{P}\mathcal{P}^T$ in (2.4), leading to $\mathbf{R} = \mathbf{I}$). Similar to the other models considered in this paper (see figure 7 and § 5.4), we use the Pareto front technique to fix the parameter γ^2 , which results in a value of $\gamma^2 = 1$ as shown in figure 15(a). Afterwards, the optimization problem is solved to reconstruct the turbulent flow field with an identical set-up to that presented in § 6. Figure 15(b) shows the convergence of the optimization algorithm, which is slower than the convergence observed for the other models in this paper (i.e. slope -1.1 versus -1.4).

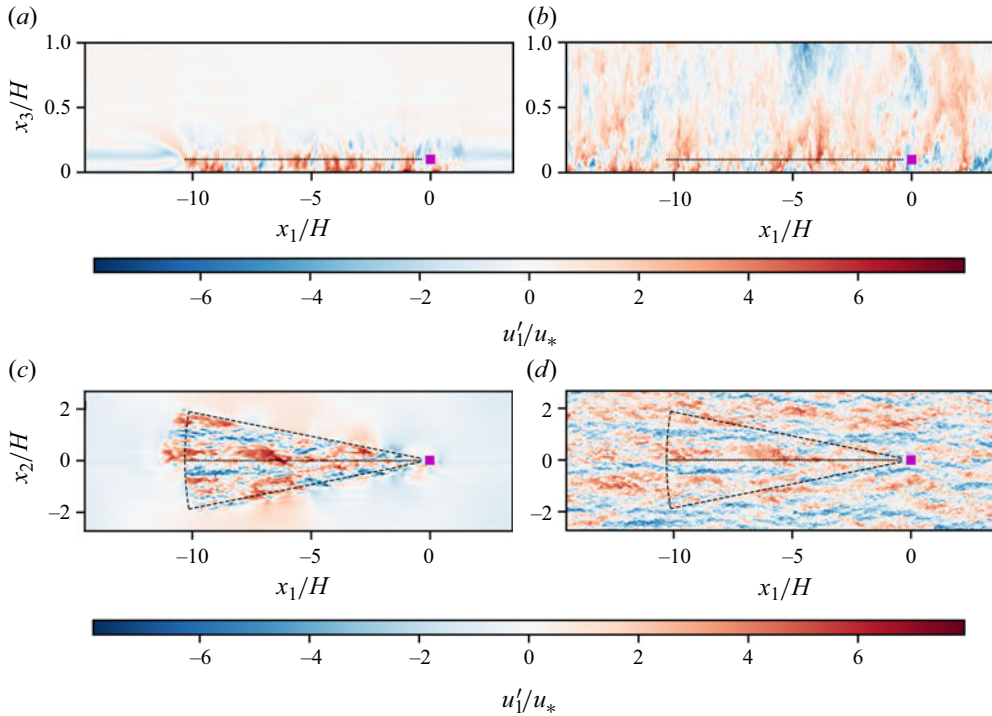


Figure 16. The streamwise velocity fluctuation component at the $x_2 = 0$ (a,b) and $x_3 = 0.1H$ (c,d) planes. The right column is the reference velocity and the left column is the reconstructed velocity using the simple ridge regression. The results are shown at $t = 0.05u_* / H$.

Examining the reconstructed field in figure 16 at $t = 0.05u_* / H$ reveals that the direct application of the simple ridge regression in this example leads to acceptable reconstruction results only at the measurement locations. However, as soon as we move outside the scanning region (beyond the convection effects), the reconstruction is solely done using the mean flow (cf. figure 12). This can be particularly seen, for example, in the vertical direction.

Finally, we note that ridge regression may be alternatively formulated by directly using $\mathbf{B} = \mathbf{I}$. We have also investigated this approach. However, we saw that it leads to consistently worse results compared with $\mathbf{R} = \mathbf{I}$ (not further shown here).

REFERENCES

- AITKEN, M.L., BANTA, R.M., PICHUGINA, Y.L. & LUNDQUIST, J.K. 2014 Quantifying wind turbine wake characteristics from scanning remote sensor data. *J. Atmos. Ocean. Technol.* **31** (4), 765–787.
- ALCAYAGA, L., LARSEN, G.C., KELLY, M. & MANN, J. 2020 Large-scale coherent structures in the atmosphere over a flat terrain. *J. Phys.: Conf. Ser.* **1618**, 062030.
- BANNISTER, R.N. 2017 A review of operational methods of variational and ensemble-variational data assimilation. *Q. J. R. Meteorol. Soc.* **143** (703), 607–633.
- BAUWERAERTS, P. & MEYERS, J. 2020 Reconstruction of turbulent flow fields from lidar measurements using large-eddy simulation. *J. Fluid Mech.* **906**, A17.
- BOU-ZEID, E., MENEVEAU, C. & PARLANGE, M. 2005 A scale-dependent Lagrangian dynamic model for large eddy simulation of complex turbulent flows. *Phys. Fluids* **17** (2), 1–18.
- BYRD, R.H., LU, P., NOCEDAL, J. & ZHU, C. 1995 A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16** (5), 1190–1208.

- CHAI, T. & LIN, C.L. 2003 Estimation of turbulent viscosity and diffusivity in adjoint recovery of atmospheric boundary layer flow structures. *Multiscale Model. Simul.* **1** (2), 196–220.
- CHAI, T., LIN, C.L. & NEWSOM, R.K. 2004 Retrieval of microscale flow structures from high-resolution Doppler lidar data using an adjoint model. *J. Atmos. Sci.* **61** (13), 1500–1520.
- CHEN, Y., GUO, F., SCHLIPF, D. & CHENG, P.W. 2022 Four-dimensional wind field generation for the aeroelastic simulation of wind turbines with lidars. *Wind Energy Sci.* **7** (2), 539–558.
- CHOUGULE, A., MANN, J., KELLY, M. & LARSEN, G.C. 2018 Simplification and validation of a spectral-tensor model for turbulence including atmospheric stability. *Boundary-Layer Meteorol.* **167**, 371–397.
- DAVIDSON, P.A. 2010 On the decay of Saffman turbulence subject to rotation, stratification or an imposed magnetic field. *J. Fluid Mech.* **663**, 268–292.
- FANG, J. & PORTÉ-AGEL, F. 2015 Large-eddy simulation of very-large-scale motions in the neutrally stratified atmospheric boundary layer. *Boundary-Layer Meteorol.* **155** (3), 397–416.
- GILLING, L. 2009 TuGen: synthetic turbulence generator, manual and user's guide. *DCE Tech. Rep.* 76. Department of Civil Engineering, Aalborg University.
- GOIT, J.P. & MEYERS, J. 2015 Optimal control of energy extraction in wind-farm boundary layers. *J. Fluid Mech.* **768**, 5–50.
- GUO, F., MANN, J., PEÑA, A., SCHLIPF, D. & CHENG, P.W. 2022 The space-time structure of turbulence for lidar-assisted wind turbine control. *Renew. Energy* **195**, 293–310.
- GUSTAFSSON, N., *et al.* 2018 Survey of data assimilation methods for convective-scale numerical weather prediction at operational centres. *Q. J. R. Meteorol. Soc.* **144** (713), 1218–1256.
- HOLMES, P., LUMLEY, J.L. & BERKOOZ, G. 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge University Press.
- HUNT, J.C.R. & GRAHAM, J.M.R. 1978 Free-stream turbulence near plane boundaries. *J. Fluid Mech.* **84** (2), 209–235.
- HUTCHINS, N. & MARUSIC, I. 2007 Evidence of very long meandering features in the logarithmic region of turbulent boundary layers. *J. Fluid Mech.* **579**, 1–28.
- IUNGO, G.V., WU, Y.T. & PORTÉ-AGEL, F. 2013 Field measurements of wind turbine wakes with lidars. *J. Atmos. Ocean. Technol.* **30** (2), 274–287.
- KAIMAL, J.C., WYNGAARD, J.C., IZUMI, Y. & COTÉ, O.R. 1972 Spectral characteristics of surface-layer turbulence. *Q. J. R. Meteorol. Soc.* **98** (417), 563–589.
- KRISHNAMURTHY, R., CHOUKULKAR, A., CALHOUN, R., FINE, J., OLIVER, A. & BARR, K.S. 2013 Coherent Doppler lidar for wind farm characterization. *Wind Energy* **16** (2), 189–206.
- KRISTENSEN, L., LENSCHOW, D.H., KIRKEGAARD, P. & COURTNEY, M. 1989 The spectral velocity tensor for homogeneous boundary-layer turbulence. *Boundary-Layer Meteorol.* **47** (1–4), 149–193.
- LE DIMET, F.-X., NAVON, I.M. & ȘTEFĂNESCU, R. 2017 Variational data assimilation: optimization and optimal control. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, vol. III, pp. 1–53. Springer International Publishing.
- LEE, M.J. & HUNT, J.C.R. 1989 The structure of sheared turbulence near a plane boundary. In *Proceedings of the 7th Symposium on Turbulent Shear Flows*, vol. 1, pp. 8.1.1–8.1.6. Springer.
- LIN, C.L., CHAI, T. & SUN, J. 2001 Retrieval of flow structures in a convective boundary layer using an adjoint model: identical twin experiments. *J. Atmos. Sci.* **58** (13), 1767–1783.
- LORENC, A.C. 1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112** (474), 1177–1194.
- MANN, J. 1994 The spatial structure of neutral atmospheric surface-layer turbulence. *J. Fluid Mech.* **273**, 141–168.
- MANN, J. 1998 Wind field simulation. *Prob. Engng Mech.* **13** (4), 269–282.
- MARUSIC, I. & HEUER, W.D.C. 2007 Reynolds number invariance of the structure inclination angle in wall turbulence. *Phys. Rev. Lett.* **99**, 114504.
- MASON, P.J. & THOMSON, D.J. 1992 Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.* **242** (28), 51–78.
- MEYERS, J. & SAGAUT, P. 2007 Evaluation of Smagorinsky variants in large-eddy simulations of wall-resolved plane channel flows. *Phys. Fluids* **19** (9), 095105.
- MORÉ, J.J. & THUENTE, D.J. 1994 Line search algorithms with guaranteed sufficient decrease. *ACM Trans. Math. Softw.* **20** (3), 286–307.
- MUNTERS, W., MENEVEAU, C. & MEYERS, J. 2016 Shifted periodic boundary conditions for simulations of wall-bounded turbulent flows. *Phys. Fluids* **28** (2), 025112.

The LES-based reconstruction of turbulence

- NEWSOM, R.K. & BANTA, R.M. 2004 Assimilating coherent Doppler lidar measurements into a model of the atmospheric boundary layer. Part I: algorithm development and sensitivity to measurement error. *J. Atmos. Ocean. Technol.* **21** (9), 1328–1345.
- NEWSOM, R.K., LIGON, D., CALHOUN, R., HEAP, R., CREGAN, E. & PRINCEVAC, M. 2005 Retrieval of microscale wind and temperature fields from single- and dual-Doppler lidar data. *J. Appl. Meteorol.* **44** (9), 1324–1344.
- NGUYEN, C.V. & SOULHAC, L. 2021 Data assimilation methods for urban air quality at the local scale. *Atmos. Environ.* **253**, 118366.
- NOCEDAL, J. & WRIGHT, S.J. 2006 *Numerical Optimization*, 2nd edn. Springer.
- PELTIER, L.J., WYNGAARD, J.C., KHANNA, S. & BRASSEUR, J.G. 1996 Spectra in the unstable surface layer. *J. Atmos. Sci.* **53** (1), 49–61.
- PEÑA, A., *et al.* 2013 Remote sensing for wind energy. *Tech. Rep.* DTU Wind Energy-E-Report-0029(EN). DTU Wind Energy, Technical University of Denmark.
- POPE, S.B. 2000 *Turbulent Flows*. Cambridge University Press.
- SABALE, A.K. & GOPAL, N.K.V. 2019 Nonlinear aeroelastic analysis of large wind turbines under turbulent wind conditions. *AIAA J.* **57** (10), 4416–4432.
- SAFFMAN, P.G. 1967 The large-scale structure of homogeneous turbulence. *J. Fluid Mech.* **27** (3), 581–593.
- SILLERO, J.A., JIMÉNEZ, J. & MOSER, R.D. 2014 Two-point statistics for turbulent boundary layers and channels at Reynolds numbers up to $\delta^+ \approx 2000$. *Phys. Fluids* **26** (10), 105109.
- SMAGORINSKY, J. 1963 General circulation experiments with the primitive equations: I. The basic experiment. *Mon. Weath. Rev.* **91** (3), 99–164.
- STUART, A.M. 2010 Inverse problems: a Bayesian perspective. *Acta Numerica* **19**, 451–559.
- STULL, R.B. 1988 *An Introduction to Boundary Layer Meteorology*. Kluwer Academic Publishers.
- SUN, J., FLICKER, D.W. & LILLY, D.K. 1991 Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci.* **48** (6), 876–890.
- WILSON, D.K. 1997 A three-dimensional correlation/spectral model for turbulent velocities in a convective boundary layer. *Boundary-Layer Meteorol.* **85** (1), 35–52.
- WILSON, D.K. 1998 Anisotropic turbulence models for acoustic propagation through the neutral atmospheric surface layer. *Tech. Rep.* ARL-TR-1519. US Army Research Laboratory.
- XIA, Q., LIN, C.L., CALHOUN, R. & NEWSOM, R.K. 2008 Retrieval of urban boundary layer structures from Doppler lidar data. Part I: accuracy assessment. *J. Atmos. Sci.* **65** (1), 3–20.