Tellus **000**, 000–000 (0000) Printed May 16, 2012 (Tellus LaTeX style file v2.2)

Power laws and inverse motion modeling: application to turbulence measurements from satellite images

By Patrick Héas[1]*, Etienne Mémin[1], Dominique Heitz[2], P.D. Mininni[3]
[1]*INRIA Bretagne Atlantique Research Center - Campus de Beaulieu - 35042 Rennes cedex - France*
[2]*CEMAGREF - Avenue de Cucillé- 35044 Rennes cedex - France*
[3]*Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and IFIBA, CONICET, Ciudad Universitaria, 1428 Buenos Aires, Argentina, and NCAR, PO Box 3000, Boulder, Colorado 80307-3000 USA.*

May 16, 2012

ABSTRACT

In the context of tackling the ill-posed inverse problem of motion estimation from image sequences, we propose to introduce prior knowledge on flow regularity given by turbulence statistical models. Prior regularity is formalized using turbulence power laws describing statistically self-similar structure of motion increments across scales. The motion estimation method minimizes the error of an image observation model while constraining second order structure function to behave as a power law within a prescribed range. Thanks to a Bayesian modeling framework, the motion estimation method is able to jointly infer the most likely power law directly from image data. The method is assessed on velocity fields of 2D or quasi-2D flows. Estimation accuracy is first evaluated on a synthetic image sequence of homogeneous and isotropic 2D turbulence. Results obtained with the approach based on physics of fluids outperforms state-of-the-art. Then, the method analyzes atmospheric turbulence using a real meteorological image sequence. Selecting the most likely power law model enables the recovery of physical quantities which are of major interest for turbulence atmospheric characterization. In particular, from meteorological images we are able to estimate energy and enstrophy fluxes of turbulent cascades, which are in agreement with previous in situ measurements.

1 Introduction

Images constitute important data for studying geophysical flows since they can characterize a large range of spatial scales in comparison to sparse information provided by standard point measurement techniques.

Correlation-based techniques are common inverse approaches to extract sparse velocity fields from image sequences. These methods penalize deviations from constant flows over a spatial window and from the brightness preservation of the tracer along its trajectory (Leese et al., 1971). Such methods have demonstrated their robustness and accuracy for velocity measurement in experimental studies (Adrian, 1991). However, correlation is not a physical-based image observation model, and is particularly limited in the context of geophysical flows (*e.g.*, because of small scale intermittency and three-dimensional effects). Moreover, the assumption of local constant motion gives rise to an intrinsic measurement scale bound associated to the correlation window size and an artificial regularity of the flow. In particular, for low-contrast observations such as satellite images, correlation-based methods are considerably limited and provide only very sparse motion fields.

In computer vision, estimating the projected apparent motion of a 3D scene onto the image plane, refereed in the literature as optical-flow, has been an intensive subject of research since the 80's and the work of Horn and Schunck (1981). The inverse motion modeling proposed in this seminal work is formulated in a differential framework enabling the easy introduction of a physical-based direct observation model (relying on brightness preservation, mass conservation, scalar transport, etc.) adapted to the particular image data at hand (Fitzpatrick, 1988; Heas et al., 2007a; Heas and Memin, 2008; Liu and Shen, 2008). The model links the motion field to the image intensity function. However, these models are insufficient for fluid flow estimation since they are ill-conditioned, as they are generally constituted at each point of a scalar equation for two unknowns. The introduction of regularization constraints is required to remove the motion ambiguities and achieve inversion. Common spatial regularizers may introduce measurement inaccuracy since they impose in a small spatial neighborhood a *prior* smoothness which may improperly describe the regularity followed by fluid flows. Moreover, they strongly depend on a tuning parameter weighting the amount of smoothing required to cure the ill-posed nature of the inverse motion problem. For

* Corresponding author.
e-mail: patrick.heas@inria.fr

large values of this parameter (which is almost unavoidable when facing noisy images or with low photometric contrast), the solution tends to be too smooth and presents sometimes lack of accuracy at large scales. Alternatively, dynamical constraints embedded in variational assimilation frameworks have recently been proposed for geophysical flow estimation (Papadakis and Memin, 2008; Corpetti et al., 2009). Unfortunately, these method rely only on a large scale dynamical model, which is not valid at the pixel level.

This work aims at showing how it is possible to alleviate these limitations in coupling two main ingredients: turbulence statistical modeling and Bayesian inference. In the one hand, important advances have been done in the statistical modeling of turbulence since the precursor work of Kolmogorov in 1941 (see e.g., Kraichnan (1967); Frisch (1995); Kurien et al. (2000); Falkovich et al. (2001); Lindborg and Cho (2001)). In particular, it has been shown from the Navier-Stokes equations for different flows that turbulent motion regularity can be characterized using some (assumed universal) scaling properties of the Probability Distribution Function (PDF) of motion increments. In the other hand, Bayesian modeling enables the reconstruction of complex physical processes in the presence of incomplete observations and prior geophysical scalings (Berliner et al., 1999; Wikle et al., 2001; Hoar et al., 2003). Moreover, this framework results in the design of non-parametrical methods and provides reliable model selection tools given some data (Gull, 1989; MacKay, 1992; Jaynes, 2003; Robert, 2007).

To be more precise, in this work we propose a method for motion estimation and turbulence characterization which exploits simultaneously two ideas: the use of turbulence scaling laws for motion regularization in optic-flow inverse problems; and the selection by Bayesian *evidence* maximization of the most appropriate power law model describing the image intensity function. The resulting frame-to-frame regularization is built from the physics of fluids, and involves neither dynamical constraints nor an adjoint model within a variational assimilation framework. It is intrinsically multiscale as it controls the solutions within a prescribed scale range. It is also non-parametric in the sense that it does not involve the tuning of any smoothing parameters. Finally, it allows as a by product the recovery of some important quantities for turbulence characterization such as energy flux across scales.

After validating the method using synthetic images from numerical simulations of 2D turbulence, we use the proposed regularization model to analyze data from meteorological image sequences. Recovered atmospheric quantities such as the energy and enstrophy flux are shown to be in good agreement with *in situ* measurements. Moreover, reconstructed second and third order structure functions and PDFs of velocity increments are consistent with previous analysis using aircraft data (Lindborg, 1999; Lindborg and Cho, 2001). We believe that those first results demonstrate the potential of image-based techniques in complementing other measurements for geophysical studies.

The paper is organized as follows. In section 2, we briefly overview optic-flow methods and review the motion scaling properties developed in theoretical works on turbulence. In section 3, self-similar regularizers are briefly introduced. Finally, the synthetic and real world experiments on meteorological image data are presented, followed by the conclusions. Details of the model, of the multiscale constraint minimization problem, and of the use of Bayesian *evidence* for the optimal selection of the *prior* self-similar model are given in the Appendices.

2 Related work

2.1 Optic-flow methods

This section briefly describes some solutions for the computation of velocity fields from image sequences. Interested readers may refer to an extended review (Heitz et al., 2010) for a more complete account of different techniques proposed in the literature and the description of some new trends based on variational assimilation processes.

2.1.1 Image observation model: Motion perceived through image intensity $I(\mathbf{x}, t)$ variations and the projection on the image plane of the velocity field $\mathbf{v} = (u, v)$ are identical when considering rigid motion and stable lighting conditions. In this situation, motion $\mathbf{v}$ respects the standard Optical Flow Constraint (OFC) equation which reads:

$$\frac{dI}{dt} = \frac{\partial I}{\partial t} + \nabla I \cdot \mathbf{v} = 0. \tag{1}$$

For fluids, this equation remains valid in the theoretical case of 2D incompressible flows. This case is relevant for many geophysical applications which are quasi-2D, as *e.g.*, in meteorology and oceanography (Pedlosky, 1987). For shallow-water flows visualized in a projected image plane, authors in Liu and Shen (2008) derived a physics-based image observation model using the scalar transport equation which reads:

$$\frac{\partial I}{\partial t} + \nabla I \cdot \mathbf{v} + I \nabla \cdot \mathbf{v} = \kappa \Delta I. \tag{2}$$

where κ denotes a diffusion coefficient. For layered geophysical flows, based on the mass conservation principle, a non-diffusive model analogous to (2) has been proposed in the literature to link the image intensity function I to a vertically averaged horizontal velocity field $\mathbf{v}$ (Heas et al., 2007a; Corpetti et al., 2009): However, all these observation models remain underconstrained, as they provide only one equation for two unknowns (u, v) at each spatio-temporal location $(\mathbf{x}, t)$.

2.1.2 Standard regularization constraints: To deal with this underconstrained estimation problem, an additional set of constraints have to be satisfied by the sought velocity field.

In the standard computer vision literature, dealing with frame-to-frame motion estimation in natural scenes, these constraints are usually chosen to enforce some spatial regularity of the solution. Global regularization schemes are convenient to enforce global coherence via local spatial dependencies. More precisely, estimation is performed through the minimization of an energy functional composed of two terms:

$$f(I, \mathbf{v}) = f_d(I, \mathbf{v}) + \alpha f_r(\mathbf{v}). \tag{3}$$

The first term, $f_d(I, \mathbf{v})$ is the observation model also called

data term. Its minimization penalizes discrepancies from the observation models. As an example, relying on the OFC equation and a quadratic penalization results in a data term[1]

$$f_d(I,\mathbf{v}) = \frac{1}{2}\int_\Omega \left(\frac{\partial I}{\partial t} + \mathbf{v}\cdot\nabla I\right)^2 d\mathbf{x}. \quad (4)$$

The second term, $f_r(\mathbf{v})$ (the "regularization term"), acts as a spatial *prior* enforcing the solution to follow some smoothness properties. The coefficient α (> 0) is a regularization parameter that balances between smoothness and global adequacy to the observation model. In this framework, Horn and Schunck (1981) proposed a first-order regularization of spatial components u and v of velocity field $\mathbf{v}$:

$$f_r(\mathbf{v}) = \frac{1}{2}\int_\Omega \left(||\nabla u||^2 + ||\nabla v||^2\right) d\mathbf{x} \quad (5)$$

However, motion gradient penalization is not well suited for fluid flows as it penalizes in a homogeneous way the curl and the divergence (Corpetti et al., 2002).

In the context of geophysical flows, second order spatial regularizers on motion vorticity and divergence have been proposed to overcome these limitations (Corpetti et al., 2002; Yuan et al., 2007; Heas et al., 2007a). It is important to outline that these spatial regularization constraints do not rely on physical kinematics properties. At best they mimic qualitative regularity properties of the physical system of interest, *i.e.*, coherent blobs of vorticity and divergence. Recently, Corpetti et al. (2009) proposed to rely on large scale dynamical constraints enforcing the temporal coherence of the solution, since in geophysics there exist some fair approximations of the physical equations governing the flows. The authors introduced a 4D variational assimilation technique to solve the motion estimation problem on a whole image sequence subject to layered shallow-water dynamical constraints. Unfortunately, this choice of constraints does not entirely remove the underdetermination of the estimation problem and still yields an infinite set of solutions. To close the problem, Corpetti et al. (2009) needed to rely on an additional local spatial regularization on motion measurements (Lucas and Kanade, 1981) in consistency with a filtered version of shallow water dynamics. By doing so, they restricted themselves to the determination of the solution at large scales (on a grid of a 100 kilometer resolution).

In summary, in frame-to-frame estimation or variational image assimilation, there is actually a lack of physical constraints to solve at small scales the underconstrained fluid motion estimation problem. However, as we will show in the following, turbulence statistical models can provide the needed closure.

[1] Let us outline that in the case of large displacements, the brightness constancy model (1) is generally replaced by its time-integrated form $I(\mathbf{x}+\mathbf{v}dt, t+dt) - I(\mathbf{x},t) = 0$ linearized around some initial solution $\mathbf{x}+\mathbf{v}dt$. For the sake of clarity we will not describe the time-integrated model further even if its application is in practice essential to recover long range velocity fields (see Heitz et al. (2010) and references therein).

2.2 Turbulence statistical modeling

Since experimental results at the beginning of the 20th century, and later Kolmogorov's theoretical works, turbulent motion increments have been known to be structured as nearly scale invariant spatial processes. In many turbulence models, a quantity of interest is the longitudinal velocity increment, which, in the direction of a unitary vector $\boldsymbol{n}_\theta$, is given by

$$\delta v_\parallel(\ell,\mathbf{x},\theta) = [\mathbf{v}(\mathbf{x}+\ell\boldsymbol{n}_\theta) - \mathbf{v}(\mathbf{x})]\cdot\boldsymbol{n}_\theta, \quad (6)$$

where the scalar ℓ represents a spatial increment. In turbulent studies homogeneity and isotropy are often assumed; that is to say, it is considered that statistical properties of the velocity field are invariant under translation of spatial location $\mathbf{x}$ and rotation of direction $\boldsymbol{n}_\theta$[2]. Consistently with these assumptions, the dependence on $\mathbf{x}$ and θ can be dropped and moments of order p of the PDF of velocity increments $p_\ell(\delta v_\parallel)$ can be approached by spatial integration:

$$\begin{aligned} S_p(\ell) &= \int \left[\delta v_\parallel(\ell)\right]^p p_\ell\left[\delta v_\parallel(\ell)\right] d\delta v_\parallel(\ell) \\ &\approx \frac{1}{2\pi|\Omega|}\int_\Omega\int_{[0,2\pi]} \left[\delta v_\parallel(\ell,\mathbf{x},\theta)\right]^p d\theta\, d\mathbf{x} \end{aligned} \quad (7)$$

in two dimensions, and where $|\Omega|$ denotes the spatial domain area. $S_p(\ell)$ are the so-called longitudinal structure functions.

For 3D isotropic turbulent flows, Kolmogorov (1941) demonstrated from the Navier-Stokes equations that the third order moment of the PDF $p_\ell(\delta v_\parallel)$, namely the third order structure function, is linear w.r.t. scale and follows the well-known "4/5 law": $S_3(\ell) = -\frac{4}{5}\epsilon\ell$ in the inertial range (see e.g., Frisch (1995) for details on this demonstration). The inertial range is defined as the range of scales $[\eta, \ell_0]$, where η represents the largest molecular dissipative scale and where ℓ_0 is smaller than the diameter L of the largest vortex. Within this range, an energy flux cascades from large to small scales. The kinetic energy dissipation rate ϵ corresponds to this energy flux passed across scales which is then evacuated at small scales by molecular viscosity.

Analogously, for pure 2D turbulence with energy injection at scale ℓ_0, Kraichnan (1967) demonstrated that two different cascades are coexisting: a direct cascade of enstrophy (the L^2 norm of the vorticity $\boldsymbol{\omega} = \nabla\times\mathbf{v}$) which satisfies $S_3(\ell) = \epsilon_\omega\ell^3/8$ within the inertial range $[\eta,\ell_0]$, and an inverse cascade of energy where $S_3(\ell) = 3\epsilon\ell/2$ within the range $[\ell_0, L]$. An enstrophy flux ϵ_ω passes in the direct cascade from large to small scales, whereas an energy flux ϵ passes in the inverse cascade from small to large scales.

In the case of atmospheric turbulence, the actual scaling laws constitute still open questions. Lindborg (1999) and Lindborg and Cho (2001) consider scaling laws inferred from aircraft measurements and propose an answer to the question: "can atmospheric flow statistics be explained by two-dimensional turbulence?". The authors showed that, under some isotropy assumption, the self-similar processes observed in the aircraft data (Nastrom et al., 1984) at small and at large scales can be modeled by the superposition of a

[2] The hypothesis of isotropy can be relaxed, see e.g. Taylor et al. (2003). The method to recover isotropic statistics described there motivates the use of different directions for the displacements in Section 3.1

3D direct energy cascade and a 2D direct enstrophy cascade so that

$$S_3(\ell) = -\epsilon\ell + \frac{1}{8}\epsilon_\omega \ell^3. \tag{8}$$

Aside from the problems to determine direction of energy and enstrophy cascades in atmospheric turbulence, it should be mentioned that there are still several independent exact relations that hold for the third-order structure function and which predict power law behavior for several flows of interest in atmospheric sciences (Monin and Yaglom, 1971; Kurien et al., 2000).

Going further, Kolmogorov assumed that the longitudinal velocity increments were strictly self-similar to explain the scaling laws observed in many experiments. If this is true, the normalized PDF of motion increments is self-similar through scales. This implies that, in a given cascade, the p-th order structure function follows a power law:

$$\gamma_p \ell^{\zeta_p} \sim S_p(\ell) = [S_3(\ell)]^{p/3} \sim \gamma_p \ell^{p\zeta_3/3}, \tag{9}$$

with universal exponents ζ_p depending on space dimension, and where the prefactors γ_p are a function of energy flux ϵ (or enstrophy flux ϵ_ω in two dimensions). A corollary of the strictly self-similar assumption is that in inertial range any 3D turbulent flow has a uniform Lipschitz regularity of $\zeta_1 = 1/3$, while a 2D turbulent flow is characterized by $\zeta_1 = 1$ (*i.e.*, it is regular). However, it is now well known that for 3D turbulence and for the direct cascade range of 2D turbulence, the Kolmogorov assumption deviates from reality because of intermittency (tendency of the flow to develop strong gradients at small scales). As a consequence, only non-strict self-similarity can in reality be assumed. It results that deviations on exponent values can be expected for $p \neq 3$, although power law behavior in inertial range still holds. The special case of the inverse cascade of 2D turbulence must be remarked, as growing evidence indicates that it is self-similar (Benzi and Scardovelli, 1995; Bernard et al., 2006).

Finally, another result that is used in the following sections is that any 2D or 3D turbulent flow is regular in the dissipative range; using Taylor expansion we then have $S_2(\ell) \sim \ell^2$ within $\ell \in [0, \eta]$. All these general power law dependencies motivate the use of self-similar *priors* for motion estimation as discussed in the next section.

3 Regularization of motion increments

Exhibiting self-similarity properties of fluid flows enables to define prior spatial regularization functionals. Besides providing a closure for motion estimation, power-law *priors* have several advantages compared to traditional smoothing functions: they constitute physical sound regularizers for fluid motion and they provide intrinsic multi-scale *prior* models which structures motion across scales. In this section we discuss the use of the self-similar regularization of motion increments with a *prior* power law, and postpone to section 4 how Bayesian modeling is used to select the most appropriate power law from the images (instead of imposing one).

3.1 Self-similar constraints

Let us first formalize the use of self-similar constraints to specify regularization functionals. Although there is no exact prediction on its scaling law for non-strictly self-similar flows, we chose to use the second order structure function $S_2(\ell)$ because it constitutes a convenient quadratic constraint for our purpose. Also, this quantity is related with the energy spectrum and the two-point correlation function, quantities of interest in experiments and simulations. Since the exponent of the power law is unknown, we take into account in section 4 deviations from the predicted laws by selecting the most likely power law defined by the two parameters γ_2 and ζ_2 given the image data (γ and ζ for simplicity in the following).

$S_2(\ell)$ is an expectation which can be obtained by spatial integration over the image domain and over all directions as presented in Eq. (7). A self-similar constraint is then defined at each scale ℓ as the difference between the 2-nd order structure function and (for the moment) a given power law depending on some parameters γ and ζ:

$$\mathcal{M}(\ell, \gamma, \zeta) = \gamma \ell^{\zeta}. \tag{10}$$

An estimated motion field should thus respect the self-similar constraint:

$$g_\ell(\mathbf{v}, \mathcal{M}) = \frac{1}{2}\left[S_2(\ell) - \mathcal{M}(\ell, \gamma, \zeta)\right] = 0. \tag{11}$$

This constraint replaces the usual smoothness constraints used in the minimization problem discussed in Sect. 2.1.

3.2 Constrained motion estimation problem

Referring to section 2.1, the minimization of the ill-conditionned optic-flow estimation problem can be written as

$$(\hat{\mathbf{v}}) = \arg\min_{\mathbf{v}} f_d(I, \mathbf{v}). \tag{12}$$

Adding the self-similar constraints, we obtain the closed constraint optic-flow minimization problem:

$$\begin{cases} \min_{\mathbf{v}} \ f_d(I, \mathbf{v}), \ \ \mathbf{v}(\mathbf{x}) \in \mathbb{R} \\ \text{subject to the constraints:} \\ g_\ell(\mathbf{v}, \mathcal{M}) = 0, \ \ \forall \ell \in \mathbf{I} \end{cases} \tag{13}$$

where $\mathbf{I}$ is the scale range of the given power law. The system (13) is discretized on a discrete set of scales, and the velocity field is recovered from the image intensity solving the constrained minimization using the Lagrangian formalism. It yields the optimal solution:

$$\mathbf{v}^* = \arg\min_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda}^*) \tag{14}$$

where the Lagrangian function is given by $L(\mathbf{v}, \boldsymbol{\lambda}^*) = f_d(I, \mathbf{v}) + \sum_\ell \lambda_\ell^* g_\ell(\mathbf{v}, \mathcal{M})$ and where $\boldsymbol{\lambda}^* = \{\lambda_\ell^*\}$ represent the Lagrangian multipliers. Further details are given in A.

4 Selection of a prior power law

So far, the proposed model recovers motion from images conditioned by some *prior* power law model $\mathcal{M}(\gamma, \zeta)$ for the second order structure function. This power law can be given

by turbulence statistical models describing the self-similar structure of turbulence, which is assumed universal in the works of Kolmogorov (1941), Kraichnan (1967), Lindborg and Cho (2001). However, we now want to relax this universal assumption by allowing ζ and γ to change, and select the most appropriate power law model for motion estimation given only the image data. With this *prior* model selection we cope with uncertainties in turbulence theoretical predictions associated to, *e.g.*, intermittency effects. Moreover, it will enable us to reveal directly from the images important physical quantities in turbulence such as power law exponents, flux across scales, or energy and enstrophy dissipation rates. To this end, we now reformulate the constrained motion estimation problem in a probabilistic framework, and we show how probability of the *prior* model $\mathcal{M}$ can be evaluated given image data. Bayes' rule provides a nice framework to evaluate this model probability.

4.1 Probabilistic model

A probabilistic reformulation of the global motion estimation problem yields a model relating the image intensity function variable I, the motion field variable $\mathbf{v}$, an observation model variance β^{-1}, and finally a *prior* power law model $\mathcal{M}(\zeta,\gamma)$. Noticing conditional independances, the joint probability can be factorized in three primary statistical models yielding the Bayesian hierarchical model:

$$p(I,\mathbf{v},\beta,\mathcal{M}) = \underbrace{p(I|\mathbf{v},\beta)}_{\text{likelihood}}\ \underbrace{p(\mathbf{v}|\mathcal{M})}_{\text{motion prior}}\ \underbrace{p(\beta)p(\mathcal{M})}_{\text{model prior}} \tag{15}$$

where $p(I|\mathbf{v},\beta)$ represents the image observation PDF or likelihood, $p(\mathbf{v}|\mathcal{M})$ denotes the motion prior PDF, and where $p(\beta)$ and $p(\mathcal{M})$ denote the prior PDF on respectively the likelihood hyper-parameter and the power-law model. Let us note that Bayesian hierarchical model can also be used to decompose multidimensional probability distributions of complex spatio-temporal processes in geophysics (Berliner et al., 1999; Wikle et al., 2001; Hoar et al., 2003). Using hierarchical models, it is then possible to infer the joint *Maximum A Posteriori* (MAP) estimate by maximization of the joint probability $p(I,\mathbf{v},\beta,\mathcal{M})$ with respect to $\mathbf{v}$, β and $\mathcal{M}$. However, it is well-known in Bayesian analysis that selection with a joint MAP estimator tends to over-fit the data, *i.e.*, to favor models with large degrees of freedom (Robert, 2007). Bayesian analysis suggests to rely instead on marginalized posterior distributions for model and hyper-parameter selection. Therefore, by successive marginalization of the joint posterior distribution $p(I,\mathbf{v},\beta,\mathcal{M})$, we perform three levels of inference yielding estimates of: 1) the velocity field, 2) the variance of the observation model error, and 3) the power law model. Although, variable β does not affect the first level of inference, we will see that it is necessary to consider the variance of the observation model to perform model selection. In the following for notational convenience, the dependence of the Lagrangian multipliers to model $\mathcal{M}$ will be omitted. The three levels are detailed in the next sections.

4.2 First inference level: motion field estimation

In this first level of inference, we consider the maximization of the joint probability with respect to $\mathbf{v}$ given some data realization I_0, the hyper-parameter estimate $\hat{\beta}$ and the selected scaling model $\hat{\mathcal{M}}$:

$$\mathbf{v}^* = \arg\max_{\mathbf{v}} p(\mathbf{v},I_0,\hat{\beta},\hat{\mathcal{M}}) = \arg\max_{\mathbf{v}} p(\mathbf{v}|I_0,\hat{\beta},\hat{\mathcal{M}})$$

We obtain the MAP estimate $\mathbf{v}^*$. Let us show that solving the problem (P) introduced in the previous section is equivalent to infering a velocity field $\mathbf{v}^*$ according to a MAP criterion. This can be shown using Bayes' relation in order to define the motion field *posterior* PDF:

$$p(\mathbf{v}|I,\beta,\mathcal{M}) = \frac{p(I|\mathbf{v},\beta)p(\mathbf{v}|\mathcal{M})}{p(I|\beta,\mathcal{M})} = \frac{\text{likelihood}\times\text{prior}}{\text{evidence}}$$
$$\propto p(I|\mathbf{v},\beta)p(\mathbf{v}|\mathcal{M}) \tag{16}$$

where the *likelihood*, the *prior* and the *posterior* are Gibbs PDF. More precisely, the quadratic observation model f_d represents the energy of an uncorrelated Gaussian *likelihood* PDF with constant variance β^{-1}:

$$p(I|\mathbf{v},\beta) = \frac{\exp\{-\beta f_d(I,\mathbf{v})\}}{Z_{f_d}(\beta)}, \tag{17}$$

where $Z_{f_d}(\beta)$ denotes a normalization constant also called partition function. The constraints $\{g_\ell\}$ weighted by inverse variances $\{\beta\lambda_\ell^*\}$ form the energy of a multiscale *prior* PDF:

$$p(\mathbf{v}|\mathcal{M}) = \frac{\exp\{-\sum_\ell \beta\lambda_\ell^* g_\ell(\mathbf{v},\mathcal{M})\}}{Z_{g_\ell}(\mathcal{M})}, \tag{18}$$

where $Z_{g_\ell}(\mathcal{M})$ denotes the partition function. At the point $\boldsymbol{\lambda}^*$, the Lagrangian function $L(\mathbf{v},\boldsymbol{\lambda}^*) = f_d(I,\mathbf{v}) + \sum_\ell \lambda_\ell^* g_\ell(\mathbf{v},\mathcal{M})$ thus defines, up to a multiplicative factor, the energy of the *posterior* PDF:

$$p(\mathbf{v}|I,\beta,\mathcal{M}) = \frac{\exp\{-\beta L(\mathbf{v},\boldsymbol{\lambda}^*)\}}{Z_L(\beta,\mathcal{M})}, \tag{19}$$

where $Z_L(\beta,\mathcal{M})$ denotes the associated partition function. It is now obvious that the maximum $\mathbf{v}^*$ of the *posterior* is reached at the minimum of the Lagrangian:

$$\mathbf{v}^* = \arg\max_{\mathbf{v}} p(\mathbf{v}|I,\beta,\mathcal{M}) = \arg\min_{\mathbf{v}} L(\mathbf{v},\boldsymbol{\lambda}^*), \tag{20}$$

which constitutes also the solution (14) of the constrained motion estimation problem presented in section 3.2.

4.3 Second inference level: estimation of observation model variance

A second level of inference considers the maximization with respect to β of the joint posterior PDF, marginalized over $\mathbf{v}$, given the data realization I_0 and the selected scaling model $\hat{\mathcal{M}}$:

$$\hat{\beta} = \arg\max_{\beta} \int_{\mathbb{R}^n} p(\mathbf{v},I_0,\beta,\hat{\mathcal{M}})d\mathbf{v} \ = \arg\max_{\beta} p(\beta|I_0,\hat{\mathcal{M}})$$

yielding the hyper-parameter estimate $\hat{\beta}$. Using again Bayes' relation one can write:

$$p(\beta|I,\mathcal{M}) = \frac{p(I|\beta,\mathcal{M})p(\beta)}{p(I|\mathcal{M})} \propto p(I|\beta,\mathcal{M})p(\beta). \tag{21}$$

Assuming a flat *prior* on variables β, the MAP $\hat{\beta}$ of (21) is simply the *Maximum Likelihood* (ML) estimate, or in other words the maximum of the *evidence* PDF $p(I|\beta,\mathcal{M})$. This probability is defined by a marginalization w.r.t. the velocity field:

$$p(I|\beta,\mathcal{M}) = \int_{\mathbb{R}^n} p(I|\mathbf{v},\beta,\mathcal{M})p(\mathbf{v}|\beta,\mathcal{M})\mathrm{d}\mathbf{v}. \qquad (22)$$

Let us recall that the *evidence* is also the normalization constant (w.r.t. velocity field $\mathbf{v}$) which has been ignored in the first level of inference (Eq. 16). Therefore, we can rewrite the *evidence* as a partition function ratio:

$$p(I|\beta,\mathcal{M}) = \frac{\text{likelihood}\times\text{prior}}{\text{posterior}} = \frac{Z_L(\beta,\mathcal{M})}{Z_{f_d}(\beta)Z_{g_\ell}(\mathcal{M})}, \qquad (23)$$

Assuming the energy of the *evidence* is convex (at least locally), the variance estimate $\hat{\beta}^{-1}$ maximizing the *evidence* is obtained by canceling out its derivative.

$$\hat{\beta} = \arg\max_{\beta} p(I|\beta,\mathcal{M}) \Leftrightarrow \frac{\partial p(I|\hat{\beta},\mathcal{M})}{\partial\beta} = 0 \qquad (24)$$

The analytical expression of the *evidence* probability, of its derivate and of the variance estimate is detailed in appendix B.

4.4 Third inference level: selection of power law

A third level of inference considers the maximization with respect to $\mathcal{M}$ of the joint posterior PDF, successively marginalized over $\mathbf{v}$ and β, given some data realisation I_0:

$$\hat{\mathcal{M}} = \arg\max_{\mathcal{M}} \int_{\mathbb{R}}\int_{\mathbb{R}^n} p(\mathbf{v},I_0,\beta,\mathcal{M})d\mathbf{v}d\beta \;= \arg\max_{\mathcal{M}} p(\mathcal{M}|I_0).$$

Again, Bayes' relation is used to select the model:

$$p(\mathcal{M}|I) = \frac{p(I|\mathcal{M})p(\mathcal{M})}{p(I)} \propto p(I|\mathcal{M})p(\mathcal{M}). \qquad (25)$$

The data-dependent term, the *likelihood* $p(I|\mathcal{M})$ for $\mathcal{M}$ (also called *evidence* for $\mathcal{M}$) appeared earlier as the normalization constant of the second level of inference (Eq. 21). Its evaluation implies marginalizing out variable β:

$$p(I|\mathcal{M}) = \int_{\mathbb{R}} p(I|\beta,\mathcal{M})p(\beta)\mathrm{d}\beta. \qquad (26)$$

This integral (26) can be approached by Laplace's approximation, and depends on the estimate $\hat{\beta}$ as detailed in appendix C. The model maximizing the *posterior* probability (26) of the power law model:

$$\hat{\mathcal{M}} = \arg\max_{\mathcal{M}} p(\mathcal{M}|I) = \arg\max_{\mathcal{M}} p(I|\mathcal{M})p(\mathcal{M}) \qquad (27)$$

is obtained by uniform sampling in $\mathbb{R}^2$ of the power law parameters (γ,ζ).

As a result, it is important to outline that, a priori, no particular power law is imposed on the data, but rather the second order structure function is constrained to follow some power law in some range of scales (which can be selected). If the smallest resolved scales in the images correspond to the dissipative scales of the flow, the constraint reduces to a multiscale first-order smoothing, as in that range $S_2(\ell)\sim\ell^2$. This kind of smoothing function constitute a multiscale generalization of regularizer (5) computed with finite differences. On the other hand, if the smallest resolved scales in images are in the inertial range of the turbulent flow, the constraint gives a better representation of the physics of the problem than *adhoc* regularizers. Finally, note that since laminar flows possess vanishing increments, their structure functions can be modeled with power laws with prefactors equal zero ($\gamma = 0$). Appendix D shows with a toy example how the proposed methodology succeeds to recover this particular case. Therefore, the proposed methodology can be expected to improve the physical representation of the flow in turbulent cases, and enforces smooth solutions if the hidden flow is smooth or laminar.

5 Numerical and experimental validation

5.1 Synthetic turbulence of horizontal velocity fields

To evaluate the performance of the reconstruction of the horizontal velocity field from intensity images using the proposed self-similar regularization, synthetic image sequences were generated based on forced 2D turbulence obtained from direct numerical simulations (DNS) of the 2D and incompressible Navier-Stokes equations with a Reynolds number of 3000. The pseudo-spectral code used for this study solved the vorticity conservation equation and the Lagrangian equation for particle tracers transported by the flow or, in the case of scalar images, the advection-diffusion equation *i.e.*, model (2) in the case of divergence-free velocities, with a Pecklet number of 2100 . The flow was stirred by an external forcing acting between wavenumbers $k = 5$ and 10. Figure 1 shows the isotropic (angle averaged) energy spectrum and the energy flux for one snapshot of the velocity field. A flat spectrum is observed at scales larger than the forcing scale, resulting from an incipient inverse cascade of energy. The direct cascade range shows a spectrum steeper than $\sim k^{-3}$ expected from Kraichnan phenomenology, which is not unusual in numerical simulations of 2D turbulence, and which in this particular case can also be associated to the moderate spatial resolution and Reynolds number considered (the slope is in fact close to -5). The simulation shows positive enstrophy flux towards small scales for wavenumbers larger than the forcing wavenumbers, and a few wavenumbers with negative energy flux associated to the beginning of the inverse cascade. Also, the energy flux in the direct cascade range is negligible when compared with the enstrophy flux (see Boffetta (2007) for examples of energy and enstrophy fluxes in high resolution simulations of 2D turbulence). A precise description of the simulation can be found in Carlier and Wieneke (2005). This database has proven its relevance in previous studies in the field of fluid flow estimation Yuan et al. (2007); Heas et al. (2007b); Papadakis and Memin (2008); Heas et al. (2009).

Figure 2 presents a sample of the synthetic particle and scalar images of 256×256 pixels generated using the numerical simulation. The corresponding DNS velocity field is also presented in Fig.2. The motion estimates were obtained by minimizing discrepancies to the brightness consistency (1) (resp. to the scalar diffusion equation, *i.e.*, model (2) in the case of divergence-free velocities) for particle images (resp. for scalar images) under power-law constraints. The self-similar regularizer was applied only in the range of $\mathbf{I} = [1, 8]$ pixels (corresponding either to dissipation or to the end of the inertial range). A multi-resolution approach was used to cope with large displacements (Bergen et al., 1992). Motion estimates can be compared with the DNS motion fields, and

with the operational correlation-based software from *LaVision* company in Fig. 2. The true and reconstructed velocity fields look similar. It can also be noticed that although correlation-based methods are accurate measurement tools for particle imagery, they usually perform badly for scalar imagery. Quantitative details of the estimation are given below.

5.1.1 Evaluation of power law selection A flat *prior* was chosen on the power law parameters in (27) and therefore, inference was performed by maximizing the *evidence* probability (26) w.r.t. γ and ζ. This probability was evaluated by sampling the exponent ζ near the theoretical value of 2 (the value expected for a dissipative range or for a direct enstrophy cascade from the theoretical predictions), and by sampling the prefactor γ around a first guest given for instance by a Least Square (LS) fitting of the solution of a rough Horn and Schunck (1981) estimation.

Fig. 3 shows the shape of the evidence according to the power-law model parameters. The model probability global maximum is located at the coordinates $\zeta = 0.0030$, $\gamma = 1.975$ (resp. $\zeta = 0.00275$, $\gamma = 1.850$) for the particle (resp. scalar) images. Three observations assess experimentally the Bayesian criterion used for model inference. First, Fig. 3 shows that, specially for particle images, the parameters yielding high model probabilities correspond to low Root Mean Square (RMS) motion reconstruction errors. Second, this correspondence is not so remarkable in the case of scalar images, because minimizing an RMS error criterion does not always lead to the underlying power law. Indeed, as shown in the plots of Fig. 4, the inferred power-law parameters are for both experiments very close to the values $\zeta = 0.003$, $\gamma = 1.948$ obtained by a least square fit on the ground truth motion field, whereas on the contrary in the case of scalar images, choosing the power-law minimizing the RMS error induces a slight bias. Remarkably, the reconstructed structure functions $S_2(\ell)$ in x- and y-directions (horizontal and vertical in the images) as well as in the diagonal directions adjust well the structure functions computed directly from the simulation (indicated by the straight lines) in a wide range of scales, and well beyond the range used for the regularizer. Third, between 1 and 8 pixels, the power law of all structure functions is close to $\sim \ell^2$, which is the expected value from the theory in the inertial or dissipative range. It is also observed that the simulation approaches isotropic statistics at small scales since structure functions calculated in horizontal-vertical and in diagonal directions are nearly identical.

Therefore, maximizing the model *evidence* constitutes a theoretically sound and reliable criterion for selecting the *prior* power law.

5.1.2 Evaluation of motion estimation A comparison with state-of-the-art methods based on the RMS reconstruction error[3] is provided in figure 5. Based on this criterion, the proposed method outperforms in average most accurate operational correlation-based techniques, first order (Horn and Schunck, 1981), and div-curl (Corpetti et al., 2002; Yuan et al., 2007) regularizers.

To assess quantitatively the accuracy of the reconstruction of the velocity increment PDFs, we plotted in Fig. 6 the exponents and the prefactors of structure functions w.r.t. their order. The exponents correspond to the dissipative range as they were computed by LS fitting of the reconstructed motion fields for increments of the structure functions between 1 and 8 pixels. Analyzing the DNS motion field, it can be noticed that, in this range, the field is smooth and the exponents are on a straight line, as the structure function of order p, $S_p(\ell)$, goes as $\sim \ell^p$. The other estimates introduce spurious deviations from the straight line that may be incorrectly interpreted as intermittency. It is striking that the proposed method provides a very good estimation of the exponents and prefactors while all other methods fail to characterize exponents above the 5^{th} order. This is a testimony of the power of the proposed motion estimation methods in recovering accurately the multiscale structure of turbulence.

5.2 Atmospheric turbulence

We now assess the inverse motion modeling approach using power laws on real data. The objective of this section is to estimate atmospheric parameters from satellite images, even when limitations associated to images and to required approximations may only give orders of magnitude of quantities of interest. We also compare results with previous observations using *in situ* techniques. The analysis is done using METEOSAT Second Generation meteorological image sequences acquired above the north Atlantic Ocean at a rate of an image every 15 min during part of one day (5-June-2004), from 12h00 to 14h30 UTC (universal time, similar to GMT, the Greenwich Mean Time). Image spatial resolution is 3×3 km^2 at the center of the whole Earth image disk. This benchmark data, which has been provided by the EUMETSAT consortium, is composed of images of top of cloud pressure and cloud-classification images. Cloud-classifications were used to segment images into two broad layers, at low and intermediate altitude[4]. Following the methodology described in Heas et al. (2007a) and in Corpetti et al. (2009), a set of sparse pressure difference images of 256×256 pixels associated to a stack of layers (low and intermediate altitude) was derived. Resulting set of input images is displayed in figure 7. As discussed before, an observation model is required for the inversion problem (see Sect. 2.1). An identical direct observation model as in (Heas et al., 2007a) was used. It is based on layer mass conservation to relate image intensity function to vertically

[3] In order to compute RMS error on the pixel grid, true velocity fields have been interpolated since they are given on a shifted (by half a pixel) grid by direct numerical simulation .

[4] Stratus and stratocumulus (resp. altostratus, atltocumulus and nimbostratus) are classified in the lower layer (resp. the intermediate layer).

averaged horizontal wind fields[5].

In the following, based on the work of Lindborg and Cho (2001), we first introduce *prior* knowledge on the sought power law using the Bayesian relation (27). The power law is chosen to model the longitudinal structure function of indifferently the zonal and meridional horizontal winds. Power law selection reduces then in this case to estimating the energy flux of the direct energy cascade towards small scales. We then relax this assumption, and perform model selection considering a flat *prior* distribution on both unknowns: the power law prefactor and the power law exponent. Finally, we provide some statistical analysis and compare wind fields obtained with the two methods.

5.2.1 Power law selection with a Lindborg's prior In this section, we assume that the scaling $\zeta = 2/3$ predicted by *Lindborg* in the direct energy cascade holds in the range $\mathbf{I} = [1, 4]$ pixels, equivalent to $\mathbf{I} = [3, 12]$ km (the direct energy cascade is visible in the 2-nd order structure function for separations of about 10 kilometers in the data of Lindborg (1999)). This hypothesis can be formulated in the Bayesian relation of (27) by choosing a *prior* such as $p(\mathcal{M}) = \delta(\zeta - 2/3)\mathcal{U}(\gamma)$, where δ and $\mathcal{U}$ denote respectively the Dirac and the flat distributions. We thus only need to infer the parameter γ. Note that no power law is imposed for separations larger than 12 km.

5.2.1.1 Direct energy flux estimation: The model proposed in Lindborg (1999) provides power law models for the 2-nd order structure function as:

$$S_2(\ell) = C_2 \epsilon^{\frac{2}{3}} \ell^{\frac{2}{3}} + b\,\ell^2 - c\,\ell^2 \log \ell, \tag{28}$$

where b and c denote parameters and $C_2 \approx 6$ is a Kolmogorov constant. The energy flux ϵ (which is equal to the energy dissipation rate) can be related in the scale range $\mathbf{I}$ to the power law prefactor by $\gamma = C_2 \epsilon^{\frac{2}{3}}$. Therefore, maximizing the power law *posterior* probability also reveals the most likely energy flux. Figure 8 shows that this probability is maximized for power laws $\hat{\mathcal{M}} = C_2 \hat{\epsilon}^{\frac{2}{3}} \ell^{\frac{2}{3}}$ with energy flux:

$$\begin{cases} \hat{\epsilon}^{low} \simeq (1.00 \pm 0.87) \times 10^{-5} m^2 s^{-3} \\ \hat{\epsilon}^{mid} \simeq (1.00 \pm 0.87) \times 10^{-6} m^2 s^{-3} \end{cases}, \tag{29}$$

which correspond to estimates at low and intermediate altitudes. These estimates have the same order of magnitude as previously reported results based on aircraft data analysis[6]. The agreement is very good considering the measure is only based on image data. It should be noted that in the proposed estimation approach, as the *evidence* maximization does not depend on motion variables, energy flux is obtained directly from the image intensity function, conversely to other approaches, which need to first extract motion observations from the images and then estimate power-law parameters. One of the main results of the studies of Lindborg (1999) and Lindborg and Cho (2001) was to determine the direction of the energy cascade from the third order structure function from *in situ* data. The third order structure function can be reconstructed from the image-based motion measurements. Figure 9 shows this function for the lower and the intermediate layer. A direct cascade with linear dependence $\sim -\hat{\epsilon}^{low} \ell$ can be identified in the scale range $[10, 40]$ km, and the agreement of the structure function for larger scales with the theoretical prediction seems reasonable.

5.2.1.2 Direct enstrophy flux estimation. A least square estimate of the enstrophy flux $\hat{\epsilon}_\omega$ can be obtained by fitting w.r.t. $\hat{\epsilon}_\omega$ the third order structure function to its model given by Eq. (8) in the converged scale range ($\ell > 50$ km). In particular, at the upper range of the cascade, positive cubic power laws can be observed for both layers. The least square average enstrophy flux estimates are

$$\begin{cases} \hat{\epsilon}_\omega^{low} \simeq (6.75 \pm 0.13) \times 10^{-15} s^{-3} \\ \hat{\epsilon}_\omega^{mid} \simeq (11.9 \pm 0.18) \times 10^{-15} s^{-3}. \end{cases} \tag{30}$$

These results are also consistent with previously published results[7]. Third order structure function models obtained by LS estimation are displayed in Fig. 9. Second order structure function and energy spectrum models given by Eq. (28) can also be adjusted in a LS sense, and the parameters b and c can be estimated for scales $\ell > 50$ km. Figure 9 thus also shows the 2-nd order structure functions and the energy spectra together with their corresponding models.

5.2.1.3 Moments convergence: In order to verify that we obtained converged statistics for the evaluation of the structure functions, we follow Gotoh et al. (2002) and examine the convergence towards a flat curve of the accumulated moments

$$C_p(z, \ell) = \int_0^z \left|\delta v_\parallel(\ell)'\right|^p p'\left[\delta v_\parallel(\ell)'\right]\, d\delta v_\parallel(\ell)', \tag{31}$$

where $p'[\delta v_\parallel(\ell)'] = \sigma_\ell p_\ell[\delta v_\parallel(\ell)/\sigma_\ell]$ is the PDF of the normalized increments $\delta v_\parallel(\ell)' = \delta v_\parallel(\ell)/\sigma_\ell$ with σ_ℓ denoting the standard deviation. To check convergence of the second and third order structure functions, we examine the accumulated moments and see if there is convergence towards flat curves for increments associated to the tails of the PDF. Curves $C_2(z, \ell)$ and $C_3(z, \ell)$ show that second and third order moments are reasonably converged. However, we observe

[5] Let us remark that one can not guarantee that the data model which has been proposed in Heas et al. (2007a) is valid for regions where the model assumptions are violated (e.g., for non-layered structures) or for too noisy input observations (e.g., due to cloud classification errors or pressure retrieval errors). However, the local weakness of the observation model is balanced in our approach by the introduction of the relevant *prior* information on turbulent motion regularity.

[6] A collection of other air-craft measurements (Nastrom et al., 1984) shows a typical value in the troposphere close to $\sim 3 \times 10^{-6} m^2 s^{-3}$ (Lindborg, 1999). Nevertheless, note that analogously to stratospheric observations (Dewan, 1997; Nastrom and Eaton, 1997; Cho and Lindborg, 2001), one should consider a significant variability in measurements

[7] These values agree well with an early estimate of $\sim 10^{-15} s^{-3}$ obtained in Charney (1971). Authors in Tung (2003) then reobtained this same result from simulations. Note however that enstrophy flux is slightly lower than the estimate $\sim 10^{-13} s^{-3}$ of Lindborg (1999) using Nastrom et al. (1984) air-craft data.

that accumulated moments show a faster convergence for the lower layer motion. This obviously results from sparsity of image data which increases with altitude. Furthermore, curves show that the third order moments are better converged for large scale increments ($\ell \gtrsim 50$ km) than for small scale increments ($\ell \lesssim 50$ km). This is in agreement with the fact that the normalized PDFs of velocity increments show stronger tails as the scale studied is decreased. These tails, which depart from Gaussian behavior, are the result of intermittency: turbulence comes in gusts, and scarce regions with strong gradients develop in the flow as the result of nonlinearities. The development of these strong events also make the convergence of high order moments slower, as more data is needed to have enough statistics to resolve the impulsive (but scarce) events. An extension of the present study to consider more images may enhance convergence.

5.2.2 Power law selection with a flat prior In this section, we now relax the previous scaling assumption and consider instead a uniform *prior* $p(\mathcal{M})$ defined on the bidimensional parameter space $(\gamma, \zeta) \in \mathbb{R}^2$ of the power law. Thus, we now also infer the most likely exponent $\hat{\zeta}$ of the power law together with the prefactor $\hat{\gamma}$. Figure 10 shows that the selected prefactor (*i.e.*, the energy flux) has the same order of magnitude as in the previous case. However, the selected power law exponent does not correspond to the model in (Lindborg and Cho, 2001). Indeed, the 2^{nd} order structure function maximizing the *evidence* probability does not scale as $\sim \ell^{\frac{2}{3}}$ but rather as ℓ^2 (at least, in the range of scales in which the constraint is applied; as will be shown next the structure functions scale in a similar way at scales larger independently of whether a scaling assumption or a uniform *prior* are used). Moreover, note that the model in (Lindborg and Cho, 2001) still corresponds to a region of high probability, *i.e.*, a valley of the probability energy (marked by an iso-contour in Fig.10). In the absence of ground truth, it is not obvious to decide whether one power law model is better than the other. One may conclude that using the images alone favors (in the particular case of these type of images) the choice of a low-complexity model which does not over-fit the data. In other words, the scaling of a regular motion is preferred, although more complex scaling are not drawn aside. When *prior* knowledge is available, the Bayesian modeling framework is able to enforce relevant regions of lower probability and recover accurately complex turbulence scaling laws.

5.2.3 Comparison of wind field estimates Deviation from strict self-similarity of the two motion estimates can be evaluated based on the behavior w.r.t. their order of the structure function exponents (obtained by LS fitting of the reconstructed motion fields in the scale range of 1 to 6 pixels, corresponding to 3 to 18 km). Analyzing the results in Fig. 11, it can be seen that we obtain a stronger deviation from linearity with the model of Lindborg and Cho (2001). However, in both cases the motion field is multifractal and not strictly self-similar.

Exponents in Fig. 11 were computed for a similar range of scales as the one used for power law selection (3 to 12 km). It is thus not surprising that the second order exponent changes substantially (as well as the other exponents), as in that range the method with the flat prior selected the scaling $\sim \ell^2$. This change in the exponents should be understood in the same light as the change in the reconstructed power law of the second order structure function: images are favoring a model of lower complexity with regular motion at small scales.

However, structure functions show similar slopes and features for both reconstructions outside the range used for power law selection (e.g., for scales larger than 40 to 80 km depending on the altitude). It is not the motivation of our work to do a detailed analysis of scaling exponents at intermediate scales for atmospheric data, but just as an example, in Fig. 11 and 12 we show the third-order structure functions for low and mid altitude using both methods, and the scaling exponents at larger scales (10 to 30 pixels, or 30 to 90 km) for the model of Lindborg and Cho (2001) and for a flat prior. Two features are worth mentioning. In Fig. 12, note that not only amplitude of structure functions is similar in both cases (as already mentioned, as the amplitude of these functions is associated to the energy flux), but also that structure functions change sign at similar scales. In Fig. 11, exponents are similar at intermediate scales independently of the method used to reconstruct the field.

Figure 13 compares motion fields obtained with our method (prior power law in ℓ^2) with a vector field obtained by a correlation-based method. It is conspicuous that, for these very sparse and noisy images, correlation yields spurious and incoherent motion estimates while our approach yields a coherent and physically plausible reconstruction. Scaling in ℓ^2 tends to produce vortex and divergent structures of larger magnitude than for scaling in $\ell^{2/3}$. Motion fields may present what could be interpreted as local abrupt discontinuities when visualized at larger scales. Indeed, the proposed optic-flow method provides a velocity vector per pixel. Therefore, small vortex and divergent structures of only a few pixels size may be confused with noise. In order to remove the confusion, we plot results at the pixel level. As shown in Fig. 14 and Fig. 15, velocity fields at this resolution appear to be very smooth and structured. In particular, vortex, sinks and sources are coherent structures of comparable order as claimed in (Lindborg, 2007).

6 Conclusions and perspectives

In this study, we presented a physically-based, multiscale method that does not involve any tuning of regularization parameters for fluid motion modeling from images. It relies on a Bayesian hierarchical model which simultaneously provides optimal solutions for two problems: motion estimation and regularization model selection. The regularization models arise from Kolmogorov's work on turbulent flow self-similarity and recent results in the study of turbulent flows. Structure functions are used to exhibit some regularity in velocity field reconstruction, without *a priori* imposed power law. Besides achieving motion inversion, the method recovers several quantities of physical interest in turbulence. Moreover, for images of turbulent flows (or for laminar flows) where dissipative range is resolved, small scales are smooth and the proposed method reduces in fact to a smoothing with, e.g., null second order structure function,

or with the second order structure function proportional to the square of the spatial increment. Experiments on a synthetic sequence shows that the method is more accurate than the best fluid-dedicated motion estimators. As a result, we believe the method constitutes a valuable tool for physical characterization of turbulence from images. In particular, we were able to recover consistent flux across scales in atmospheric turbulence at different altitudes from a meteorological image sequence. Results are also consistent with theoretical predictions and with previous studies using *in situ* measurements.

We believe this work opens interesting perspectives for observational characterization of turbulence. Characterization of turbulence cascades could also be performed by analyzing other meteorological (or laboratory) images, depicting for instance evolution of passive scalars, water vapor, or temperature fields. Of course, the method requires a proper direct image observation model for each problem of interest, and as a result may have to be complemented with other measurement techniques for validation. However, we believe that the results presented here are promising and give a useful tool to complement other measurements in the difficult problem of turbulence characterization. Inferring directly the scale range on which the prior power-law should be exhibited according to the observed images represent an interesting perspective which is in principle straightforward considering a Bayesian framework. It would present the advantage to extend the set of possible power-law multi-scale models considered for selection to traditional single scale regularizers, which represent power-law particular cases. Another interesting perspective could concern the addition of a *prior* dynamical model to ensure dynamical consistency along time. Several works relying on variational image assimilation techniques have recently been proposed in this direction (Papadakis and Memin, 2008; Corpetti et al., 2009).

APPENDIX A: Solution of the constrained motion estimation problem

In this appendix, we show how to solve optimally the constrained problem given by Eq. (13).

A1 Problem discrete formulation

Let us first express this problem in its discrete form. The derivatives $\nabla_{\mathbf{v}} f_d(I,\mathbf{v})$ related to any quadratic data term w.r.t motion $\mathbf{v}$ (*e.g.*, the model of (4)) can be expressed in the matricial form $A_0\mathbf{v}-b_0$, when discretized on an image grid S of m points. The two discretized components of $\mathbf{v}\in\mathbb{R}^n$ now represent a field of $n=2m$ variables supported by the grid S, A_0 is $n\times n$ symmetric positive-definite, and $\mathbf{b}_0\in\mathbb{R}^n$ represents a vector of size n. The discrete data term can then be rewritten as:

$$f_d(I,\mathbf{v}) = \frac{1}{2}\mathbf{v}^T A_0\mathbf{v} - \mathbf{b}_0^T\mathbf{v} + c_0, \tag{A1}$$

where $c_0\in\mathbb{R}$ denotes a scalar. The discretization of the self-similar constraints defined in Eq. (11) implies the discretization of the second order structure function defined in Eq. (7). The integral over θ yields a sum on 8 directions: 4 horizontal-vertical directions $\mathbf{n}_h=\{(-1,0),(1,0)\}$ and $\mathbf{n}_v=\{(0,1),(0,-1))\}$ and 4 diagonal directions $\sqrt{2}\mathbf{n}_d=\{(1,1),(1,-1),(-1,-1),(-1,1)\}$ of the bidimensional plane. The continuous motion field discretization yields to rewrite the integral over the spatial domain Ω with a sum over the pixel grid $\boldsymbol{S}$. On this regular grid, the longitudinal velocity increment function is available at scale ℓ either for $\ell\in\mathbb{N}^+$ on horizontal $\mathbf{n}_h$ and vertical $\mathbf{n}_v$ directions, or for $\sqrt{2}\ell\in\mathbb{N}^+$ on diagonal directions $\sqrt{2}\mathbf{n}_d$. To avoid using boundary conditions, we exclude of the sum the grid points $\overline{S}(\ell)$ which belongs to image borders of width ℓ. A node subset $S^\ell=\{S-\{\overline{S}(\ell)\}\}$ is thus defined depending on scale. Therefore, at scale ℓ, the discrete second order structure function reads $S_2\|(\ell)=:$

$$\begin{cases} \frac{1}{4\Gamma}\sum\limits_{\mathbf{s}\in S^\ell}\left\{\sum\limits_{\mathbf{n}_h}[u(\mathbf{s})-u(\mathbf{s}+\ell\mathbf{n})]^2+\sum\limits_{\mathbf{n}_v}[v(\mathbf{s})-v(\mathbf{s}+\ell\mathbf{n})]^2\right\}, \text{ if } \ell\in\mathbb{N}^+ \\ \frac{1}{4\Gamma}\sum\limits_{\mathbf{s}\in S^\ell}\sum\limits_{\mathbf{n}_d}[u(\mathbf{s})-u(\mathbf{s}+\ell\mathbf{n})+v(\mathbf{s})-v(\mathbf{s}+\ell\mathbf{n})]^2, \text{ if } \frac{\ell}{\sqrt{2}}\in\mathbb{N}^+ \end{cases} \tag{A2}$$

where we have denoted the number of nodes of the grid S^ℓ by $\Gamma=|S^\ell|$. The quadratic constraint derivatives can be expressed in the vectorial form $A_\ell\mathbf{v}-\mathbf{b}_\ell$, where A_ℓ are symmetric positive semi-definite matrices and $\mathbf{b}_\ell$ are vectors of size n. Indeed, manipulating the derivative of the expectation of Eq. (A2) w.r.t. the motion field $\mathbf{v}$, one obtains for grid points in the subset S^ℓ a new expression for the $\mathbf{s}$-th component of the constraint derivatives $\nabla_{\mathbf{v}(\mathbf{s})}g_\ell(\mathbf{v})=$

$$\begin{cases} -2\Gamma^{-1}\ell^2\left[D^\ell_{xx}u(\mathbf{s}),D^\ell_{yy}v(\mathbf{s})\right]^T, \text{ if } \ell\in\mathbb{N}^+ \\ -\Gamma^{-1}\ell^2\left[L^\ell_{\frac{\pi}{4}}u(\mathbf{s})+D^\ell_{xy}v(\mathbf{s}),D^\ell_{xy}u(\mathbf{s})+L^\ell_{\frac{\pi}{4}}v(\mathbf{s})\right]^T \text{ if } \frac{\ell}{\sqrt{2}}\in\mathbb{N}^+ \end{cases} \tag{A3}$$

where $L^\ell_{\frac{\pi}{4}}$ represents a discrete laplacian operator with a centered 2-nd order finite difference scheme defined on a grid rotated of $\pi/4$ with mesh point spacing equal to ℓ and where D^ℓ_{xx}, D^ℓ_{yy} and D^ℓ_{xy} are similar to the discretization of 2-nd order spatial derivatives in a centered 2-nd order finite difference scheme on a mesh with spacing of ℓ. The constraints can finally be written in their discrete form as

$$g_\ell(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T A_\ell\mathbf{v} = 0,\ \ \forall\ell\in\mathbf{I}, \tag{A4}$$

where $A_\ell\mathbf{v}=\nabla_{\mathbf{v}}g_\ell(\mathbf{v})$. Let us remark that the discretization of the self-similar constraints does not rely on any approximation, conversely to standard regularization schemes such as in Horn and Schunck (1981) where continuous spatial derivatives have to be approached by discrete operators. The self-simlar model introduced here consists thus in a multi-scale generalization of the first order regularizer where the smoothing parameters are in this case related to the power law and given by the Lagrange multipliers. The constraint motion estimation problem defined in Eq. (13) can thus be rewritten in its discrete form as,

$$(P)\begin{cases} \min_{\mathbf{v}}\ f_d(I,\mathbf{v}) = \frac{1}{2}\mathbf{v}^T A_0\mathbf{v} - \mathbf{b}_0^T\mathbf{v} + c_0. \\ \text{s.t. } g_\ell(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T A_\ell\mathbf{v} = 0,\ \ \forall\ell\in\mathbf{I} \\ \mathbf{v}\in\mathbb{R}^n. \end{cases} \tag{A5}$$

A2 Dual problem and optimal solution

We are now interested in defining optimality conditions taken into account the constraints. To this end, the La-

grangian function $L(\mathbf{v}, \boldsymbol{\lambda})$ is introduced as

$$L(\mathbf{v}, \boldsymbol{\lambda}) = f_d(I, \mathbf{v}) + \sum_\ell \lambda_\ell g_\ell(\mathbf{v}), \;\; \boldsymbol{\lambda} = \{\lambda_\ell\}. \qquad \text{(A6)}$$

In the Lagrangian duality formalism, the optimal solutions of the so-called *primal* problem P are obtained by searching saddle points of the Lagrangian function. The saddle points, denoted by $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$, are defined as the solutions of the so-called *dual* problem D, with

$$(D) \left\{ \begin{array}{l} L(\mathbf{v}^*,\boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}} w(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}}\{\min_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda})\} \\ \lambda_\ell \in \mathbb{R}, \; \forall \ell \in \mathbf{I} \end{array} \right. , \qquad \text{(A7)}$$

where $w(\boldsymbol{\lambda})$ denotes the *dual* function. As the functions f_d and g_ℓ are convex and as the constrained group is not empty, for positive Lagrangian multipliers λ_ℓ, L is convex and the minimization problem (P) has a unique *saddle point*, *i.e.*, an optimal solution $\mathbf{v}^*$ which is unique. More details of convex functionals and the convergence of the related minimization problems can be found in Weickert and Schnörr (2004). Lagrangian multipliers $\{\lambda_\ell\}$, which represent the regularization parameters at scales $\{\ell\}$, are then optimally given by the coordinates of the saddle point. Note that for negative Lagrangian multipliers, the convexity of the functional is no longer insured, and there is no guarantee of the solution unicity. Nevertheless, there still exist local optimal solutions.

A3 Convex optimization

We focus now in the strategy to obtain the saddle point solution of the constrained estimation problem. The minimum $\hat{\mathbf{v}}'$ of a locally convex Lagrangian function at point $\boldsymbol{\lambda}$ can be obtained by canceling the gradient

$$\nabla_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda}) = \nabla_{\mathbf{v}} f_d(I, \mathbf{v}) + \sum_\ell \lambda_\ell \nabla_{\mathbf{v}} g_\ell(\mathbf{v}) = 0, \qquad \text{(A8)}$$

which reduces (using (A1) and (A4)) to solving the large linear system

$$(A_0 + A)\hat{\mathbf{v}} = \mathbf{b}_0, \qquad \text{(A9)}$$

with $A = \sum_\ell \lambda_\ell A_\ell$. In appendix A1 we showed that the linear system components A_ℓ are constituted by the superposition of a collection of discrete operators obtained in a centered 2-nd order finite difference scheme on a grid of mesh ℓ, which corresponds to 2-nd order derivatives at different scales. Since we have no guarantee that the matrix $A_0 + A$ is positive-definite depending on the sign of Lagrangian multipliers λ_ℓ, the solution of the large system given by Eq. (A9) is efficiently achieved using a Conjugate Gradient Squared (CGS) method with an incomplete LU preconditioner. The dual function is then given by

$$w(\boldsymbol{\lambda}) = \frac{1}{2}\hat{\mathbf{v}}^T \Big(A_0 + A\Big)\hat{\mathbf{v}} - \mathbf{b}_0^T \hat{\mathbf{v}} + c_0. \qquad \text{(A10)}$$

The dual function is by definition concave and possesses so-called *sub-gradients* equal to $g_\ell(\hat{\mathbf{v}})$. We employ a classical gradient method to find $\boldsymbol{\lambda}^*$ which maximizes the dual function and thus obtain the solution $\mathbf{v}^*$. Finally, the constrained motion estimation method results in a *Uzawa* algorithm, which is used to converge towards the saddle point $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$, *i.e.*, the optimal motion estimate under self-similar constraints. An important remark is that once the regularization coefficient vector $\boldsymbol{\lambda}^*$ has been estimated for two consecutive images of the sequence, assuming motion stationarity, only one step of the *Uzawa* algorithm is needed to process the following image pairs. Therefore, the complexity of the algorithm reduces to the resolution of the linear system by CGS, that is $\mathcal{O}(\kappa n)$, where κ is the conditioning number of $A_0 + A$. The Uzawa algorithm is briefly presented herefater.

- Iterate until convergence from point $\boldsymbol{\lambda}^{(k=0)} \neq 0$:

  – At iteration k, compute motion $\hat{\mathbf{v}}$ solution of Eq. (A9).

  – Define $\boldsymbol{\lambda}^{(k+1)}$ by a gradient ascent step:

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} + \rho^{(k)} \mathbf{P}^{(k)} g(\hat{\mathbf{v}}),$$

- At convergence: $(\mathbf{v}^*, \boldsymbol{\lambda}^*) = (\hat{\mathbf{v}}, \boldsymbol{\lambda}^{(k=\infty)})$.

where $g(\hat{\mathbf{v}})$ represents a vector whose components are $g_\ell(\hat{\mathbf{v}})$, and where $\rho^{(k)}$ and $\mathbf{P}^{(k)}$ denote the step size and a positive-definite matrix defined at iteration k. The latter parameter and matrix can be fixed at each iteration in order to formalize a Quasi-Newton algorithm satisfying the strong Wolf conditions (Nocedal and Wright, 1999).

Uzawa algorithm converging towards saddle point $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$**.**

APPENDIX B: Maximum likelihood estimate of the observation model error variance

In order to obtain an analytical expression of the variance $\hat{\beta}^{-1}$ of the observation model error maximizing the *evidence*, we first detail analytically the three different partition function composing the *evidence*, then give an analytical expression of its derivate w.r.t. β, and finally provide an expression for the maximum *likelihood* estimate $\hat{\beta}$ canceling out the *evidence* derivative.

B1 Evaluation of the evidence for β

First, the *likelihood* PDF $p(I|\mathbf{v}, \beta)$ energy related to a quadratic observation model $\beta f_d(I, \mathbf{v})$ is composed of the energy of a multidimensional Gaussian with m uncorrelated components of variance β, where m is the number of point composing the image grid. Its partition function reads:

$$Z_{f_d}(\beta) = \left(\frac{2\pi}{\beta}\right)^{m/2}. \qquad \text{(B1)}$$

Therefore, we obtain:

$$-\log Z_{f_d}(\beta) = -\frac{m}{2}\log(2\pi) + \frac{m}{2}\log(\beta) \qquad \text{(B2)}$$

Secondly, the partition function $Z_L(\beta, \mathcal{M})$ of the *posterior* PDF of (19) is an integral which can be calculated using Laplace's approximation (see for example MacKay (2003)):

$$Z_L(\beta,\mathcal{M}) = \int_{\mathbb{R}^n} \exp\{-\beta L(\mathbf{v},\boldsymbol{\lambda}^*)\}\mathrm{d}\mathbf{v} \simeq \exp^{\{-\beta(f_d(I,\mathbf{v}^*)+\sum_\ell \lambda_\ell^* g_\ell(\mathbf{v}^*,\mathcal{M}))\}} 2\pi^{\frac{n}{2}} \det\{\beta(A_0+\sum_\ell \lambda_\ell^* A_\ell)\}^{-\frac{1}{2}}. \quad \text{(B3)}$$

where we recall that $\mathbf{v}^*$ is the MAP estimate, and $\boldsymbol{\lambda}^*$ is the associated set of Lagrangian multipliers. Noting that the term $\sum_\ell \lambda_\ell^* g_\ell(\mathbf{v}^*,\mathcal{M})$ vanishes since constraints are fulfilled, this yields:

$$-\log Z_L(\beta,\mathcal{M}) = \beta f_d(I,\mathbf{v}^*) - \frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\beta) + \frac{1}{2}\log\det\{A_0+\sum_\ell \lambda_\ell^* A_\ell\} \quad \text{(B4)}$$

where $n = 2m$ denotes the number of unknown velocity variables. For Gaussian distributions, the Laplace equality is exact and for other mono-modal distributions it still constitutes a good approximation (Gull, 1989). The determinant of such large and sparse matrices can be efficiently approximated via an incomplete LU decomposition.

Finally, the partition function $Z_{g_\ell}(\mathcal{M})$ of the self-similar *prior* of (18) does not exist since the PDF is degenerated and has an infinity set of maxima, corresponding to the infinite set of admissible velocity field solutions respecting the self-similar constraint. To make this *prior* well-defined, we use Dirichlet boundary conditions for evaluating the *evidence*[8]. Considering these boundary conditions, we get a Hessian matrix A which has a slightly changed form but which is full rank. As previously, the partition function can be calculated using Laplace's approximation:

$$Z_{g_\ell}(\mathcal{M}) = \int_{\mathbb{R}^n} \exp\{-\beta\sum_\ell \lambda_\ell^* g_\ell(\mathbf{v},\mathcal{M})\}\mathrm{d}\mathbf{v} = \max_{\mathbf{v}}\Big(\exp^{\{-\beta\sum_\ell \lambda_\ell^* g_\ell(\mathbf{v},\mathcal{M})\}}\Big) 2\pi^{\frac{n}{2}}\det\{\beta\sum_\ell \lambda_\ell^* A_\ell\}^{-\frac{1}{2}} \quad \text{(B5)}$$

As the set of admissible solutions for the self-similar constraint is not empty, the maximum value of the exponential term in (B5) is equal to 1, and one gets:

$$-\log Z_{g_\ell}(\mathcal{M}) = -\frac{n}{2}\log(2\pi) + \frac{n}{2}\log(\beta) + \frac{1}{2}\log\det\{\sum_\ell \lambda_\ell^* A_\ell\} \quad \text{(B6)}$$

Replacing (B2), (B4), (B6) in (23), one obtains an analytical expression for minus the *log evidence*:

$$-\log p(I|\beta,\mathcal{M}) = \underbrace{\beta f_d(I,\mathbf{v}^*)}_{\text{-log likelihood}} + \underbrace{\frac{1}{2}\log\frac{\det\{A_0+\sum_\ell \lambda_\ell^* A_\ell\}}{\det\{\sum_\ell \lambda_\ell^* A_\ell\}} - \frac{m}{2}\log(\frac{\beta}{2\pi})}_{\text{-log Occam's factor}} \quad \text{(B7)}$$

[8] Note that the precise values on the boundaries do not need to be specified since they modify the content of vector $\mathbf{b}$ but do not have any impact on the Hessian matrix A. Indeed, the form of the matrix is slightly changed to code Dirichlet boundary conditions but its content is independent of the boundary function.

The *evidence* energy can be interpreted as the balance between three terms. The left term is the *log likelihood*. It represents the misfit of the data to the observation model. The two other terms appearing on the second line penalize the model complexity. They are known as the log of *Occam's factor* (Jaynes, 2003; MacKay, 2003). Of these, the first is the ratio of the *posterior* accessible volume on the *prior* accessible volume in $\mathbf{v}$ (a variance ratio in 1D). The term most to the right is a normalization term of *Occam's factor* for non-unitary variance β.

B2 Derivative of the evidence and maximum likelihood estimate $\hat{\beta}$

To finally achieve the estimation of the observation model variance, we now want to minimize the energy of the *evidence* PDF. Noting that for any matrix C

$$\frac{d}{d\beta}\log\det C = \mathrm{trace}(C^{-1}\frac{dC}{d\beta}), \quad \text{(B8)}$$

one obtains the following equality:

$$\frac{d}{d\beta}\log\frac{\det\{A_0+\sum_\ell \lambda_\ell^* A_\ell\}}{\det\{\sum_\ell \lambda_\ell^* A_\ell\}} = \frac{d}{d\beta}\log\det\{\beta A_0+\sum_\ell \lambda_\ell^*(\mathcal{M})\beta A_\ell\} = \frac{1}{\beta}\mathrm{trace}((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1}A_0). \quad \text{(B9)}$$

Note that in the last expression $\lambda_\ell^*(\mathcal{M})\beta$ (resp. $\lambda_\ell^*(\mathcal{M})$) does not (resp. does) depend on β. Using (B9), the derivative of *log evidence* (B7) w.r.t. β can be written as:

$$\frac{\partial}{\partial\beta}(\log p(I|\beta,\mathcal{M})) = \frac{1}{2}\left(\frac{m}{\beta} - 2f_d(I,\mathbf{v}^*) - \frac{1}{\beta}\mathrm{trace}((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1}A_0)\right). \quad \text{(B10)}$$

The *log evidence* is maximum at the point where the derivative cancels. Therefore, an analytical expression of the maximum *likelihood* estimate of the observation model inverse variance reads:

$$\hat{\beta} = \frac{m - \mathrm{trace}((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1}A_0)}{2f_d(I,\mathbf{v}^*)}. \quad \text{(B11)}$$

Calculating the trace of the high-dimensional matrices involved in the previous equations can be very demanding. In order to overcome this problem, we use a trace randomization technique Skilling (1989) which is efficiently achieved using the CGS algorithm. More precisely, $L \gg 1$ random samples of a n-dimensional normalized and centered Gaussian distribution $r_j \sim \mathcal{N}(0,\mathbf{1}^n)$ are used to approximate the

trace operator:

$$\begin{aligned}\text{trace}((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1}A_0) &= \text{trace}\mathbb{E}[(A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1} r_j r_j^t A_0] \\ &= \mathbb{E}[((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1} r_j)^t (A_0 r_j)] \\ &\simeq \frac{1}{L}\sum_{j=1}^{L}[(A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1} r_j]^t A_0 r_j \end{aligned} \quad \text{(B12)}$$

where the vector $(A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1} r_j$ is the solution provided by the CGS algorithm of the inverse problem $(A_0+\sum_\ell \lambda_\ell^* A_\ell)\mathcal{X} = r_j$, where $\mathcal{X}$ is the unknown.

APPENDIX C: Likelihood probability of the prior power law

Since we have assumed that the *prior* $p(\beta)$ is flat, the *evidence* is maximum at $\hat{\beta}$. Using Laplace's Gaussian approximation, the *prior* model *evidence* reads:

$$p(I|\mathcal{M}) \propto p(I|\hat{\beta},\mathcal{M})\sqrt{2\pi}\sigma_\beta, \quad \text{(C1)}$$

where the inverse variance σ_β^2 is defined as a second order derivative w.r.t. β of the *evidence* energy:

$$\begin{aligned}(\sigma_\beta^2)^{-1} &= \frac{\partial^2}{\partial\beta^2}\left(-log p(I|\beta,\mathcal{M})\right) \\ &= \frac{1}{2}\left(\frac{\partial}{\partial\beta}\left(\frac{\text{trace}((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1}A_0)}{\beta}\right)+\frac{m}{\beta^2}\right) \\ &\simeq \frac{1}{2\beta^2}\left(-\text{trace}((A_0+\sum_\ell \lambda_\ell^* A_\ell)^{-1}A_0)+m\right) \quad \text{(C2)}\end{aligned}$$

Using (B11), we obtain an approximation of the standard deviation σ_β at $\hat{\beta}$:

$$\sigma_\beta = \sqrt{\frac{\hat{\beta}}{f_d(I,\mathbf{v}^*)}} \quad \text{(C3)}$$

Finally, we obtain the energy of the *evidence* for model $\mathcal{M}$:

$$\begin{aligned}-\log p(I|\mathcal{M}) \propto & 2\hat{\beta} f_d(I,\mathbf{v}^*) + \log\frac{\det\{A_0+\sum_\ell \lambda_\ell^* A_\ell\}}{\det\{\sum_\ell \lambda_\ell^* A_\ell\}} \\ & -(m+1)\log\hat{\beta} + \log(f_d(I,\mathbf{v}^*)) \quad \text{(C4)}\end{aligned}$$

The *prior* model canceling out the derivative of the previous energy can not be analytical obtained as it involves a inverse matrix operator. Therefore, in order to minimize the energy of the *evidence* for $\mathcal{M}$, the parameter vector (γ,ζ) is sampled uniformly in $\mathbb{R}^2$. The maximum *likelihood* estimate $\hat{\mathcal{M}}$ is the minimizer of (C4) and represent the selected model.

APPENDIX D: Power law inference for laminar flows

Laminar flows are characterized by flat structure functions. Indeed, for such flows, increments are equal to zero. Therefore, power-laws with vanishing prefactor $\gamma = 0$ constitute proper models for their structure functions. In the following we use a toy example in order to demonstrate the capacity of the proposed algorithm to recover accurately this particular case. Wu used a synthetic image (of 64 × 64 pixels) seeded with particles, which is evolved by a laminar flow (a translation by 1 pixel) to form the second image of the pair. Figure 16 shows that the inference algorithm properly recovers the flat power law characterizing such motions. It obviously corresponds to low RMS motion reconstruction errors as shown in the right part of the figure.

Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant MS-DAG (ANR-08-SYSC-014) "Multiscale Data Assimilation for Geophysics"

REFERENCES

Adrian, R., 1991: Particle imaging techniques for experimental fluid mechanics. *Annal Rev. Fluid Mech.*, 23, 261–304.

Baker, S. and coauthors, 2007: A database and evaluation methodology for optical flow. In *Int. Conf. on Comp. Vis., ICCV 2007*.

Benzi, R. and R. Scardovelli, 1995: Intermittency of two-dimensional decaying turbulence. *Europhys. Lett.*, 29, 371–376.

Bergen, J., P. Burt, R. Hingorani, and S. Peleg, 1992: A 3-frame algorithm for estimating two-component image motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(9), 886–895.

Berliner, M., C. Wikle, and R. Milliff, 1999: Multiresolution wavelet analyses in hierarchical bayesian turbulence models. *Bayesian Inference in Wavelet Based Models, Lecture Notes in Statistics*, 141, 341–349.

Bernard, D., G. Boffetta, A. Celani, and G. Falkovich, 2006: Conformal invariance in two-dimensional turbulence. *Nature Physics*, 2, 124–128.

Boffetta, G., 2007: Energy and enstrophy fluxes in the double cascade of two-dimensional turbulence. *J. Fluid Mech.*, 589, 253–260.

Carlier, J. and B. Wieneke, 2005: Report 1 on production and diffusion of fluid mechanics images and data. *Fluid project deliverable 1.2. http://www.fluid.irisa.fr*.

Charney, J., 1971: Geostrophic turbulence. *J. Atmos. Sci*, 28, 1087–1095.

Cho, J. and E. Lindborg, 2001: Horizontal velocity structure functions in the upper troposphere and lower stratosphere 1. observations. *J. Geophysical Research.*, 106, 223–232.

Corpetti, T., P. Heas, E. Memin, and N. Papadakis, 2009: Pressure image assimilation for atmospheric motion estimation. *Tellus Series A: Dynamic Meteorology and Oceanography*, 61, 160–178.

Corpetti, T., E. Mémin, and P. Pérez, 2002: Dense estimation of fluid flows. *Pattern Anal Mach Intel*, 24(3), 365–380.

Dewan, E., 1997: Saturated-cascade similitude theory of gravity wave spectra. *Journal of Geophysical Research*, 102, 799–818.

Falkovich, G., K. Gawędzki, and M. Vergazzola, 2001: Particles and fields in fluid turbulence. *Rev. Mod. Phys.*, 73, 913–975.

Fitzpatrick, J., 1988: The existence of geometrical density-image transformations corresponding to object motion. *Comput. Vision, Graphics, Image Proc.*, 44(2), 155–174.

Frisch, U., 1995: *Turbulence : the legacy of A.N. Kolmogorov*. Cambridge university press.

Gotoh, T., D. Fukayama, and T. Nakano, 2002: Velocity field statistics in homogeneous steady turbulence using a high-resolution direct numerical simulation. *Physics of Fluids*, 14, 1065–1081.

Gull, S. F., 1989: Developments in maximum-entropy data analysis. In *Maximum Entropy and Bayesian Methods*, Kluwer Academic, 53–71.

Heas, P. and E. Memin, 2008: Three-dimensional motion estimation of atmospheric layers from image sequences. *IEEE trans. on Geo. and Rem. Sensing*, 46(8), 2385–2396.

Heas, P., E. Memin, D. Heitz, and P. Mininni, 2009: Bayesian selection of scaling laws for motion modeling in images. In *International Conference on Computer Vision (ICCV'09)*, Kyoto, Japan.

Heas, P., E. Memin, N. Papadakis, and A. Szantai, 2007a: Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *IEEE trans. on Geo. and Rem. Sensing*, 45(12), 4087–4104.

Heas, P., E. Memin, N. Papadakis, and A. Szantai, 2007b: Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *IEEE trans. on Geo. and Rem. Sensing*, 45(12), 4087–4104.

Heitz, D., J. Carlier, and G. Arroyo, 2007: Final report on the evaluation of the tasks of the workpackage 2. Technical report, Deliverable 5.4, Cemagref, Rennes, France.

Heitz, D., E. Memin, and C. Schnoerr, 2010: Variational fluid flow measurements from image sequences: synopsis and perspectives. *Exp. Fluids*, 48(3), 369–393.

Hoar, T. J., R. F. Milliff, D. Nychka, C. K. Wikle, and L. M. Berliner, 2003: Winds from a bayesian hierarchical model: Computation for atmosphere-ocean research. *Journal of Computational and Graphical Statistics*, 12(4), pp. 781–807.

Horn, B. and B. Schunck, 1981: Determining optical flow. *Artificial Intelligence*, 17, 185–203.

Jaynes, E. T., 2003: *Probability Theory: The Logic of Science*. G. Larry Bretthorst.

Kolmogorov, A., 1941: The local structure of turbulence in inompressible viscous fluid for very large reynolds number. *Dolk. Akad. Nauk SSSR*, 30, 301–5.

Kraichnan, R., 1967: Inertial ranges in two-dimensional turbulence. *Phys. Fluids*, 10, 1417–1423.

Kurien, S., V. S. L'vov, I. Procaccia, and K. R. Sreenivasan, 2000: The scaling structure of the velocity statistics in atmospheric boundary layer. *Physical Review E (Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics)*, 61(1), 407–421.

Leese, J., C. Novack, and B. Clark, 1971: An automated technique for obtained cloud motion from geosynchronous satellite data using cross correlation. *Journal of applied meteorology*, 10, 118–132.

Lindborg, E., 1999: Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence. *J. Fluid Mech.*, 388, 259–288.

Lindborg, E., 2007: Horizontal wavenumber spectra of vertical vorticity and horizontal divergence in the upper troposphere and lower stratosphere. *Journal of the Atmospheric Sciences*, 64(3), 1017–1025.

Lindborg, E. and J. Cho, 2001: Horizontal velocity structure functions in the upper troposphere and lower stratosphere 2. theoretical considerations. *J. Geophysical Research.*, 106, 233–241.

Liu, T. and L. Shen, 2008: Fluid flow and optical flow. *Journal of Fluid Mechanics*, 614, 253.

Lucas, B. and T. Kanade, 1981: An iterative image registration technique with an application to stereovision. In *Int. Joint Conf. on Artificial Intel. (IJCAI)*, 674–679.

MacKay, D. J. C., 1992: Bayesian interpolation. *Neural Computation*, 4(3), 415–447.

MacKay, D. J. C., 2003: *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.

Monin, A. and A. Yaglom, 1971: *Statistical Fluid Mechanics: Mechanics of Turbulence*. JDover Pubns.

Nastrom, G., W. Jasperson, and K. Gage, 1984: Kinetic energy spectrum of large and mesoscale atmospheric processes. *Nature*, 310, 36–38.

Nastrom, G. D. and F. D. Eaton, 1997: Turbulence eddy dissipation rates from radar observations at 5-20 km at White Sands Missile Range, New Mexico. *Journal of geophysical research*, 102, 19495–19506.

Nocedal, J. and S. J. Wright, 1999: *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, New York, NY.

Papadakis, N. and E. Memin, 2008: Variational assimilation of fluid motion from image sequences. *SIAM Journal on Imaging Science*, 1(4), 343–363.

Pedlosky, 1987: *Geophysical Fluid Dynamics*. Springer,New York.

Robert, C. P., 2007: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, 577.

Skilling, J., 1989: The eigenvalues of mega-dimensional matrices. In *Maximum Entropy and Bayesian Methods*, 455-466.

Taylor, M. A., S. Kurien, and G. L. Eyink, 2003: Recovering isotropic statistics in turbulence simulations: The kolmogorov 4/5th-law. *Physical review E*, 2, 026310.

Tung, K. K., 2003: The k 3 and k 5/3 energy spectrum of the atmospheric turbulence: quasi-geostrophic two level model simulation. *J. Atmos. Sci*, 60, 824–835.

Weickert, J. and C. Schnörr, 2004: A theoretical framework for convex regularizers in pde-based computation of image motion. *Int. J. Computer Vision* 245–264.

Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner, 2001: Spatio-temporal hierarchical bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association*, 96, 382–397.

Yuan, J., C. Schnoerr, and E. Memin, 2007: Discrete orthogonal decomposition and variational fluid flow estimation. *Journ. of Math. Imaging & Vison*, 28, 67–80.

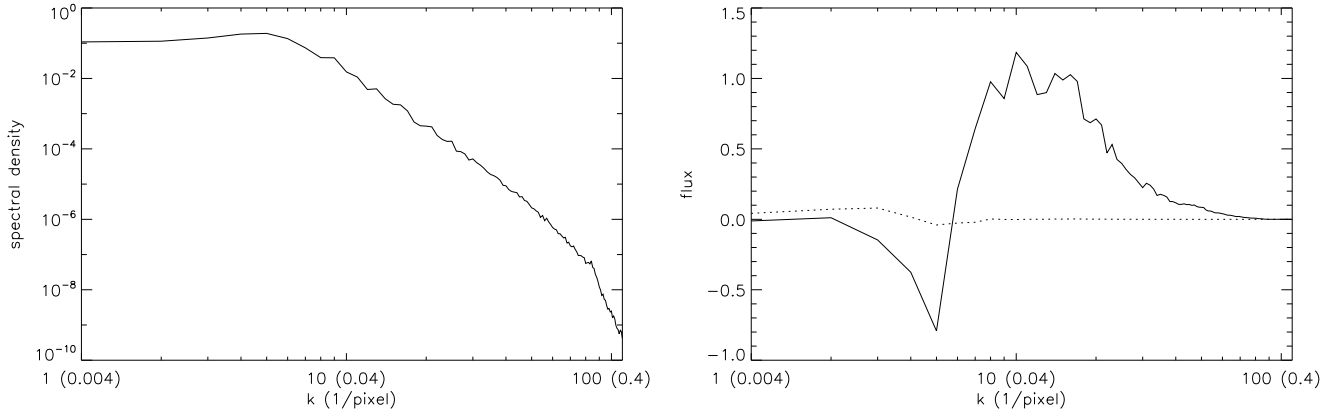

Fig. 1. Isotropic energy spectrum and flux in the numerical simulation. Left: energy spectrum, as a function of the wavenumber in units of the inverse of the length of the box, and in units of 1/pixel (between parenthesis). Right: enstrophy flux (solid) and energy flux (dotted), as a function of the wavenumber using the same notation.

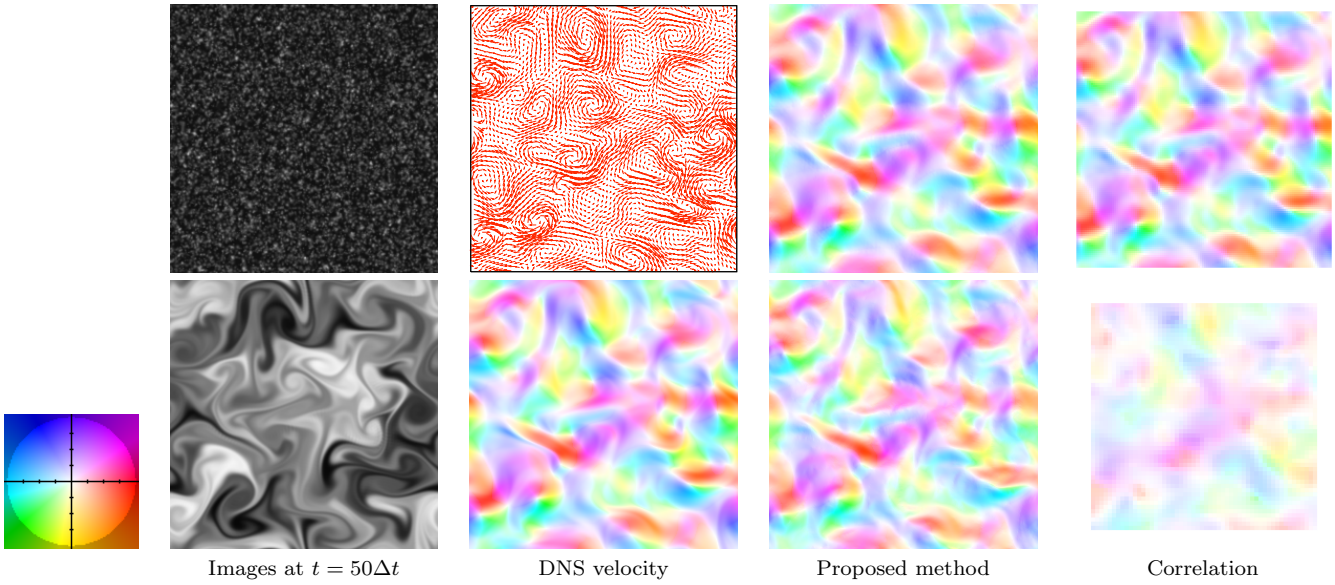

Fig. 2. Particle and scalar images, DNS and motion estimates. From left to right: Particle (above) and scalar (below) image at time $t = 50\Delta t$; Turbulent motion field obtained by DNS in vectorial (above) and color (below) representation; Motion field inferred by the proposed method and correlation-based velocity fields provided by LaVision in the context of the *Fluid FET Open* european project (Heitz et al., 2007) for particles (above) and scalar (below). For a better visual analysis, the velocity fields are displayed in a color representation where color and intensity code respectively vector orientations and magnitudes (Baker and coauthors, 2007). The corresponding color map circle is displayed on the left.

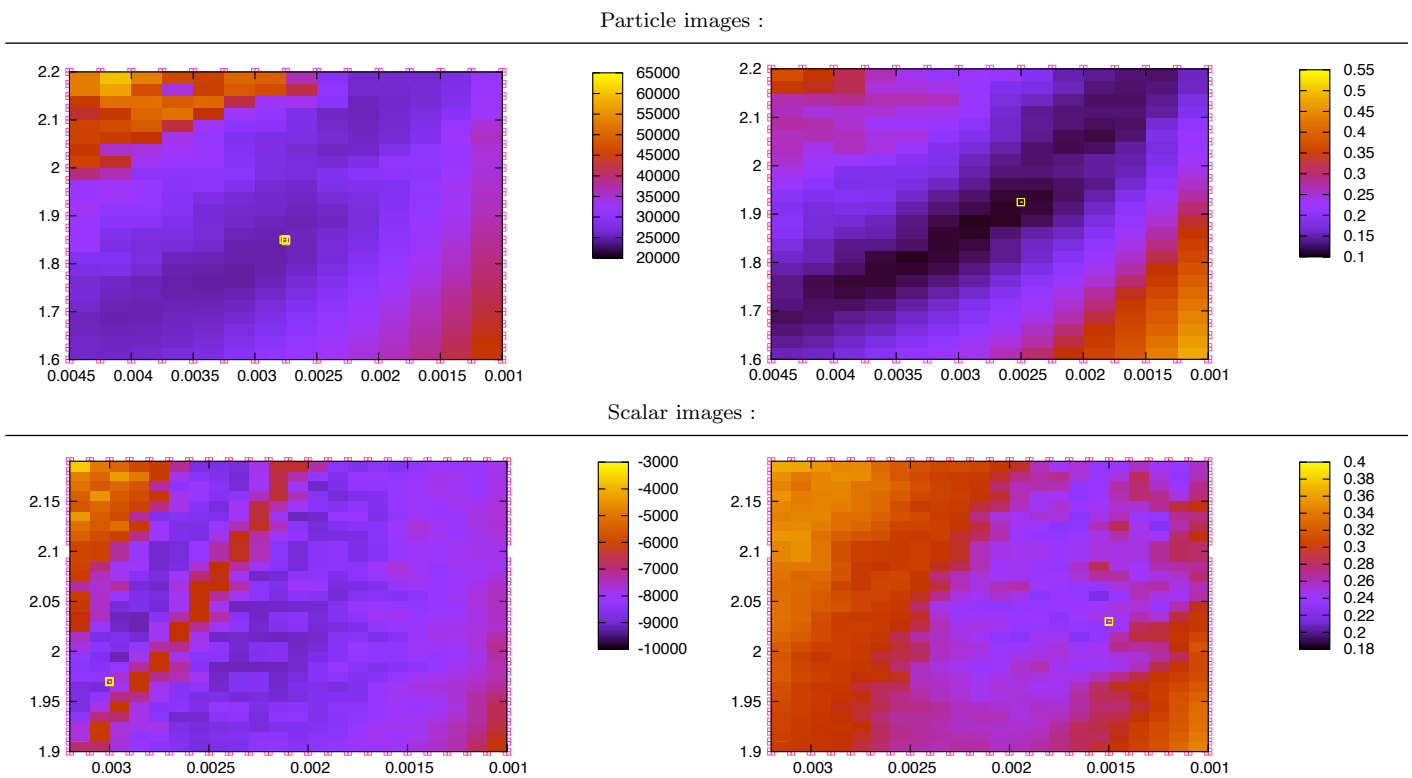

Fig. 3. Energy of probability $p(\mathcal{M}|I)$ and RMS error behavior w.r.t. power-law parameters. Plot of minus of the log of the power-law model evidence probability (left) and the motion RMS error (right) w.r.t. power-law exponent ζ (ordinate) and prefactor γ (absciss) for the particle (above) and scalar (below) image sequences. The coordinates of the minima are plotted with yellow squares. These coordinates have to be compared with the true power law parameters $(\gamma, \zeta) = (0.003, 1.948)$, given by a least square fit of the DNS data.

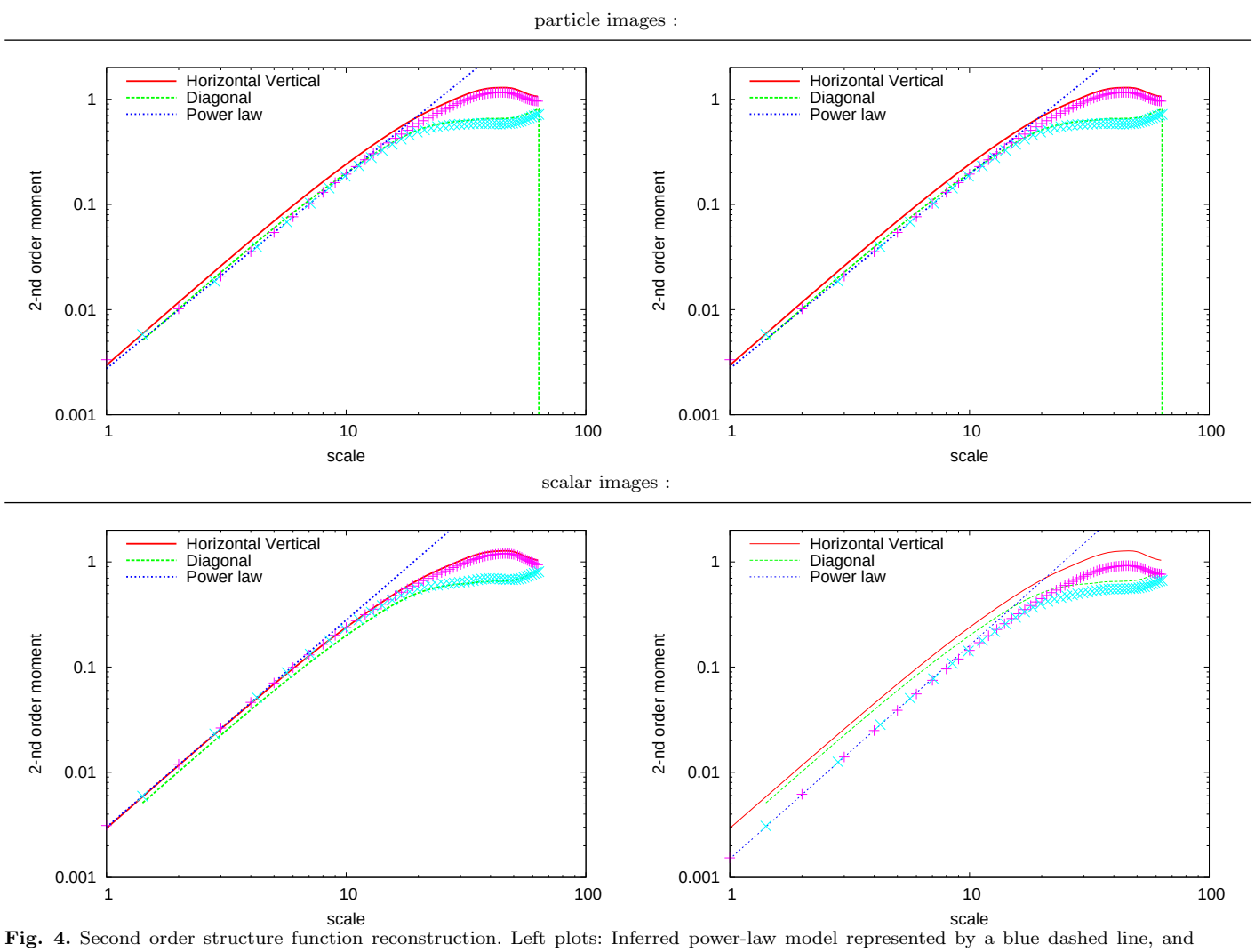

Fig. 4. Second order structure function reconstruction. Left plots: Inferred power-law model represented by a blue dashed line, and estimated (resp. true) second order structure function in horizontal-vertical directions plotted with stars (resp. continuous line) and in diagonal directions with cross (resp. dashed line). Right plots: identical legend for the motion estimate obtained with the prior power-law model minimizing the RMS error.

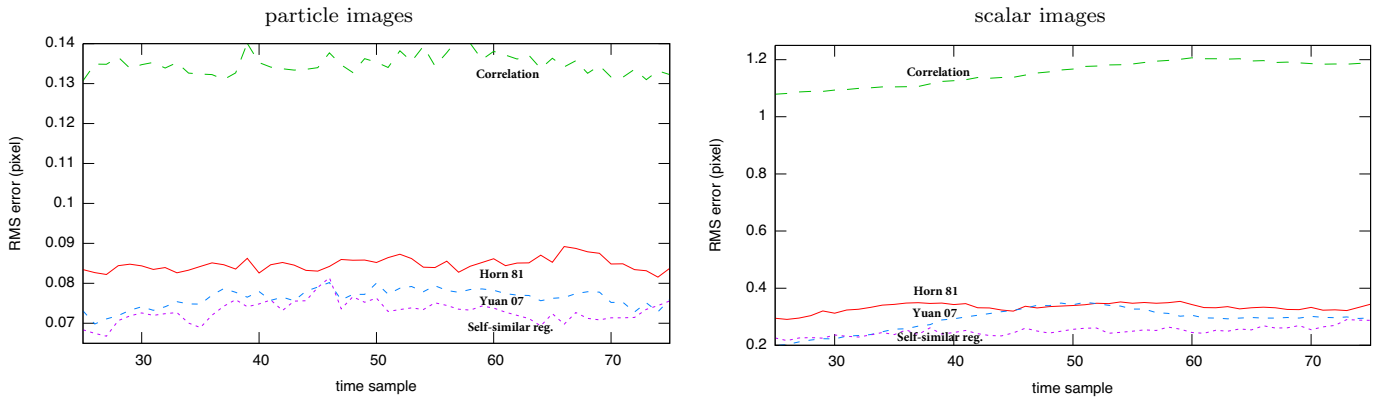

Fig. 5. Motion estimation accuracy. Evolution of RMS error w.r.t. time index for an operational correlation-based method, a robust first order regularizer (Horn and Schunck, 1981), a second order regularizer (Yuan et al., 2007), and the inferred self-similar constraints for the particle (left) and scalar (right) image sequences. The results obtained with correlation approach were provided by LaVision (*www.lavision.de*) company from their operational PIV software (Davis) in the context of the *Fluid* FET Open european project.

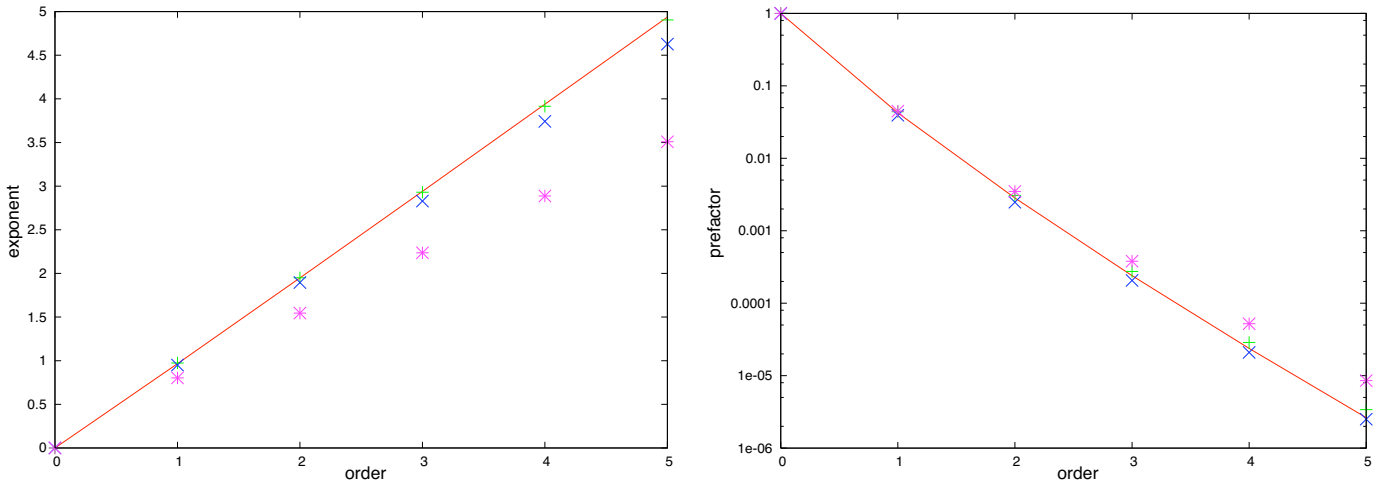

Fig. 6. Longitudinal structure function exponents and prefactor w.r.t. their order for the scalar image sequence. The proposed self-similar regularization (crosses +) provides exponents (left plots) and prefactors (right plots) (in a least square sense) very close to the ground truth (solid line) in comparison to estimates of (Horn and Schunck, 1981) (stars $*$) or (Yuan et al., 2007) ($\times$ symbols). Correlation-based measurements are not presented here since they do not provide motion increments in the bottom of the scale range of [1,8] pixels.

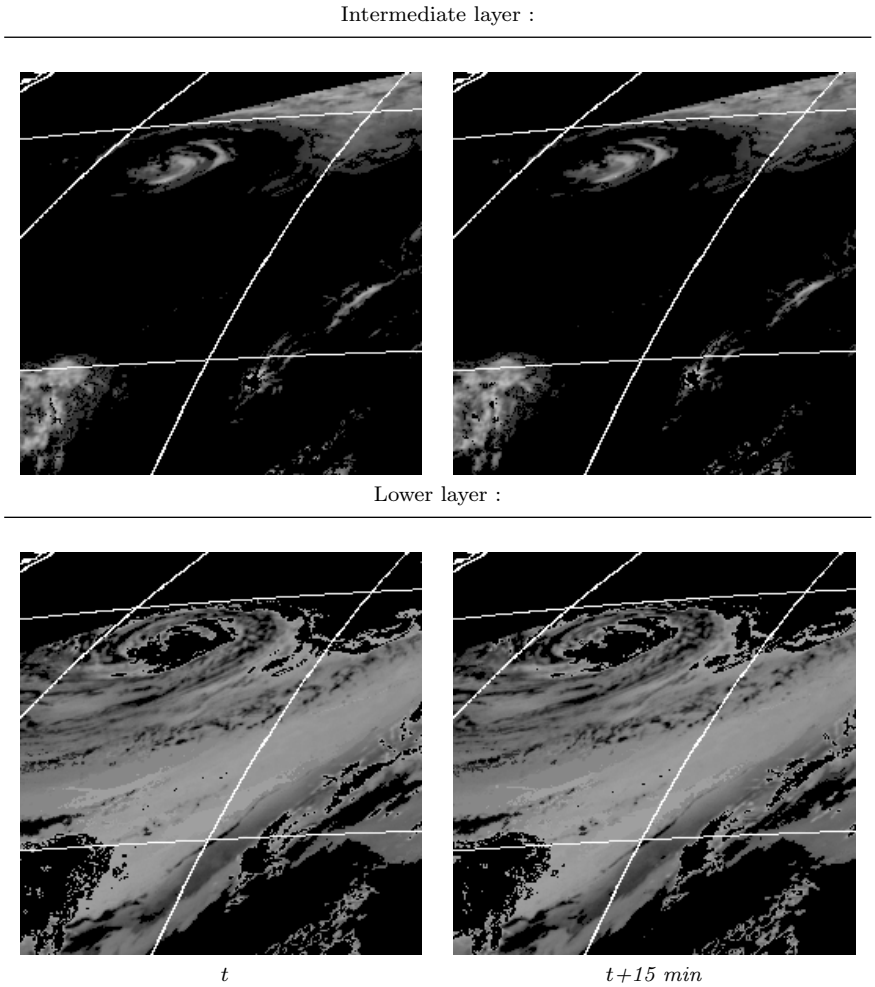

Fig. 7. Depression over the north-Atlantic Ocean. Sequence of images depicting sparse pressure difference maps of layers at intermediate (above) and low (below) altitude. The images characterize the layers evolution in a time interval of 15 minutes and with an average spatial resolution of 3 km. Black regions correspond to missing observations and white lines represent meridians (20^o, 30^o and 40^o), parallels (50^o and 60^o) and the coastal contours of south of Greenland (upper left corner).

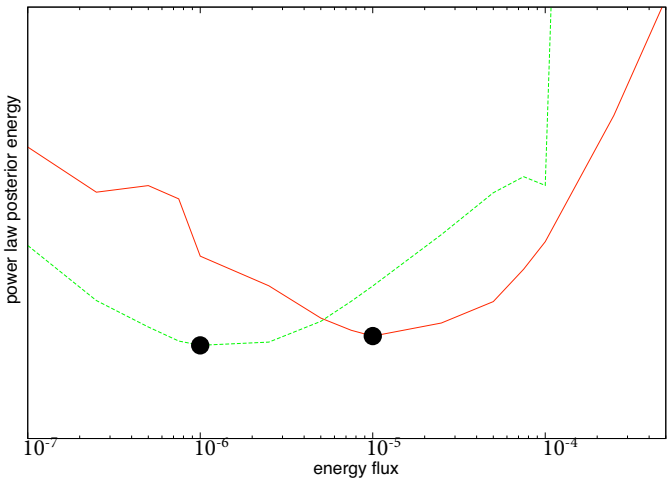

Fig. 8. Selection of most likely energy flux for *Lindborg*'s model. Minus log of the power law *posterior* probability $p(\mathcal{M}|I)$ vs. energy flux ϵ in $m^2 s^{-3}$ for horizontal winds at low (solid line) and at intermediate (dashed line) altitude

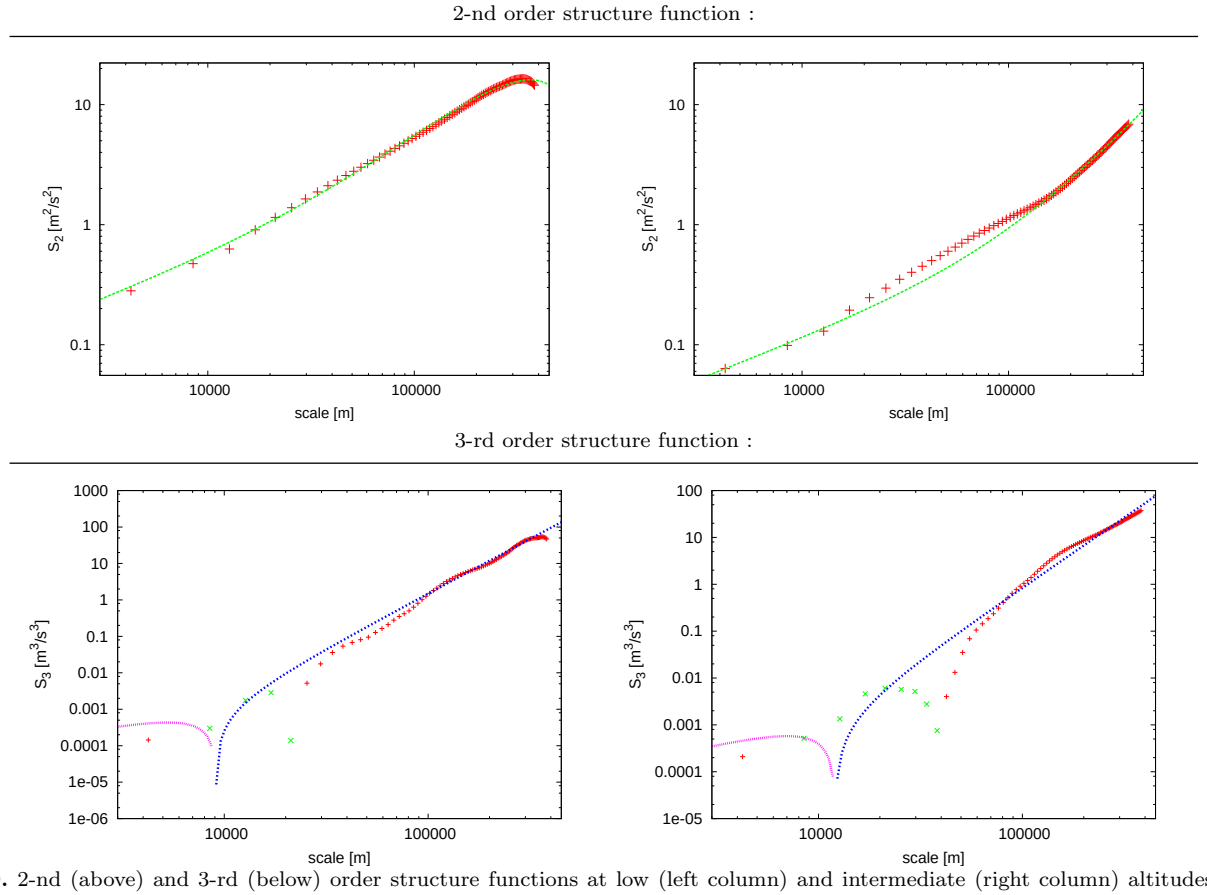

Fig. 9. 2-nd (above) and 3-rd (below) order structure functions at low (left column) and intermediate (right column) altitudes. 2-nd order structure functions (+ symbols) are plotted with their associate models (dashed line). 3-rd order structure functions (plotted with × symbols for positive values and with + symbols for negative values) can be compared to their associate models (fine dashed line for positive values and coarse dashed line for negative values).

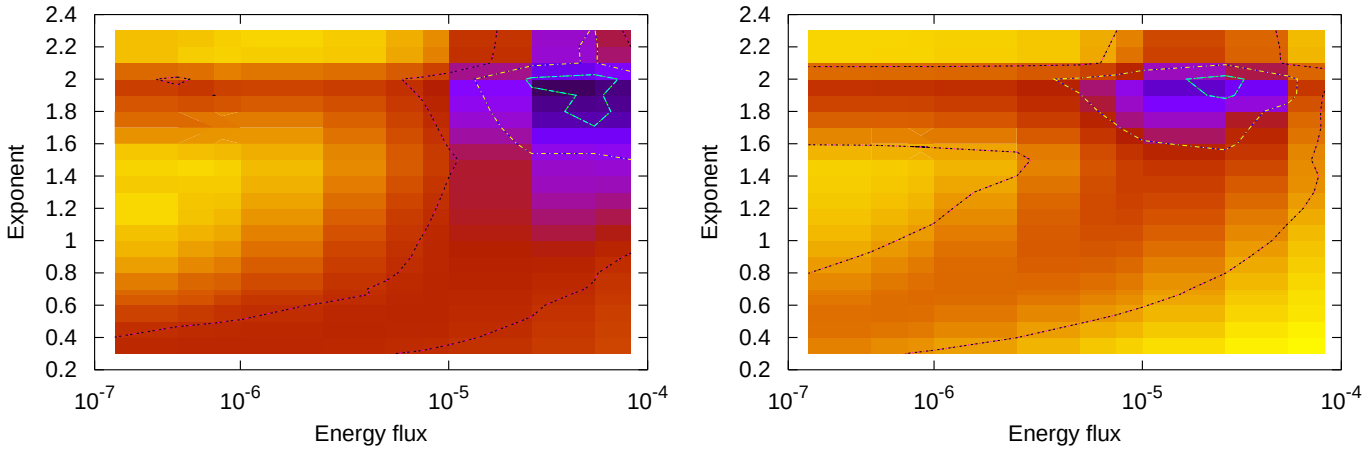

Fig. 10. Power law evidence probability. Minus log of model probability $p(\mathcal{M}|I)$ w.r.t. parameters: exponent ζ (ordinate) and energy flux ϵ in $m^2 s^{-3}$ (absciss) (*i.e.*, prefactor $\gamma = C_2 \epsilon^{\zeta}$), for horizontal winds at low (left) and at intermediate (right) altitude. Iso-contours of decreasing values around the minima are plotted in dark blue, yellow and turquoise.

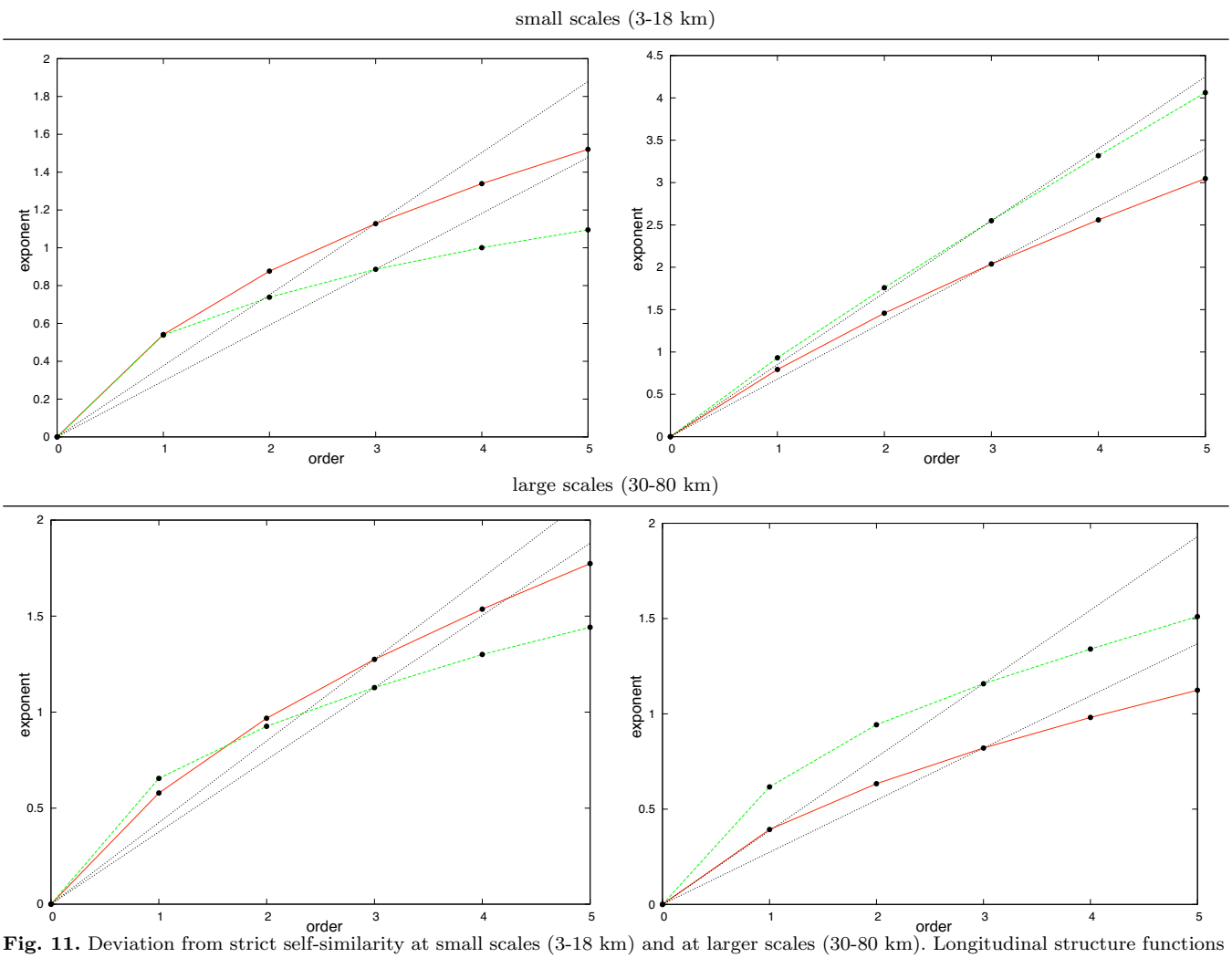

Fig. 11. Deviation from strict self-similarity at small scales (3-18 km) and at larger scales (30-80 km). Longitudinal structure functions exponents w.r.t. their order at low (solid curve) and intermediate (dashed curve) altitude using the model in (Lindborg and Cho, 2001) (on the left) or a flat *prior* (on the right) for power law models. The dashed straight lines represent strict self-similar behavior, *i.e.*, a linear relation of exponents w.r.t. order for the two models.

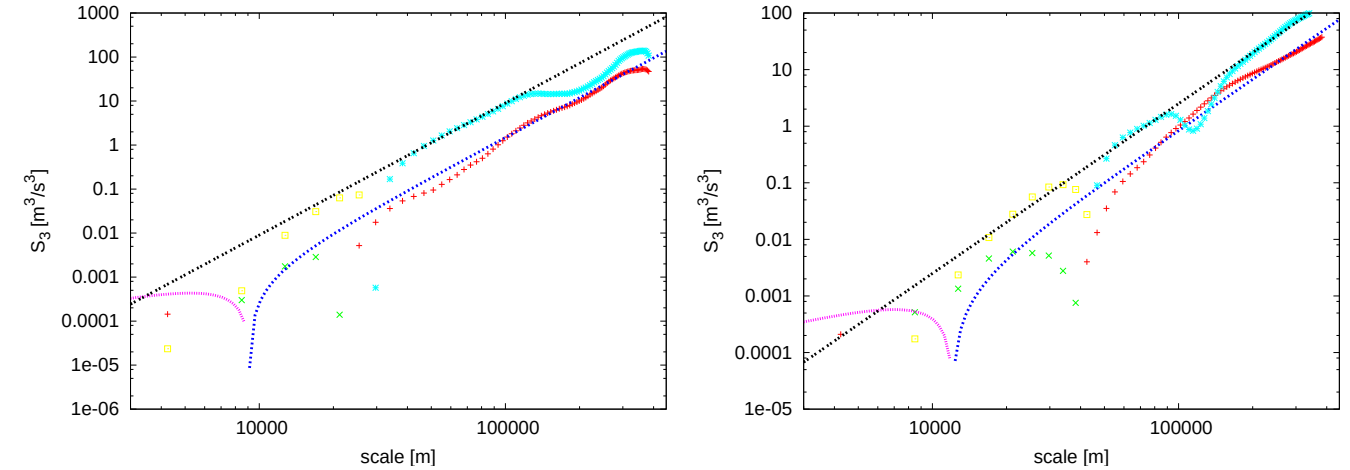

Fig. 12. 3-rd order structure functions' absolute value at low (left) and intermediate (right) altitudes for the two methods. Structure functions obtained with a (Lindborg and Cho, 2001) (resp. a flat) *prior* are plotted with × symbols (resp. □ symbols) for positive values and with + symbols (resp. ∗ symbols) for negative values. Estimate obtained with the (Lindborg and Cho, 2001) prior can be compared to their associate models (fine dashed curve for positive values and coarse dashed curve for negative values). At large scales, estimate obtained with both priors scale as $\sim \ell^3$ (straight dashed line).

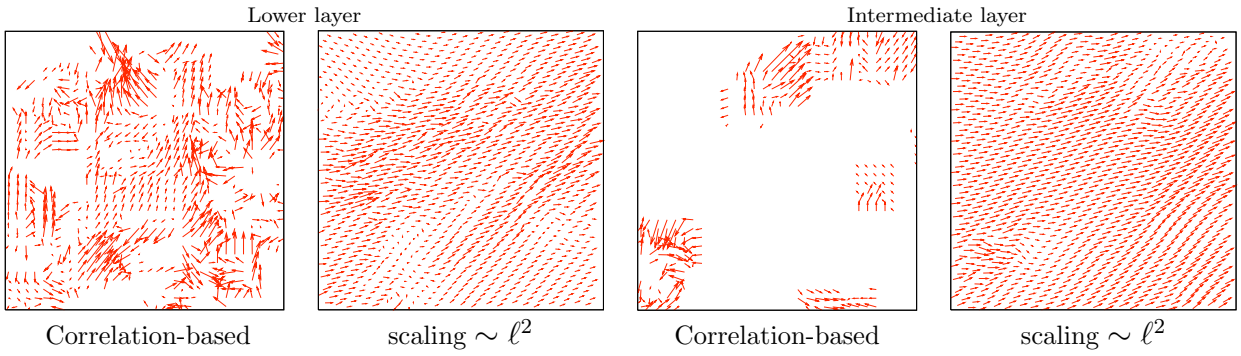

Fig. 13. Estimated horizontal wind fields compared to correlation results. Dense wind fields at low (left) and intermediate (right) altitude where obtained using the flat prior, *i.e.*, a scaling in ℓ^2. The correlation results were obtained using the operational PIV software *(Davis)* from *LaVision* (*www.lavision.de*) company.

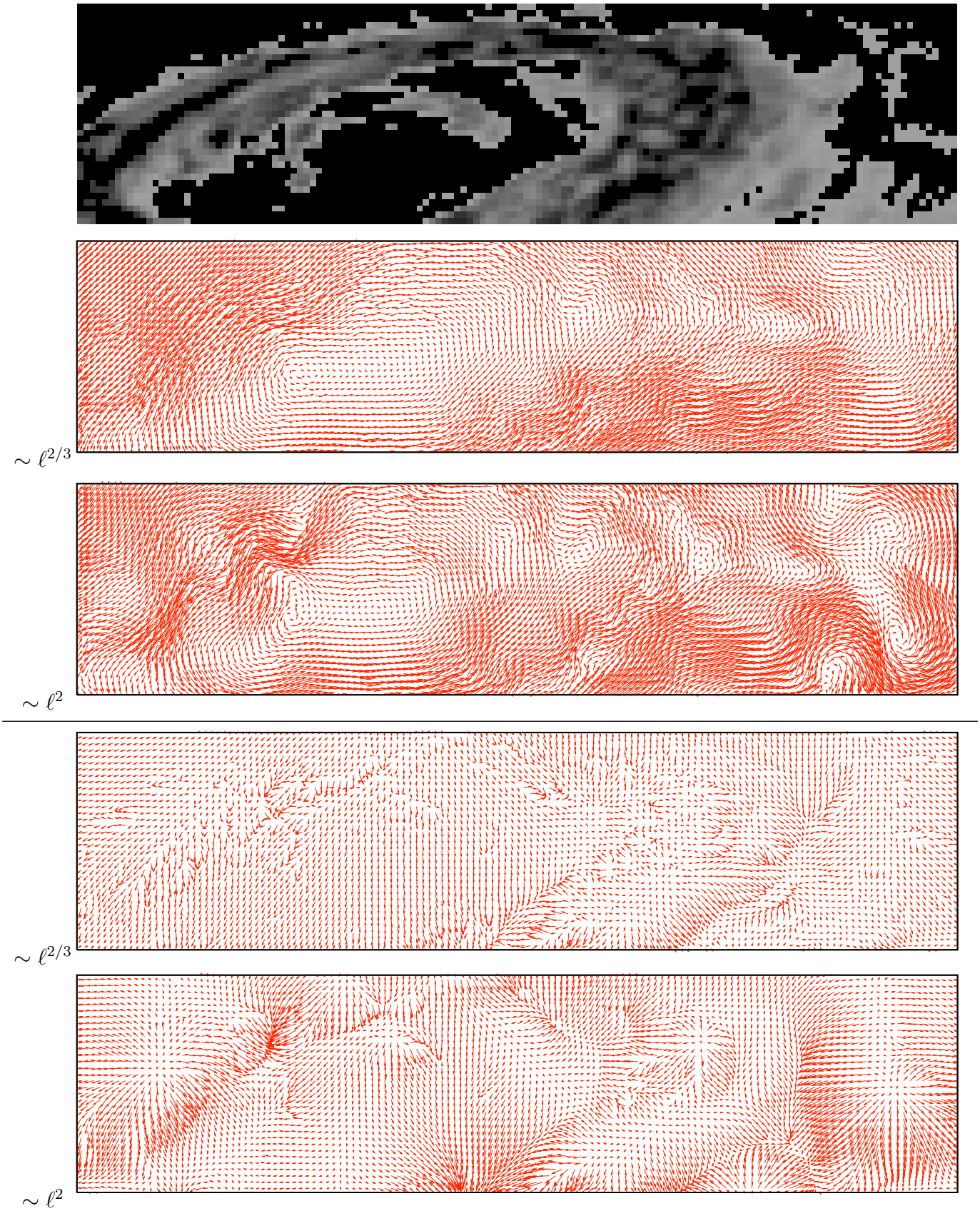

Fig. 14. Detail of horizontal winds at low altitude. From top to bottom: input image (zoom between the 20^o and 30^o meridians and the 50^o and 60^o parallels), solenoidal (2 following lines) and divergent part (2 last lines) of motion estimated with a scaling of $\ell^{2/3}$ (2-nd and 4-th line) or ℓ^2 (3-rd and 5-th line).

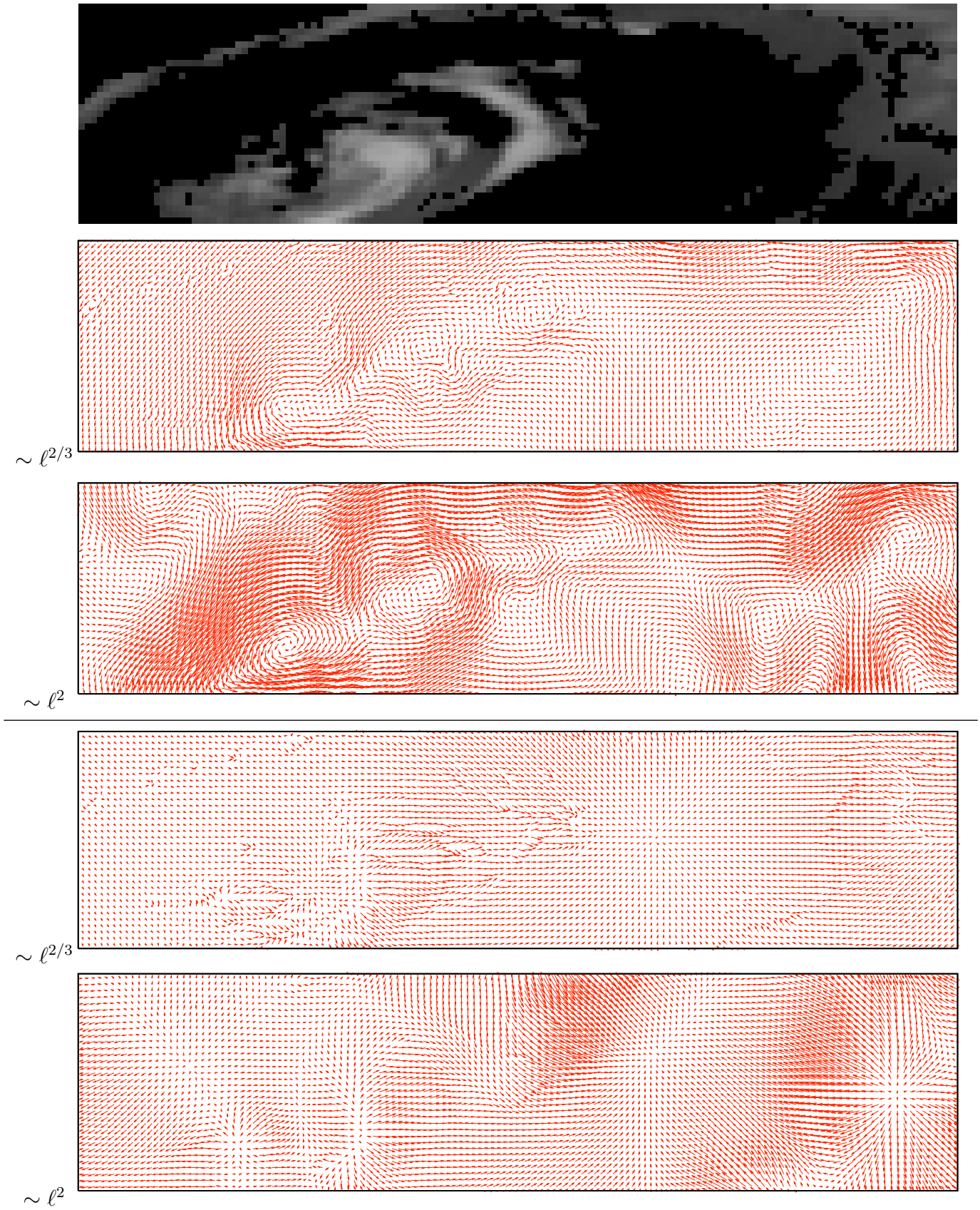

Fig. 15. Detail of horizontal winds at intermediate altitude. From top to bottom: input image (zoom between the 20^o and 30^o meridians and the 50^o and 60^o parallels), solenoidal (2 following lines) and divergent part (2 last lines) of motion estimated with a scaling of $\ell^{2/3}$ (2-nd and 4-th line) or ℓ^2 (3-rd and 5-th line).

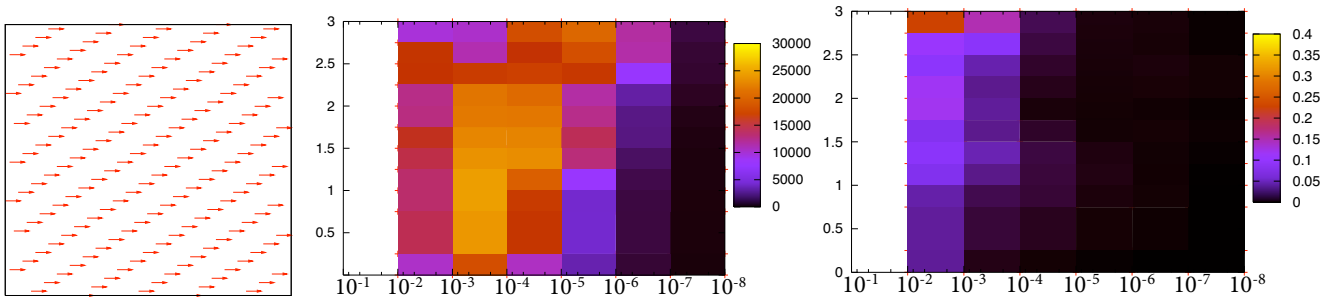

Fig. 16. Power-law evidence probability for laminar flows. True velocity field (left), plot of minus of the log of the power-law model probability (center) and the RMS motion reconstruction error (right) w.r.t. power-law exponent ζ (ordinate) and prefactor γ (abscissa). The line of minima which is selected by Bayesian inference corresponds to the true underlying power law ($\gamma = 0$ for any ζ).

LIST OF FIGURES

15 Detail of horizontal winds at intermediate altitude. From top to bottom: input image (zoom between the 20^o and 30^o meridians and the 50^o and 60^o parallels), solenoidal (2 following lines) and divergent part (2 last lines) of motion estimated with a scaling of $\ell^{2/3}$ (2-nd and 4-th line) or ℓ^2 (3-rd and 5-th line).

16 Power-law evidence probability for laminar flows. True velocity field (left), plot of minus of the log of the power-law model probability (center) and the RMS motion reconstruction error (right) w.r.t. power-law exponent ζ (ordinate) and prefactor γ (absciss). The line of minima which is selected by Bayesian inference corresponds to the true underlying power law ($\gamma = 0$ for any ζ).