Bayesian selection of scaling laws for motion modeling in images

Patrick Héas, Etienne Mémin, Dominique Heitz, P.D. Mininni

Bayesian selection of scaling laws for motion modeling in images

Patrick Héas
INRIA
Rennes, France
pheas@irisa.fr

Etienne Mémin
INRIA
Rennes, France
memin@irisa.fr

Dominique Heitz
CEMAGREF
Rennes, France
dominique.heitz@cemagref.fr

Pablo D. Mininni
UBA
Buenos Aires, Argentina
mininni@df.uba.ar

Abstract

Based on scaling laws describing the statistical structure of turbulent motion across scales, we propose a multiscale and non-parametric regularizer for optic-flow estimation. Regularization is achieved by constraining motion increments to behave through scales as the most likely self-similar process given some image data. In a first level of inference, the hard constrained minimization problem is optimally solved by taking advantage of lagrangian duality. It results in a collection of first-order regularizers acting at different scales. This estimation is non-parametric since the optimal regularization parameters at the different scales are obtained by solving the dual problem. In a second level of inference, the most likely self-similar model given the data is optimally selected by maximization of Bayesian evidence. The motion estimator accuracy is first evaluated on a synthetic image sequence of simulated bi-dimensional turbulence and then on a real meteorological image sequence. Results obtained with the proposed physical based approach exceeds the best state of the art results. Furthermore, selecting from images the most evident multiscale motion model enables the recovery of physical quantities, which are of major interest for turbulence characterization.

1. Introduction

The inverse modeling of fluid motion in images is an important issue in various application areas like meteorology, oceanography or turbulence studies and experimental fluid mechanics. Image analysis and data-assimilation methods are particularly important for studying multi-scale geophysical dynamical systems since they can characterize a large range of scales in comparison to sparse information contained in standard "in situ" data. For such turbulent flows, motion cannot be represented by a single spatial polynomial model. Instead, at each point a direct optic-flow observation model (relying on mass conservation, scalar transport, ...) links a motion vector to the image intensity function [7, 11, 17]. For these dense representations, regularization models are required to remove the motion ambiguities and achieve inversion. However, actual regularizers are all insufficient since they impose in a small scale spatial neighborhood a prior smoothness which describes improperly the regularity of turbulent flows. Moreover, they rely on the choice of a regularization weight, which may yield too smooth or on the contrary over-fitted solutions.

Nevertheless, it is possible to overcome such limitations. Firstly, important advances have been achieved in statistical modeling of turbulence since the precursor work of *Kolmogorov* in 1941 [8, 13, 16]. In particular, it has been shown from the Navier-Stokes equations for different kind of flows that turbulent motion regularity can be characterized using some universal scaling properties of the Probability Distribution Function (PDF) of motion increments. Secondly, Bayesian modeling provides a reliable framework for the design of non-parametrical methods and for selecting the regularization model given some data [18]. Therefore, this work 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 scaling law model describing the image intensity function based on a variable hierarchy linking 'image' to 'motion' to 'scaling laws'. The resulting regularization is built from the physics of fluid. It is multiscale, non-parametric (in the sense that it does not involve the tuning of any parameters) and allows the recovering of quantities such as energy flux across scales.

2. Related work

2.1. Optic-flow methods

Aperture problem The apparent motion $\mathbf{v} = (u, v)^T$, perceived through image intensity $I(\mathbf{s}, t)$ variations, respects the standard Optical Flow Constraint (OFC) observation model. Apparent motion and the real underlying velocity field are identical when considering rigid motion and stable lighting conditions. For fluids, this identity remains valid in the case of 2D incompressible flows. Based on mass conservation, the integrated continuity equation has been proposed in the literature for various 3D fluid flows visualized in a projected image plane in order to link the image intensity function I to a vertically averaged horizontal velocity field $\mathbf{v}$ [5, 7, 10, 11, 17]. However, the observation models are underconstrained, as they provide only one equation for two unknowns at each spatio-temporal location $(\mathbf{s}, t)$. This constitute the so-called aperture problem.

Standard regularization To deal with this problem, the most common assumption consists in enforcing locally spatial coherence. Global regularization schemes over the entire image domain Ω are convenient to model global coherence via local spatial dependencies. More precisely, esti-

mation is performed through the minimization of an energy composed of two terms: $f(I, \mathbf{v}) = f_d(I, \mathbf{v}) + \alpha f_r(\mathbf{v})$. The first one $f_d(I, \mathbf{v})$, called the data term, penalizes discrepancies from the observation models. For example, discretizing in time the OFC equation results in the data term:

$$f_d(I, \mathbf{v}) = \frac{1}{2} \int_\Omega (\tilde{I} - I + \mathbf{v} \cdot \nabla \tilde{I})^2 d\mathbf{s} \tag{1}$$

where $\tilde{I}$ denotes the image $I(t + \Delta t)$. The second component $f_r(\mathbf{v})$, called the regularization term, acts as a spatial *prior* enforcing the solution to follow some smoothness properties. In the previous expression, $\alpha > 0$ denotes a regularization parameter controlling the balance between the smoothness and the global adequacy to the observation model. In this framework, *Horn and Schunck* [12] proposed a first-order regularization of the two spatial components u and v of velocity field $\mathbf{v}$:

$$f_r(\mathbf{v}) = \frac{1}{2} \int_\Omega (||\nabla u||^2 + ||\nabla v||^2) d\mathbf{s} \tag{2}$$

However, motion gradient penalization is not adapted to fluid flows as it penalizes spatially non-homogeneous velocity fields. Second order regularizers on motion vorticity and divergence have been proposed to overcome those limitations [6]. Regularizations including additional constraints on vanishing divergence have also recently been proposed [21]. However, state of the art models only mimic qualitatively physical behavior but are not precisely related to the physics of fluid motion. They all depend on the tuning of α which can be chosen optimally [14, 18, 19].

Multi-resolution approach One major problem with observation models is the estimation of large displacements. Indeed, these equations are only valid if the solution remains in the region of linearity of the image intensity function. A standard approach for tackling non-linearity consists to apply successive linearizations around a current estimate and to warp a multi-resolution representation of the data accordingly. More explicitly, a large displacement field $\tilde{\mathbf{v}}$ is first estimated with the original data term at coarse resolution, where the linearity assumption is valid. Then, introducing the decomposition:

$$\mathbf{v} = \tilde{\mathbf{v}} + \mathbf{v}', \tag{3}$$

motion is refined through an incremental fields $\mathbf{v}'$ estimated using a linearized motion-compensated data term while going down the resolution levels of an image pyramid [3].

2.2. Turbulence statistical modeling

Since *Kolmogorov*'s works, turbulent motion increments are known to be structured as nearly scale invariant spatial processes. To review turbulence models, we need to introduce a central quantity: the longitudinal velocity increment function. It is defined in the direction of unitary vector $\boldsymbol{n}$ by:

$$\delta v_\parallel(\ell, \mathbf{s}, \boldsymbol{n}) = (\mathbf{v}(\mathbf{s} + \ell \boldsymbol{n}) - \mathbf{v}(\mathbf{s})) \cdot \boldsymbol{n} \tag{4}$$

where the scalar ℓ represents a spatial increment. As a classical hypothesis in turbulence studies, we assume homogeneity and isotropy, that is to say we consider that the statistical properties of the velocity field are invariant under translation of spatial location $\mathbf{s}$ and rotation of direction $\boldsymbol{n}$. In agreement with these assumptions, index to $\mathbf{s}$ and $\boldsymbol{n}$ can be dropped and moments of PDF of velocity increments $P_\ell(\delta v_\parallel)$, the so-called *structure functions*, can be approached by spatial integration:

$$\begin{aligned}
\mathrm{E}[\delta v_\parallel(\ell)^p] &= \int_\mathbb{R} \delta v_\parallel(\ell)^p p_\ell(\delta v_\parallel(\ell)) d\delta v_\parallel(\ell) \tag{5} \\
&\approx \frac{1}{2\pi|\Omega|} \int_\Omega \int_{[0,2\pi]} \left(\delta v_\parallel(\ell, \mathbf{s}, \boldsymbol{n}) \right)^p d\boldsymbol{n}\, d\mathbf{s}
\end{aligned}$$

where $|\Omega|$ denotes the spatial domain area.

For three-dimensional isotropic turbulent flows, *Kolmogorov* [8] demonstrated from the Navier-Stokes equations that the third order moment of PDF $p_\ell(\delta v_\parallel)$, namely the third order structure function, is linear w.r.t scale. That is $\mathrm{E}[\delta v_\parallel(\ell)^3] = -\frac{4}{5}\epsilon\ell$ in a so-called *inertial range*. The inertial range is defined as $[\eta, \ell_0]$, where η represents the largest molecular dissipative scale and where ℓ_0 is much smaller than the diameter L of the largest vortex. Within this range, an energy flux cascades from large to small scales. The energy dissipation rate ϵ is this energy flux passed accross scales which is evacuated at small scales by molecular viscosity. Analogously, for pure bi-dimensional turbulence with energy injection at scale ℓ_0, *Kraichnan* [13] demonstrated that there exist two different self-similar processes: a direct cascade where $\mathrm{E}[\delta v_\parallel(\ell)^3] = \frac{1}{8}\epsilon_\omega\ell^3$ within the inertial range $[\eta, \ell_0]$, and an inverse cascade where $\mathrm{E}[\delta v_\parallel(\ell)^3] = \frac{3}{2}\epsilon\ell$ within range $[\ell_0, L]$. An enstrophy flux ϵ_ω (L^2 norm of vorticity) 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. Concerning atmospheric turbulence, there are still open questions on the observed scaling laws. *Lindborg* [15, 16] proposed an answer to the question: "can atmospheric flow statistics be explained by two-dimensional turbulence?". He showed that the self-similar processes observed in aircraft data [20] at small scales and at large scales could be modeled by the superposition of a 3D direct energy cascade and a 2D direct enstrophy cascade so that:

$$\mathrm{E}[\delta v_\parallel(\ell)^3] = -\epsilon\ell + \frac{1}{8}\epsilon_\omega\ell^3. \tag{6}$$

Going further, *Kolmogorov* assumed that the longitudinal velocity increment functions were *strictly self-similar* processes. In this case the normalized PDF of motion increments is self-similar through scales as illustrated in Fig. 1. This implies that, in a given cascade, the p-th order structure function follows a power law:

$$\beta_p \ell^{\zeta_p} \sim \mathrm{E}[\delta v_\parallel(\ell)^p] = \mathrm{E}[\delta v_\parallel(\ell)^3]^{\frac{p}{3}} \sim \beta \ell^{\frac{p\zeta_3}{3}} \tag{7}$$

with universal exponents ζ_p depends on space dimension and where factor β_p is a function of the energy flux ϵ or the

enstrophy flux ϵ_ω. A corollary of the strictly self-similar assumption is that in the 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.* is regular). However, it is now well known that *Kolmogorov* assumption deviates from reality because of intermittency (coherent structures appearing in turbulence). As a consequence, only *non-strict self-similarity* can in reality be assumed for turbulent flows. It results that deviations on exponent values can be expected for $p \neq 3$. Finally, any 2D or 3D turbulent flow is regular in the dissipative range and using Taylor expansion we have $\mathrm{E}[\delta v_\parallel(\ell)^2] \sim \ell^2$ within $\ell \in [0, \eta]$.
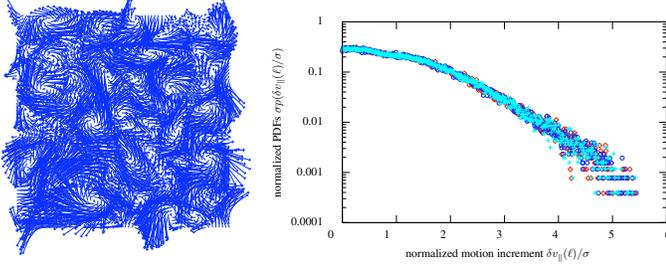

Figure 1. *2D turbulent motion and similarity through scales of its normalized PDFs* $\sigma_\ell p_\ell(\frac{\delta v_\parallel(\ell)}{\sigma_\ell})$ *for scale increment* $\ell = 1, 2, 4, 8$.

3. Self-similar regularization of optic-flow

Besides providing a closure for motion estimation, self-similar priors can yield several benefits:

- first, they constitute physical sound regularizers for fluid motion,
- second, they are prior multi-scale models which structures motion across scales,
- third, they lead to non-parametric estimation methods with no regularization parameter α.

3.1. Self-similar constraints

Let us first formalize self-similar constraints. Although it does not provide an exact prediction on scaling laws for non-strictly self-similar flows, we chose to use the second order structure function $\mathrm{E}[\delta v_\parallel(\ell)^2]$ because it constitutes a convenient quadratic constraint. Nevertherless, in section 4 we take into account deviations from the predicted law by selecting the most likely scaling law defined by parameters (β, ζ) given the image data. $\mathrm{E}[\delta v_\parallel(\ell)^2]$ is an expectation which can be obtained by spatial integration over the image domain as presented in Eq. 5. The integral on all directions $\mathbf{n}$ is approched by a sum on the 4 horizontal-vertical directions $\mathbf{n}_h = \{(-1,0), (1,0), (0,1), (0,-1))\}$ and on the 4 diagonal directions $\sqrt{2}\mathbf{n}_d = \{(1,1), (1,-1), (-1,-1), (-1,1)\}$ of the bi-dimensional plane :

$$\mathrm{E}[\delta v_\parallel(\ell)^2] = \frac{1}{8|\Omega - \{\overline{\Omega}(\ell)\}|} \int_{\Omega - \{\overline{\Omega}(\ell)\}} d\mathbf{s} \sum_{\mathbf{n}} \left(\delta v_\parallel(\ell, \mathbf{s}, \boldsymbol{n}) \right)^2 \tag{8}$$

In order to avoid using boundary conditions when calculating $\mathrm{E}[\delta v_\parallel(\ell)^2]$, we have excluded image borders $\overline{\Omega}(\ell)$ of width ℓ in the integral of Eq. 8. A self-similar constraint $g_\ell(\mathbf{v})$ is then defined at each scale ℓ as the difference between the 2-nd order structure function and a given power law. Thus, an estimated motion field should respect the constraint:

$$g_\ell(\mathbf{v}) = \frac{1}{2}(\mathrm{E}[\delta v_\parallel(\ell)^2] - \beta \ell^\zeta) = 0 \tag{9}$$

for given scaling exponent ζ and factor β.

3.2. Constrained motion estimation problem

Referring to section 2.1, the minimization of the ill-conditionned optic-flow estimation problem reads:

$$(\hat{\mathbf{v}}) = \arg\min_{\mathbf{v}} f_d(I, \mathbf{v}). \tag{10}$$

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

$$\begin{cases} \min_{\mathbf{v}} f_d(I, \mathbf{v}), \quad \mathbf{v} \in \mathbb{R}^n \\ \text{subject to the constraints:} \\ g_\ell(\mathbf{v}) = 0, \quad \forall \ell \in \mathbf{I} \end{cases} \tag{11}$$

where $\mathbf{I}$ is the scale range of the given power law.

3.3. Discrete problem formulation

Let us now express the constraint problem in its discrete form. The derivatives $\nabla_{\mathbf{v}} f_d(I, \mathbf{v})$ related to any motion-compensated data term (which is quadratic with respect to motion increments $\mathbf{v}'$) can be expressed in the matricial form $A_0 \mathbf{v}' - b_0$, when discretized on an image grid $\mathbf{S}$ of m points with a finite difference scheme. The two discretized components of $\mathbf{v}' \in \mathbb{R}^n$ now represent a field of $n = 2m$ variables supported by the grid $\mathbf{S}$, A_0 is $n \times n$ symmetric positive-definite, $\mathbf{b}_0 \in \mathbb{R}^n$ represents a vector of size n. The discrete data term can be rewritten as:

$$f_d(I, \mathbf{v}) = \frac{1}{2}\mathbf{v}'^T A_0 \mathbf{v}' - \mathbf{b}_0^T \mathbf{v}' + c_0., \tag{12}$$

where $c_0 \in \mathbb{R}$ denotes a scalar. For the self-similar constraints, 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. Thus, the constraints reads using variables $\mathbf{v}'$:

$$g_\ell(\mathbf{v}) = \frac{1}{2}\mathbf{v}'^T A_\ell \mathbf{v}' - \mathbf{b}_\ell^T \mathbf{v}' + c_\ell = 0, \quad \forall \ell \in \mathbf{I}, \tag{13}$$

where $c_\ell \in \mathbb{R}$ are scalars. The constraint motion estimation problem defined in Eq. 11 can thus be rewritten in its discrete form as:

$$(P) \begin{cases} \min_{\mathbf{v}} f_d(I, \mathbf{v}) = \frac{1}{2}\mathbf{v}'^T A_0 \mathbf{v}' - \mathbf{b}_0^T \mathbf{v}' + c_0 \\ \text{subject to:} \\ g_\ell(\mathbf{v}) = \frac{1}{2}\mathbf{v}'^T A_\ell \mathbf{v}' - \mathbf{b}_\ell^T \mathbf{v}' + c_\ell = 0, \quad \forall \ell \in \mathbf{I} \\ \mathbf{v} = \mathbf{v}' + \tilde{\mathbf{v}} \in \mathbb{R}^n. \end{cases} \tag{14}$$

3.4. Dual problem and optimal solution

To define optimality conditions, we now introduce the lagrangian function $L(\mathbf{v}, \boldsymbol{\lambda})$ associated to (P):

$$L(\mathbf{v}, \boldsymbol{\lambda}) = f_d(I, \mathbf{v}) + \sum_\ell \lambda_\ell g_\ell(\mathbf{v}), \quad \boldsymbol{\lambda} = \{\lambda_\ell\}. \tag{15}$$

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. *Saddle points* denoted by $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$ are defined as the solutions of the so-called *dual* problem:

$$(D) \begin{cases} L(\mathbf{v}^*, \boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}} w(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}}\{\min_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda})\} \\ \lambda_\ell \in \mathbb{R}, \ \forall \ell \in \mathbf{I} \end{cases}, \quad (16)$$

where $w(\boldsymbol{\lambda})$ denotes the *dual* function. As the functions f 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. Lagrangian multipliers λ_ℓ, which represent the regularization parameter at scale ℓ, 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. Nevertherless, there exists local optimal solutions.

3.5. Convex optimization

The minimum $\hat{\mathbf{v}}'$ of a locally convex lagrangian function at point $\boldsymbol{\lambda}$ can be obtained by solving the following Euler-Lagrange equations:

$$\nabla_{\mathbf{v}} L(\mathbf{v}, \boldsymbol{\lambda}) = \nabla_{\mathbf{v}} f_d(I, \mathbf{v}) + \sum_\ell \lambda_\ell \nabla_{\mathbf{v}} g_\ell(\mathbf{v}) = 0, \quad (17)$$

which reduce (using Eq. 12 and Eq. 13) to solve the large linear system:

$$(A_0 + A)\hat{\mathbf{v}}' = \mathbf{b}_0 + \mathbf{b}, \quad (18)$$

with $A = \left(\sum_\ell \lambda_\ell A_\ell\right)$ and $\mathbf{b} = \sum_\ell \lambda_\ell \mathbf{b}_\ell$. It can be shown that the linear system components A_ℓ and $\mathbf{b}_\ell$ 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 represent 2-nd order derivatives at different scales. Since we have no guarantee that matrix $A_0 + A$ is positive-definite depending on the sign of lagrangian multipliers λ_ℓ, the resolution of the large system of Eq. 18 is efficiently achieved using a Conjugate Gradient Squared (CGS) method with an incomplete LU preconditionner. The dual function is then given by:

$$w(\boldsymbol{\lambda}) = \frac{1}{2}\hat{\mathbf{v}}'^T\left(A_0 + A\right)\hat{\mathbf{v}}' - \left(\mathbf{b}_0 + \mathbf{b}\right)^T \hat{\mathbf{v}}' + c_0 + c, \quad (19)$$

where the constant $c = \sum_\ell \lambda_\ell c_\ell$. The dual function is by definition concave and possesses so-called *sub-gradients* equal to $g_\ell(\hat{\mathbf{v}}' + \tilde{\mathbf{v}})$. We employ a classical gradient method to find $\boldsymbol{\lambda}^*$ which maximizes the dual function and thus obtain the solution $\mathbf{v}^*$. Finally, the constraint 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. To cope with non-linearity of the data term, incremental motion fields are estimated using motion compensated images and a multiresolution approach[1]. An important remark is that once

---

[1]Note that the scaling law factor β must be scaled by the multiplicative factor $(2^j)^{-\zeta+2}$ at each scale j of the multi-resolution pyramid.

---

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 multiresolution Uzawa algorithm is presented below.

- Iterate until resolution j reaches the finest level:
  - Compensate image $I_{(j)}$ with coarse motion estimate $\tilde{\mathbf{v}}_{(j)}$
  - Iterate until convergence from initial point $(\tilde{\mathbf{v}}_{(j)}, \boldsymbol{\lambda}^0_{(j)})$ with $\boldsymbol{\lambda}^0_{(j)} > 0$:
    * At iteration k, find increment $\hat{\mathbf{v}}'_{(j)}$ by solving Eq. 18
    * Define $\boldsymbol{\lambda}^{k+1}_{(j)}$ by a gradient ascent step:
      $$\forall \ell \in \mathbf{I}, \ \lambda^{k+1}_{\ell,(j)} = \lambda^k_{\ell,(j)} + \rho^k g_\ell(\hat{\mathbf{v}}'_{(j)} + \tilde{\mathbf{v}}_{(j)})$$
  - $(\mathbf{v}^*_{(j)}, \boldsymbol{\lambda}^*_{(j)}) = (\hat{\mathbf{v}}'_{(j)} + \tilde{\mathbf{v}}_{(j)}, \boldsymbol{\lambda}^k_{(j)})$
  - Define $\tilde{\mathbf{v}}_{(j-1)}$ by projection of $\mathbf{v}^*_{(j)}$ on level $j-1$
  - $j = j - 1$

ρ^k denotes the displacement step at iteration k. The latter parameter is relaxed at each iteration.

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

4. Selection of a multiscale prior model

In the previous section, we have proposed to model motion in images conditioned by a prior scaling law model (defined by power law factor β and exponent ζ or slope in log-log coordinates). We now want to select the most appropriate scaling law model for motion estimation given the image data. Model selection will yield several advantages:

- first, prior model inference will result in a non-parametric method,
- second, the modeling will deal with uncertainties in turbulence theoretical predictions,
- third, inference of $(\hat{\beta}, \hat{\zeta})$ will reveal important physical quantities in turbulence such as power law exponents (linked to motion regularity), flux across scales, or the energy and enstrophy dissipation rates.

After reformulating the constrained motion estimation problem in a probabilistic framework, we show how the multiscale prior model likelihood probability given the image data can be evaluated.

4.1. Bayesian hierarchical modeling

Bayes' rule provides a nice framework to evaluate this model likelihood probability, the so called *evidence*. Indeed, a probabilistic reformulation of the global motion estimation problem yields a 3-level hierarchical model linking image, motion and scaling laws:

$$I \to \mathbf{v} \to \beta, \zeta \quad (20)$$

Note that regularization weights $\boldsymbol{\lambda}^*(\zeta, \beta)$ do not appear in the variable hierarchy as they are deterministically given for fixed (ζ, β). Applying Bayes' rule, we obtain two levels of inference in this hierarchy [18]:

- *Scaling model fitting.* We assume some scaling model parameters (ζ, β) i.e regularization weights $\boldsymbol{\lambda}^*(\zeta, \beta)$ (lagrangian multipliers) given by the dual formalism. Solving the primal problem in the previous section is equivalent to infer a velocity field $\mathbf{v}^*$ according to a *Maximum A Posteriori* (MAP) criterion. The *posterior* PDF of this first level of inference is given by Bayes' relation:

$$p(\mathbf{v}|I, \zeta, \beta) = \frac{p(I|\mathbf{v}, \zeta, \beta)p(\mathbf{v}|\zeta, \beta)}{p(I|\zeta, \beta)} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$
$$\propto p(I|\mathbf{v}, \zeta, \beta)p(\mathbf{v}|\zeta, \beta) \quad (21)$$

and is a Gibbs PDF which reads $p(\mathbf{v}|I, \zeta, \beta) =$

$$\frac{\exp\{-\frac{1}{2}\mathbf{v}'^T(A_0 + A(\zeta, \beta))\mathbf{v}' + (\mathbf{b}_0 + \mathbf{b}(\zeta, \beta))^T \mathbf{v}' - c_0 - c(\zeta, \beta)\}}{Z_L(\zeta, \beta)} \quad (22)$$

where $Z_L(\zeta, \beta)$ denotes a normalization constant.

- *Scaling model selection.* A second level of inference can be performed on the scaling law model parameters (ζ, β) using Bayes' relation:

$$p(\zeta, \beta|I) = \frac{p(I|\zeta, \beta)p(\zeta, \beta)}{p(I)} \propto p(I|\zeta, \beta)p(\zeta, \beta) \quad (23)$$

For a flat prior $p(\zeta, \beta)$, the MAP of Eq. 23 w.r.t self-similar model parameters (ζ, β) is simply the *Maximum Likelihood* (ML) estimate or in other words the maximum of the *evidence* $p(I|\zeta, \beta)$. The *evidence* can be obtained by marginalization w.r.t. the velocity field:

$$p(I|\zeta, \beta) = \int_{\mathbb{R}^n} p(I|\mathbf{v}, \zeta, \beta)p(\mathbf{v}|\zeta, \beta)d\mathbf{v} \quad (24)$$

Direct calculation of this integral is prohibitive since it is in huge dimensions. However, let us recall that the *evidence* is the normalization constant (w.r.t. velocity field $\mathbf{v}$) which has been ignored in the first level of inference (Eq. 21). Therefore, we can rewrite the *evidence* as a normalization constant ratio:

$$p(I|\zeta, \beta) = \frac{\text{likelihood} \times \text{prior}}{\text{posterior}} = \frac{Z_L(\zeta, \beta)}{Z_{f_d}Z_{g_\ell}(\zeta, \beta)} \quad (25)$$

where Z_{f_d} and Z_{g_ℓ} denote the normalization constants of the Gibbs *likelihood* and *prior* PDF.

4.2. Scaling model selection by evidence

The scaling law model evidence can now be evaluated as a normalization constant ratio. First, the *likelihood* PDF $p(I|\mathbf{v}, \zeta, \beta)$ related to a quadratic optic-flow data term f_d is a normalized m dimensional Gaussian with uncorrelated components. Thus its normalization constant reads:

$$Z_{f_d} = \int_{\mathbb{R}^m} \exp\{-f_d(I, \mathbf{v})\}dI = (2\pi)^{m/2}, \quad (26)$$

where m denotes the number of pixels. Therefore Z_{f_d} is a constant w.r.t. (ζ, β) which can be ignored. Then, the *posterior* PDF of Eq. 22 may be a multivariate Gaussian or not depending on the positivity of lagrangian multipliers λ_ℓ. For a multivariate Gaussian, the normalization constant integral Z_L can be exactly calculated using the Laplace's relation:

$$Z_L(\zeta, \beta) = \int_{\mathbb{R}^n} \exp\{-L(\mathbf{v}, \boldsymbol{\lambda}^*(\zeta, \beta))\}d\mathbf{v} \quad (27)$$
$$= \exp\{-L(\mathbf{v}^*(\zeta, \beta), \boldsymbol{\lambda}^*(\zeta, \beta))\} 2\pi^{\frac{n}{2}} \det(A_0 + A(\zeta, \beta))^{-\frac{1}{2}},$$

where we recall that $\mathbf{v}^*$ is the MAP estimate, $\boldsymbol{\lambda}^*$ is the associated set of lagrangian multipliers and where $n = 2m$ denotes the number of unknown velocity variables. For other distributions, Laplace's relation stills constitutes a good approximation [9]. The determinant of such large and sparse matrices can be efficiently approximated via an incomplete LU decomposition. Finally, as the *prior* PDF normalization constant can not be evaluated analytically such as in [18, 19], we expand it around its maximum using Laplace's relation. The prior reads:

$$p(\mathbf{v}|\zeta, \beta) = \frac{\exp\{-\frac{1}{2}\mathbf{v}'^T A(\zeta, \beta)\mathbf{v}' + \mathbf{b}^T(\zeta, \beta)\mathbf{v}' - c(\zeta, \beta)\}}{Z_{g_\ell}(\zeta, \beta)}. \quad (28)$$

This self-similar *prior* 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 (only for evaluating the evidence). Note that the precise value on the boundaries need not to be specified since it modifies vector $\mathbf{b}$ but does not have impact on the hessian matrix A (only the size of A is reduced with dirichlet boundary conditions). Considering these boundaries, we get a slightly changed hessian matrix A which is of full rank. As previously, the normalization constant can be calculated using Laplace's approximation:

$$Z_{g_\ell}(\zeta, \beta) = \int_{\mathbb{R}^n} \exp\{-\sum_\ell \lambda_\ell^*(\zeta, \beta)g_\ell(\mathbf{v})\}d\mathbf{v} \quad (29)$$
$$= \underbrace{\max_{\mathbf{v}}\left(\exp\{-\sum_\ell \lambda_\ell^*(\zeta, \beta)g_\ell(\mathbf{v})\}\right)}_{=1} 2\pi^{\frac{n}{2}} \det A(\zeta, \beta)^{-\frac{1}{2}}.$$

As the set of admissible solution for $\mathbf{v}$ for the self-similar constraint is not empty, the exponential term in Eq. 27 has a maximum value equal to 1. Finally, using Eq. 25, Eq. 27 and Eq. 29, the log evidence of the scaling model reads:

$$\log p(I|\zeta, \beta) \propto \underbrace{-f_d(\mathbf{v}^*, I)}_{\text{data term}} \underbrace{-\frac{1}{2}(\log \frac{\det(A_0 + A)}{\det(A)})}_{\text{log Occam factor}}, \quad (30)$$

where for simplification we have dropped the dependance to parameters (ζ, β). The last terms, known as *Occam factor*, penalizes the model complexity. It is the ratio of the posterior accessible volume on the prior accessible volume in $\mathbf{v}$ (a variance ratio in 1D). Note that term $\sum_\ell \lambda_\ell^* g_\ell(\mathbf{v}^*)$ does not appear in Eq. 30 as the constraints vanish at the saddle point.

5. Experiments

Simulated 2D turbulence. To evaluate the performance of the self-similar regularization, a synthetic particle image sequence was generated based on forced two-dimensional turbulence obtained by direct numerical simulation (DNS) of Navier-Stokes equations with a Reynold number of 3000,
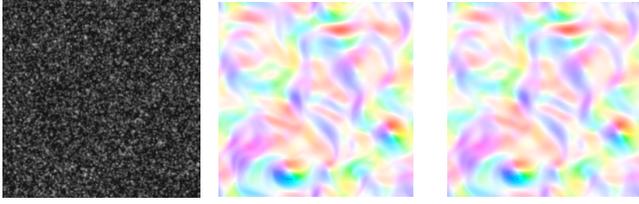
Figure 2. **2D turbulence.** *Particle image (left). Scalar visualization [1] of true (middle) and estimated (right) velocity fields*
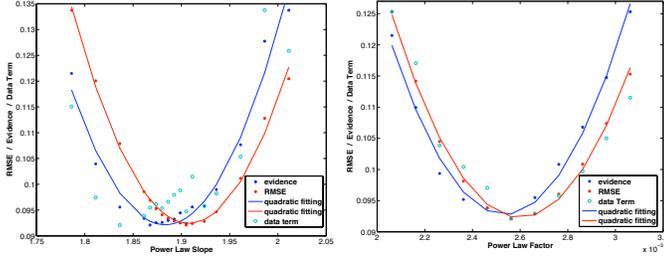

Figure 3. **Scaling model selection by evidence maximization.** *Behavior of data term and minus the evidence logarithm w.r.t slope ζ (left) and factor β (right) in comparison to the RMSE. For comparison, evidence and data term ranges of value have been rescaled on the RMSE.*

and a particle image generator [4]. Fig. 2 presents one of the particle images of 256×256 pixels and a scalar representation of the true underlying velocity field displayed in Fig. 1. In such visualizations color and intensity code vector orientations and magnitudes [1]. The motion estimate minimizing the quadratic OFC based data-term (Eq. 1) under self-similarity constraints is displayed for comparison in Fig. 2.

Evaluation of model selection. The evidence of the self-similar model in the scale range of $\mathbf{I} = [1, 10]$ pixels (corresponding either to dissipation or the enstrophy cascade) is evaluated by sampling ζ respectively around the theoretical value of 2 for the power law exponent, and by sampling factor β around a Least Square (LS) estimate given by any rough estimator (*e.g.* [12]). Fig. 3 shows the behavior of the log-evidence and the Root Mean Square Error (RMSE) w.r.t the scaling law slope and factor. Note that the evidence reaches its maximum in $\hat{\zeta} = 1.90$ and $\hat{\beta} = 0.0026$ which corresponds to the RMSE minimum but not to the exponent fitting ground truth ($\zeta = 1.94$ in the LS sense). As RMSE and minus the logarithm of the evidence seem to define parabolas around their minimum, we fit them with quadratic functions for visualization convenience. While parabolas share roughly the same minimum in β, one can notice a slight shift of 0.018 between the two parabolas in ζ. This shift has however a minor incidence in the error increase. The power law minimizing the evidence probability is plotted in Fig. 4. It is identical in the scale range $\mathbf{I}$ to the true second order structure function $\mathrm{E}[\delta v_{\parallel}(\ell)^2]$. Bayesian evidence constitutes a theoretical sound and therefore reliable criteria for selecting the scaling model. In particular, as shown in Fig. 3 it is more efficient than the data-term error.
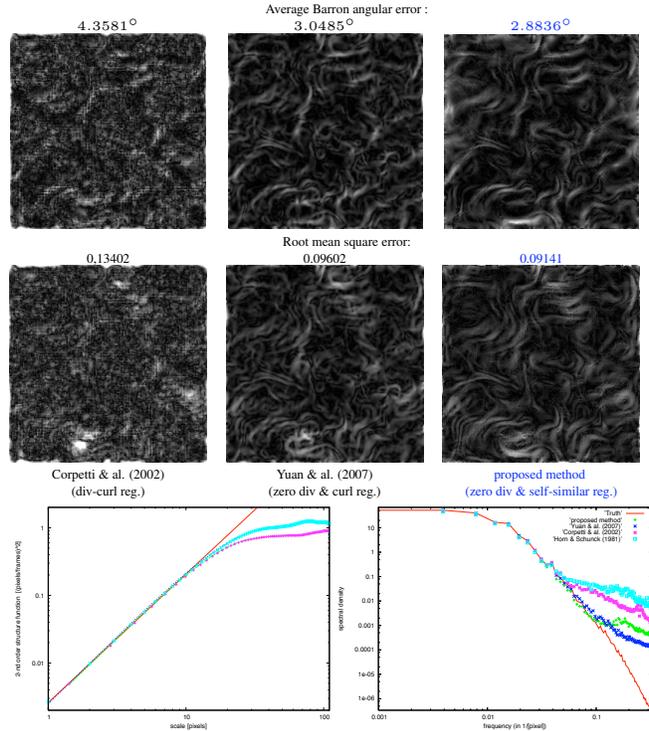
Evaluation of motion estimation. A comparison of end

Figure 4. **Motion estimation accuracy & 2-nd order statistics. Images:** *Barron's angular error (1-st line) and end point errors (2-nd line) comparisons with state of the art. RMSE and average Barron angular error are displayed above the figures.* **Log-log plots:** *power law maximizing evidence (continuous red line). True (dash line) and estimated (crosses) 2-nd order structure functions in horizontal-vertical (in blue and turquoise) and diagonal (in pink and green) directions (left). Energy spectra $E(k)$ of first order (in turquoise), div-curl (in blue or pink) and proposed (in green) regularizers compared to the true (in red) spectrum.(right).*

point errors (L^1 norm of velocity vector difference) and Barron's angular errors [2] with state of the art estimators is presented in Fig. 4. Based on these criteria, the proposed method outperforms in average most accurate operational correlation-based techniques[2], first order[3] [12] and div-curl [6, 21] regularizers. Let us remark that for this incompressible bi-dimensional experiment, we have subtracted the divergent component of the estimated flow using Helmoltz decomposition in order to make results comparable to [21]. The error maps comparison displayed in the same figure shows that the proposed regularization enhances in particular the estimation of small scale displacements compared to other approaches. At larger scales, the method also outperforms other approaches. Nevertheless, medium structures very similar to those obtained in [21] can be observed. The 2-nd order structure function log plots show that, especially in the scale interval where constraints have been applied, the estimation fits perfectly the ground truth. The power law maximizing the model evidence fits in this interval also very well the ground truth.

---

[2]Operational software from *LaVision* compagny (*www.lavision.de*) gave a RMSE of 0.1313

[3]*Horn&Schunk* estimator gave a RMSE of 0.1385

One can observe that the turbulent flow possess isotropic statistics at small scales since structure functions calculated in horizontal-vertical and in diagonal directions are nearly identical. We also compute the average L_2 norm of the fourier transform over each line of the horizontal velocity component as it produces a 1D energy spectrum $E(k)$ representation which enables to analyze motion at the different scales. The log plot of the ground truth energy spectrum exhibits a slope close to -5 which is much steeper than the -3 slope expected according to *Kraichnan*'s theory, but which is not unusual for DNS of bi-dimensional turbulence. Unfortunately, for slopes outside the interval $]-1,-3[$, no correspondence can be made with the slope of the related 2-nd order structure function [8]. This explains that, although the power law of the 2-nd order structure function has been accurately estimated, the tale of the estimated spectrum is not constraint to fit to the -5 slope. Nevertheless, the proposed regularization restitutes the spectrum at higher frequencies.
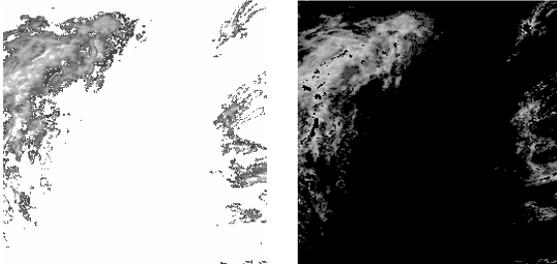

Figure 5. *Meteorological turbulence : sparse input images of layer at intermediate altitude for consecutive times. White or black regions correspond to missing observations.*

Atmospheric flow. The multi-scale regularizer has then been assessed on real data. A benchmark constituted with METEOSAT Second Generation meteorological image sequences acquired above the north Atlantic Ocean at a rate of an image every 15 min has been used. The image spatial resolution is $3 \times 3 \, \text{km}^2$ at the center of the whole Earth image disk. According to the physical-based methodology proposed in [11], sparse image of 256×256 pixels related to a layer at intermediate altitude have been derived. A robust data term using semi-quadratic M-estimator and relying on a layer mass conservation model has been used to relate the image intensity function to a vertically averaged horizontal wind field. The two input images are displayed in figure 5.

Energy flux ϵ selection by evidence maximization. We assume that the exponent $\zeta = 2/3$ 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 only visible for the 2-nd order structure function up to separation of about 10 kilometers [15]). We thus only need to infer parameter β by evidence maximization. Fig. 6 shows that the evidence maximum is around $\hat{\beta} = 0.0024$. This plot also illustrates the shared contribution of the *Occam*
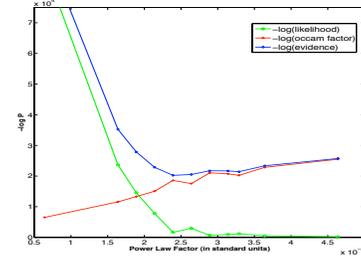

Figure 6. **Scaling model selection by evidence maximization.** *Minus log of evidence (blue), likelihood (green) and Occam factor (red) v.s. power law factor β*
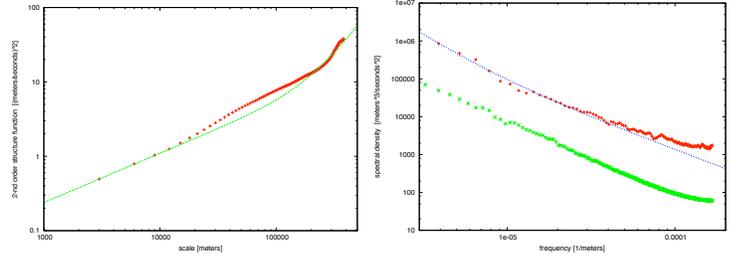

Figure 7. **2-nd order statistics. Left :** *2-nd order structure function (in red) and its parametrical model (in green).* **Right :** *Parametrical model (blue) and 1D energy spectrum estimated by our approach (red) compared to [11] (green).*

factor and the data term in the evidence. Note that alone, the data term is an insufficient criteria for model selection as it vanishes almost completely for large values of β. The model proposed in [15] provides an expression for the 2-nd order structure function and the energy spectrum:

$$\begin{cases} \mathrm{E}[\delta v_\parallel(\ell)^2] = C_2 \epsilon^{\frac{2}{3}} \ell^{\frac{2}{3}} + b\,\ell^2 - c\,\ell^2 \log \ell \\ E(k) = C_2 \epsilon^{\frac{2}{3}} k^{-\frac{5}{3}} + c\,k^{-3}/2 \end{cases} \quad (31)$$

In Eq. 31, the energy flux (or the energy dissipation rate) can be related in the scale range $\mathbf{I}$ to the power law factor by $\beta = C_2 \epsilon^{\frac{2}{3}}$. In the previous equations, b and c denote parameters and $C_2 \simeq 6$ is a Kolmogorov constant. The evidence maximum provides the most likely energy flux:

$$\hat{\epsilon} \simeq 0.79 \times 10^{-5} m^2 s^{-3}.$$

This estimate has the same order of magnitude as previous reported results based on aircraft data [15, 16]. Thus we find the agreement very good as the measure is only based on image data. Parameters b and c can be estimated in a LS sense fitting the parametrical models of Eq. 31 to the estimated energy spectrum and to the 2-nd order structure functions for scales $\ell > 50$ km. As shown in figure 7, those second order statistics are consistent with their associate parametrical models. For comparison, the energy spectrum obtained by [11] is displayed on the same plot. It can be noticed here that there is a clear under estimation of the energy flux ϵ (magnitude about 10 times smaller in [11]).

Evaluation of atmospheric wind estimation. In Fig. 8, one can visualize estimated wind fields for different factor β (*i.e.* energy flux ϵ) superimposed on sparse image observations. The wind fields are all visually consistent. It clearly appears that the smoothness of the wind fields decrease with

$\epsilon = 0.1e-5$

$\hat{\epsilon} = 0.8e-5$

$\epsilon = 2.1e-5$
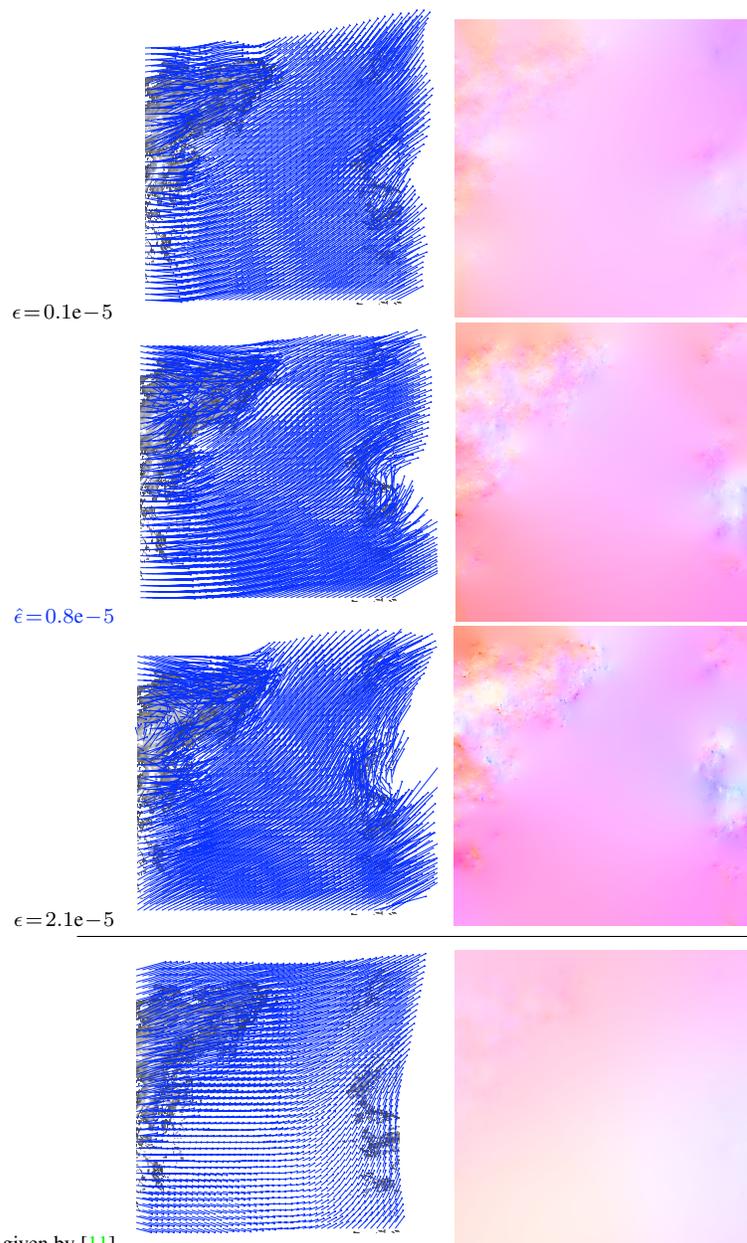
given by [11]

Figure 8. ***Winds & energy flux.*** *Vector (left) and scalar (right) motion representations for increasing energy flux (including the most likely flux $\hat{\epsilon}$ selected by Bayesian evidence) superimposed on the sparse image data compared to smooth results of [11].*

factor β. In agreement with the previous spectral comparison, the most likely wind field selected by evidence maximization is much more structured than the estimate provided in [11], which is displayed in Fig. 8.

6. Conclusions

We have presented a physical-based, multi-scale and non-parametric (no tuning of any parameters) method for fluid motion modeling in images. It relies on a Bayesian hierarchical model which simultaneously provides optimal solutions for two problems: motion estimation and regularization model selection. Regularization models rise from *Kolmogorov*'s theoretical work on turbulent flow self-

similarity. Experiments on a synthetic sequence shows that the method is more accurate than the best motion estimators. Moreover, the method constitutes a valuable tool for physical characterization of turbulence from images. In particular, consistent flux across scales in atmospheric turbulence are recovered from a meteorological image sequence.

References

[1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *Int. Conf. on Comp. Vis., ICCV 2007*, 2007. 6

[2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *Int. J. Computer Vision*, 12(1):43–77, 1994. 6

[3] J. Bergen, P. Burt, R. Hingorani, and S. Peleg. A 3-frame algorithm for estimating two-component image motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(9):886–895, Sept. 1992. 2

[4] J. Carlier and B. Wieneke. Report 1 on production and diffusion of fluid mechanics images and data. *Fluid project deliverable 1.2. http://www.fluid.irisa.fr*, 2005. 6

[5] T. Corpetti, P. Heas, E. Memin, and N. Papadakis. Pressure image assimilation for atmospheric motion estimation. *Tellus A*, 61(3):160–178, 2008. 1

[6] T. Corpetti, E. Mémin, and P. Pérez. Dense estimation of fluid flows. *Pattern Anal Mach Intel*, 24(3):365–380, 2002. 2, 6

[7] J. Fitzpatrick. The existence of geometrical density-image transformations corresponding to object motion. *Comput. Vision, Graphics, Image Proc.*, 44(2):155–174, Nov. 1988. 1

[8] U. Frisch. *Turbulence : the legacy of A.N. Kolmogorov*. Cambridge university press, 1995. 1, 2, 7

[9] S. F. Gull. Developments in maximum-entropy data analysis. In *Maximum Entropy and Bayesian Methods*, pages 53–71. Kluwer Academic, 1989. 5

[10] P. Heas and E. Memin. Three-dimensional motion estimation of atmospheric layers from image sequences. *IEEE trans. on Geo. and Rem. Sensing*, 46(8):2385–2396, 2008. 1

[11] P. Heas, E. Memin, N. Papadakis, and A. Szantai. Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *IEEE trans. on Geo. and Rem. Sensing*, 45(12):4087–4104, 2007. 1, 7, 8

[12] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981. 2, 6

[13] R. Kraichnan. Inertial ranges in two-dimensional turbulence. *Phys. Fluids*, 10:1417–1423, 1967. 1, 2

[14] K. Krajsek and R. Mester. A maximum likelihood estimator for choosing the regularization parameters in global optic flow methods. In *ICIP*, pages 1081–1084, 2006. 2

[15] E. Lindborg. Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence. *J. Fluid Mech.*, 388:259–288, 1999. 2, 7

[16] E. Lindborg and J. Cho. Horizontal velocity structure functions in the upper troposphere and lower stratosphere 2. theoretical considerations. *J. Geophysical Research.*, 106:233–241, 2001. 1, 2, 7

[17] T. Liu and L. Shen. Fluid flow and optical flow. *Journal of Fluid Mechanics*, 614:253, Oct. 2008. 1

[18] D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, 1992. 1, 2, 4, 5

[19] R. Molina, A. K. Katsaggelos, and J. Mateos. Bayesian and regularization methods for hyperparameter estimation in image restoration. *IEEE Trans. Image Processing*, 8:231–246, 1999. 2, 5

[20] G. Nastrom, W. Jasperson, and K. Gage. Kinetic energy spectrum of large and mesoscale atmospheric processes. *Nature*, 310:36–38, 1984. 2

[21] J. Yuan, C. Schnoerr, and E. Memin. Discrete orthogonal decomposition and variational fluid flow estimation. *Journ. of Math. Imaging & Vison*, 28:67–80, 2007. 2, 6