INFERENCE ON GIBBS OPTIC-FLOW PRIORS : APPLICATION TO ATMOSPHERIC TURBULENCE CHARACTERIZATION

P. Héas, E. Mémin

INRIA Bretagne Atlantique Research Center, Rennes, France

ABSTRACT

In this paper, bayesian inference is used to select the most evident Gibbs prior model for motion estimation given some image sequence. The proposed method supplements the maximum *a posteriori* motion estimation scheme proposed in (1). Indeed, in this recent work, the authors have introduced a family of multiscale spatial priors in order to cure the ill-posed inverse motion estimation problem. We propose here a second level of inference where the most likely prior model is optimally chosen given the data by maximization of bayesian evidence. Model selection and motion estimation are assessed on Meteorological Second Generation (MSG) image sequences. Selecting from images the most evident multiscale model enables the recovery of physical quantities which are of major interest for atmospheric turbulence characterization.

Index Terms— Bayesian evidence, Gibbs random fields, optic-flow, self-similar process, atmospheric turbulence

1. INTRODUCTION

The inverse modeling of fluid motion in images is an important issue in meteorology and turbulence studies. Image analysis and data-assimilation methods are particularly important for studying multiscale 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, etc) links a motion vector to the image intensity function (2; 3). For these dense representations, regularization models are required to remove the motion ambiguities and achieve inversion.

1.1. Optic-flow standard soft regularization

To deal with this underconstrained optic-flow estimation problem, the most common setting consists in enforcing some spatial coherence of motion field $\mathbf{v} = (u, v)^T$. This coherence is imposed either globally through a regularization functional defined over the whole image I domain Ω or locally relying on polynomial models defined on local bounded supports. Global regularization schemes are convenient to model global coherence via local spatial dependencies. More precisely, the estimation is performed through the minimization of an energy functional composed of two terms:

$$f(I, \mathbf{v}) = f_d(I, \mathbf{v}) + \alpha f_r(\mathbf{v}). \tag{1}$$

The first one $f_d(I, \mathbf{v})$, the data term (or minus the log *likelihood*), penalizes discrepancies from the observation models. For example, the integrated continuity equation observed in the image plane results in the data term (3):

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

where $\tilde{I}$ denotes the image $I(t + \Delta t, \mathbf{s} + \mathbf{v}\Delta t)$. The second component $f_r(\mathbf{v})$, 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, (4) proposed a first-order regularization of the two spatial components u and v of velocity field $\mathbf{v}$:

$$f_r(\mathbf{v}) = \frac{1}{2}\int_\Omega (||\nabla u||^2 + ||\nabla v||^2) d\mathbf{s} \tag{3}$$

However, motion gradient penalization is not adapted to fluid flows as it comes to penalize in an homogeneous way the curl and the divergence of the solution. Second order regularizers on motion vorticity and divergence have been proposed to overcome such limitations (2). All these approaches depend however on the tuning of α and only mimic qualitatively physical behavior but are not precisely related to the physics.

1.2. Hard-constraints & non-parametric multiscale prior

The authors in (1) removed the weighting prior parameter α by using physical-based multiscale self-similar models (power laws) describing the motion field regularity and a hard-constrained penalization scheme. More precisely, let us first defined a discrete data term (quadratic with respect to motion $\mathbf{v}$), expressed it in a matricial form after discretization on an image grid $\mathbf{S}$ of m points with a finite difference scheme. The two discretized components of $\mathbf{v} \in \mathbb{R}^n$ thus 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:

$$f_d(I, \mathbf{v}) = \frac{1}{2}\mathbf{v}^T A_0 \mathbf{v} - \mathbf{b}_0^T \mathbf{v} + c_0., \tag{4}$$

where $c_0 \in \mathbb{R}$ denotes a scalar. Then, the authors employed self-similar constraints $g_\ell(\mathbf{v})$ defined at each scale $\ell \in \mathbf{I}$ as the difference between a given power law and the 2-nd order moment of the motion increment Probability Distribution Function (PDF):

$$\mathrm{E}[\delta v(\ell)^2] \simeq K \int_\Omega \int_{\boldsymbol{\theta}} d\boldsymbol{\theta} ds (\mathbf{v}(\mathbf{s}+\ell\boldsymbol{\theta}) - \mathbf{v}(\mathbf{s}))^2 \cdot \boldsymbol{\theta}, \quad (5)$$

where the scalar ℓ represents a given scale, $\boldsymbol{\theta}$ denotes a unitary vector direction and K a normalization constant. They thus forced the estimated motion field to respects the statistical constraint at scale ℓ: $g_\ell(\mathbf{v}) = \frac{1}{2}(\mathrm{E}[\delta v(\ell)^2] - \beta \ell^\zeta) = 0$, where (β, ζ) are parameters of the power law defining the prior model. They then showed that those quadratic constraints could be written after discretization as:

$$g_\ell(\mathbf{v}, \beta, \zeta) = \frac{1}{2}\mathbf{v}^T A_\ell \mathbf{v} - \mathbf{b}_\ell^T \mathbf{v} + c_\ell(\beta, \zeta) = 0, \ \ \forall \ell \in \mathbf{I}, \quad (6)$$

where $c_\ell \in \mathbb{R}$ are scalars and A_ℓ are symmetric positive semi-definite matrices and $\mathbf{b}_\ell$ are vectors of size n. They finally showed that solving the hard constrained minimization problem is equivalent to minimize the Gibbs posterior energy:

$$L(\mathbf{v}, \beta, \zeta) = f_d(I, \mathbf{v}) + \sum_\ell \lambda_\ell^*(\beta, \zeta) g_\ell(\mathbf{v}, \beta, \zeta), \ \ \boldsymbol{\lambda}^* = \{\lambda_\ell^*\}. \quad (7)$$

where weights λ_ℓ^* can be optimally obtained in a dual formalism. However, a crucial issue still remains : how to chose optimally the multiscale prior model, *i.e.* infer parameters (β, ζ) of the power law given the image data?

2. PRIOR MODEL INFERENCE

In the previous section, we have introduced prior multiscale models (defined by power law factor β and exponent ζ or slope in log-log coordinates) for optic-flow non-parametric regularization. We now want to select the most appropriate multiscale model for motion estimation given only the image.

2.1. Bayesian hierarchical modeling

Bayes' rule provides a nice framework to evaluate multiscale prior model likelihood probability given the image data, 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:

$$I \rightarrow \mathbf{v} \rightarrow \beta, \zeta \quad (8)$$

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 (5):

- *Scaling model fitting.* We assume some scaling model parameters (ζ, β), i.e regularization weights $\boldsymbol{\lambda}^*(\zeta, \beta)$ (lagrangian multipliers) provided 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:

$$p(\mathbf{v}|I, \zeta, \beta) = \frac{p(I|\mathbf{v}, \zeta, \beta) p(\mathbf{v}|\zeta, \beta)}{p(I|\zeta, \beta)} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}} \propto p(I|\mathbf{v}, \zeta, \beta) p(\mathbf{v}|\zeta, \beta) \quad (9)$$

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

$$\frac{\exp\{-\frac{1}{2}\mathbf{v}^T(A_0 + A(\zeta,\beta))\mathbf{v} + (\mathbf{b}_0 + \mathbf{b}(\zeta,\beta))^T\mathbf{v} - c_0 - c(\zeta,\beta)\}}{Z_L(\zeta, \beta)} \quad (10)$$

  where $Z_L(\zeta, \beta)$ denotes the normalization constant also called the partition function.

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

$$p(\zeta, \beta|, I) = \frac{p(I|\zeta, \beta) p(\zeta, \beta)}{p(I)} \propto p(I|\zeta, \beta) p(\zeta, \beta). \quad (11)$$

  For a flat prior on variables (ζ, β), the MAP of Eq. 11 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:

$$p(I|\zeta, \beta) = \int_{\mathbb{R}^n} p(I|\mathbf{v}, \zeta, \beta) p(\mathbf{v}|\zeta, \beta) \mathrm{d}\mathbf{v}. \quad (12)$$

  Direct calculation of this integral is impractical due to its huge dimension. 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. 9). Therefore, we can rewrite the *evidence* as a normalization constant ratio:

$$p(I|\zeta, \beta) = \frac{\text{likelihood} \times \text{prior}}{\text{posterior}} = \frac{Z_L(\zeta, \beta)}{Z_{f_d} Z_{g_\ell}(\zeta, \beta)}, \quad (13)$$

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

2.2. 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:

$$Z_{f_d} = \int_{\mathbb{R}^m} \exp\{-f_d(I, \mathbf{v})\} \mathrm{d}I = (2\pi)^{m/2}, \quad (14)$$

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 normalization constant integral Z_L of the *posterior* PDF of Eq. 10 can be calculated using a Gaussian approximation:

$$\begin{aligned} Z_L(\zeta, \beta) &= \int_{\mathbb{R}^n} \exp\{-L(\mathbf{v}, \boldsymbol{\lambda}^*(\zeta, \beta))\} \mathrm{d}\mathbf{v} \quad (15) \\ &\simeq \exp^{\{-L(\mathbf{v}^*(\zeta,\beta), \boldsymbol{\lambda}^*(\zeta,\beta))\}} 2\pi^{\frac{n}{2}} \det(A_0 + A(\zeta,\beta))^{-\frac{1}{2}}, \end{aligned}$$

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. The determinant of such large and sparse matrices can be efficiently approximated via an incomplete LU decomposition. Finally, the *prior* PDF can be written as:

$$p(\mathbf{v}|\zeta, \beta) = \frac{\exp\{-\frac{1}{2}\mathbf{v}^T A(\zeta,\beta)\mathbf{v} + \mathbf{b}^T(\zeta,\beta)\mathbf{v} - c(\zeta,\beta)\}}{Z_{g_\ell}(\zeta, \beta)}. \quad (16)$$

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 does not need to be specified since it modifies the vector $\mathbf{b}$ but does not have impact on the hessian matrix A (even if the form of A is changed considering 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 a gaussian approximation:

$$\begin{aligned} Z_{g_\ell}(\zeta,\beta) &= \int_{\mathbb{R}^n} \exp\{-\sum_\ell \lambda_\ell^*(\zeta,\beta) g_\ell(\mathbf{v})\} d\mathbf{v} \qquad (17) \\ &= \underbrace{\max_{\mathbf{v}} \left(\exp^{\{-\sum_\ell \lambda_\ell^*(\zeta,\beta) g_\ell(\mathbf{v})\}} \right)}_{=1} 2\pi^{\frac{n}{2}} \det A(\zeta,\beta)^{-\frac{1}{2}} \end{aligned}$$

As the set of admissible solution for $\mathbf{v}$ for the self-similar constraint is not empty, the exponential term in Eq. 17 has a maximum value equal to 1. Finally, using Eq. 13, Eq. 15 and Eq. 17, the log evidence of the scaling model reads:

$$\log p(I|\zeta,\beta) \propto \underbrace{-f_d(\mathbf{v}^*, I)}_{\text{data term}} \underbrace{-\frac{1}{2}(\log \frac{\det(A_0+A)}{\det(A)})}_{\text{log Occam factor}} \qquad (18)$$

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 the term $\sum_\ell \lambda_\ell^* g_\ell(\mathbf{v}^*)$ does not appear in Eq. 18 as the constraints vanish at the saddle point.

3. ATMOSPHERIC TURBULENCE

The prior model inference has been assessed on 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. The image spatial resolution is 3×3 km^2 at the center of the whole Earth image disk. According to the physical-based methodology proposed in (3), a set of sparse pressure difference images of 256×256 pixels related to a stack of layers (low and intermediate altitude) have been derived. As detailed in this paper pressure-based cloud classification and images of top of cloud pressure have been used to create the set of input images displayed in figure 1. We then used the direct observation model designed by the authors for those sparse images. It is based on a layer continuity equation which relates the image intensity functions to vertically averaged horizontal wind fields.

3.1. Model selection : inference of inter-scales energy flux

In these experiments, 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 moment up to separation of about 10 kilometers (6)). We thus

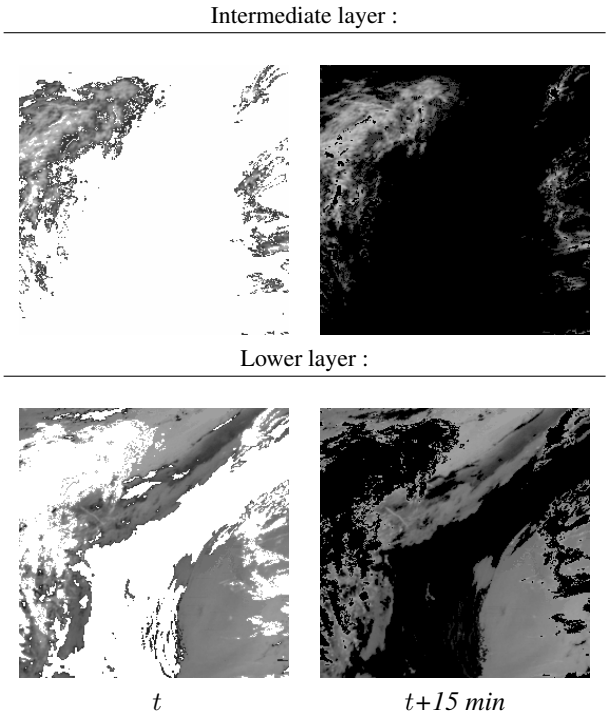

Fig. 1. ***Input meteorological images.*** *Sparse pressure difference maps of layers at intermediate (above) and low (below) altitude. White pixels of image at time t and black pixels at time $t+15$ minutes are areas with no observations. The images characterize the layers evolution in a time interval of 15 minutes.*

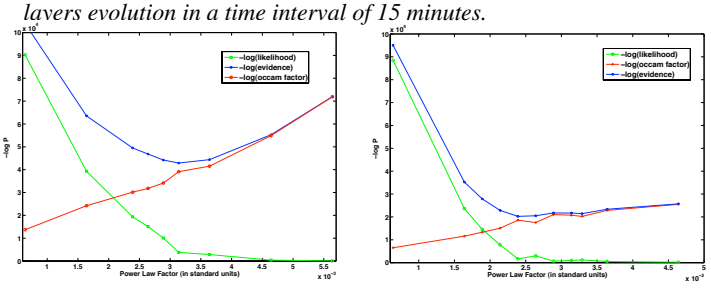

Fig. 2. ***Evidence maximization w.r.t energy flux***. *Minus log of evidence (blue), likelihood (green) and Occam factor (red) v.s. power law factor β (i.e. energy flux ϵ) for horizontal winds at low (**left**) and at intermediate (**right**) altitude*

only need to infer the parameter β by evidence maximization. Figure 4 shows that the evidence maximum is around $\hat{\beta}_{mid} = 0.0024$ for the middle layer and around $\hat{\beta}_{low} = 0.0031$ for the lower layer. This plot also illustrates the shared contribution of the *Occam* factor and the *likelihood* (or 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 (6) provides an expression for the 2-nd order moment where the energy flux can be related in the scale range $\mathbf{I}$ to the power law factor by $\beta = C_2 \epsilon^{\frac{2}{3}}$. Parameter $C_2 \simeq 6$ denotes a Kolmogorov constant. Therefore, the maximum of the evidence also provides the most likely energy flux at low and medium altitude $\hat{\epsilon}$:

$$\begin{cases} \hat{\epsilon}^{mid} \simeq 0.79 \times 10^{-5} m^2 s^{-3} \\ \hat{\epsilon}^{low} \simeq 1.20 \times 10^{-5} m^2 s^{-3}. \end{cases} \qquad (19)$$

These estimates have the same order of magnitude as previ-

ous reported results based on aircraft data analysis[1]. Thus this agreement is in our opinion very good as the measure is only based on image data. An energy spectrum comparison in figure 4 shows that on the contrary to the motion estimator proposed in (3), the present method does not underestimate the energy flux. 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 approach which need to first extract pseudo motion observations from the images and then estimate independently those physical quantities.

3.2. Wind field estimation

Figure 3 displays wind fields at intermediate altitude estimated with the most evident prior model (*i.e.* for $\hat{\beta}$). They are superimposed on the sparse image observations. Figure 4 shows the statistical consistency of the estimated motion fields by comparison with the turbulence modeling proposed in (7). In particular, one can notice that the estimated second order structure functions and the power spectra are in good agreement.

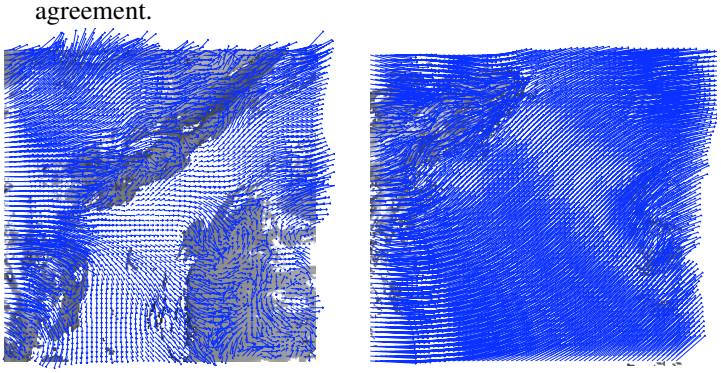

Fig. 3. ***Motion at different altitude.*** *Estimated horizontal wind fields at low (left) and intermediate (right) altitude.*

4. CONCLUSION

We have presented a bayesian hierarchical model which simultaneously provides optimal solutions for two problems: motion estimation and prior model selection from image sequences. Multiscale optic-flow prior models describe turbulent flow self-similarity resulting in power law behaviors for the second order moment of the PDF of motion increment. A hard constraint minimization scheme is used to provide an *a posteriori* motion estimate which does not involve the tuning of any regularization weight. Bayesian *evidence* provides a reliable criteria to select the most likely prior model given the image sequence data. Experiments and comparison with *in situ* measurements prove that the method constitutes an interesting tool for physical characterization of atmospheric motion. In particular, we show that bayesian *evidence* enables to recover directly from a meteorological image sequence consistent flux across scales.

[1]Authors in (7) estimated an average energy flux value of $\epsilon \sim 6 \times 10^{-5} m^3 s^{-3}$ for the stratosphere. A collection of other *in situ* measurements shows a typical value close to $\sim 10^{-5} m^3 s^{-3}$ (8)

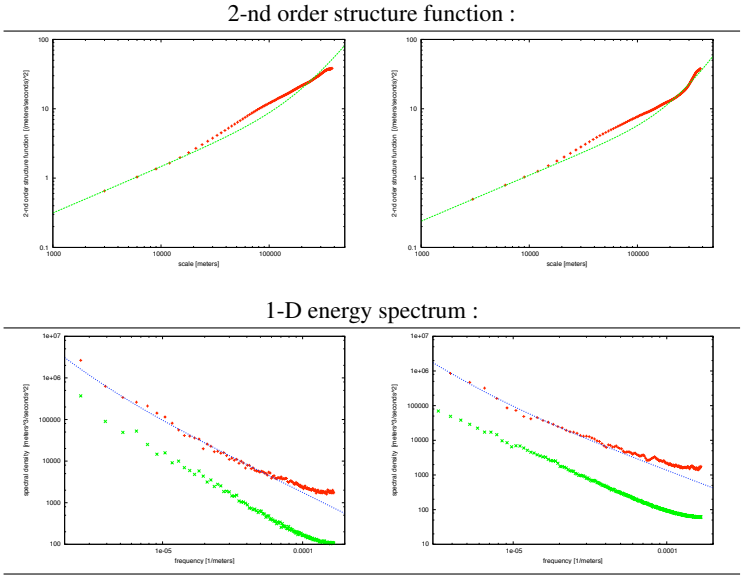

Fig. 4. ***Energy spectra (line below) and second order structure functions (line above) at low (left column) and intermediate (right column) altitudes.*** *2-nd order structure functions (red crosses) are plotted with their associate models given by (7) (green dashed line). 1D energy spectra obtained by our approach (red crosses) can be compared to their models given by (7) (blue dashed line) and to results from (3) (green stars).*

References

[1] P. Héas, E. Memin, and D. Heitz, "Self-similar regularization of optic-flow for turbulent motion estimation," in *ECCV'09 first International Workshop on Machine Learning for Vision-based Motion Analysis - MLVMA08*, Marseille France, 2008.

[2] T. Corpetti, E. Mémin, and P. Pérez, "Dense estimation of fluid flows," *Pattern Anal Mach Intel*, vol. 24, no. 3, pp. 365–380, 2002.

[3] 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*, vol. 45, no. 12, pp. 4087–4104, 2007.

[4] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[5] David J. C. MacKay, "Bayesian interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415–447, 1992.

[6] E. Lindborg, "Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence," *J. Fluid Mech.*, vol. 388, pp. 259–288, 1999.

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

[8] E. Dewan, "Saturated-cascade similitude theory of gravity wave spectra," *Journal of Geophysical Research*, vol. 102, pp. 799–818, 1997.