

# An ensemble filter estimation scheme for Lagrangian trajectory reconstruction

Yin Yang, Dominique Heitz, Etienne Mémin

► **To cite this version:**

Yin Yang, Dominique Heitz, Etienne Mémin. An ensemble filter estimation scheme for Lagrangian trajectory reconstruction. 16ème Congrès Francophone de Techniques Laser pour la mécanique des fluides, CNRS, CentraleSupélec, Université Paris Saclay, IRSN, Sep 2018, Dourdan, France. pp.8. hal-02097724

**HAL Id: hal-02097724**

**<https://hal.archives-ouvertes.fr/hal-02097724>**

Submitted on 12 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# An ensemble filter estimation scheme for Lagrangian trajectory reconstruction

Yin YANG<sup>1</sup>, Dominique HEITZ<sup>1,2</sup>, Etienne MEMIN<sup>2</sup>

<sup>1</sup>Irstea, UR OPAALE, F-35044 Rennes Cedex, France

<sup>2</sup>Inria, Fluminance group, Campus universitaire de Beaulieu, F-35042 Rennes Cedex, France

Email auteur correspondant : yin.yang@irstea.fr

Particle tracking velocimetry (PTV), also referred to as Lagrangian particle tracking (LPT) has recently gained considerable revival. The trend started with the Iterative Particle Reconstruction (IPR) method that applied a projection-matching scheme, to reconstruct 3D particles' positions rather than voxel-based intensity, like in Tomographic PIV. Recently, IPR has given rise to the Shake-The-Box (STB) method able to tackle densely seeded flows with considerably high accuracies and reasonable computational efforts. However, in most of 3D turbulent flows, image-based experiments can only provide sparse spatiotemporal data, for which STB is not able to track particles. If more robust estimations are possible, something useful may be learnt from the coupling between dynamical models and image data. In responding to these problems, we introduce a novel approach originated from the data assimilation technique comprising a sampling-based optimal estimation algorithm, namely a group of ensemble-based filtering variational schemes. We found that employing such an ensemble-based optimal estimation method helped tackling the problems associated with STB: the inaccurate predictor and/or the robustness of the optimization procedure. The proposed method (ENS) was quantitatively evaluated with synthetic particle image data built by transporting virtual particles in a turbulent cylinder wake-flow at Reynolds number equal to 3900. We examined the mean positional error of the reconstructed particles, the fraction of track lost particles as well as the required CPU time/memory. We observed that even at large ppp levels ( $>0.1$ ), the mean positional error of ensemble method was considerably lower than the one given by the STB method. Besides ENS performed equally well in terms of data series of relatively large time separation. These preliminary results indicates that the ensemble-based method was indeed effective.

## 1 Introduction

In the fields of fluid dynamics research and application, reconstructing the 3D flow through imaging particles driven by the flow, followed by inferring spatiotemporal flow structures from particle images, is an important method to investigate the small scale flow phenomena related to the turbulence. Carrying out such investigations using CFD tools is hardly applicable to such problems due to lack of exact knowledge to calibrate the dynamic model.

The algorithms used to infer spatiotemporal flow fields from particle images can be classified under two categories. Tomographic particle image velocity (TomoPIV) [1] centers on a tomographic reconstruction of particles in voxel basis as a discretization of the 3D object space. Then the Eulerian velocity field is obtained by applying the classic cross-correlation technique inherited from standard 2D PIV technique to 3D intensity voxels. Thus the central task is to solve for the intensity at voxel coordinate given particle images. We can solve the problem iteratively using algebraic reconstruction technique (ART), MART and SMART algorithms [2]. The TomoPIV technique has gained considerable success since the last decade due to its robustness and its capability of dealing with particle image of high seeding density (particle per pixel, ppp around 0.05). Nevertheless the TomoPIV technique still suffers from several flaws such as large amount of ghost particles, relatively large particle positional error and computationally demanding.

On the other hand, 3D particle tracking velocimetry (3D-PTV), also referred to as Lagrangian particle tracking (LPT), aims at following particle positions through time. The 3D-PTV has been proposed since 1990s [3, 4] but this approach is limited by the fact that it can only tackle images of rather low seed-

ing densities. But 3D-PTV recently has regained popularity pioneered by the advanced iterative particle reconstruction (IPR) approach proposed in [5]. The main contribution of IPR is its using of active image matching scheme inspired by the success of TomoPIV in terms of particle tracking. The central idea following this paradigm is to find the best estimation of particles' positions through iteratively matching the projection between their prior distributions and their observational records captured by cameras. This method is different from the classic 2D PIV or aforementioned voxel-based TomoPIV technique, in which the local velocity is calculate through cross-correlation operation on the interrogation domain composed either by pixels or voxels. In pure Lagrangian particle tracking method, once we are able to reconstruct the particle trajectory, the Lagrangian velocity field is readily available through simple difference method. Nevertheless, in practice, we prefer to work with Eulerian field define on fixed grids, so additional interpolation procedure must be applied [6, 7].

Recently the Shake-The-Box (STB) algorithm proposed in [8] has shown its superiority in terms of reconstruction accuracy, occurrence of ghost particles as well as usage of computational resources compared to TomoPIV. The STB method is rooted in IPR combined with an additional prediction phase able to introducing temporal information. Consequently the difficulties of finding the uniqueness solution in terms of the optimization procedure is alleviated especially dealing with particle images of high seeding densities. However, we have found that STB is less effective when the Wiener filter or polynomial extrapolations fails to predict the particles' positions. This corresponds to the cases of sparse temporal data or data extracted from complex flows. Furthermore, large prediction errors lead to severe convergence issues where the simple "Shaking" optimization algorithm is trapped by local minimum.

In responding to these problems, we consider reformulating the particle reconstruction procedure under the framework of data assimilation (DA) since the core of STB essentially shares similarities with DA. The difference is that STB focused on reconstruction from data by adding an adhoc model whereas DA focused on controlling the governing dynamical equations by inferring the optimal state and/or parameter from data.

Therefore, we propose to integrate the reconstruction strategies used in IPR/STB approaches into data assimilation framework. Our motivation is twofold, on the one hand, the image matching scheme linking the 3D object space and 2D image space is proven to be quite accurate that provides us a good observation operator allowing extracting as much as information from data, on the other hand, introducing dynamics into the data assimilation framework allows the unobserved variables (e.g. velocity, pressure) to be easily inferred. We are particularly interested in ensemble-based method which provides an efficient solution of the complex nonlinear optimization problem. The complexity is either due to the nonlinearities of dynamics [9] or due to the nonlinearities of the observation operator. Ensemble methods, based on Monte-Carlo approximations, provides a stable and solid solution in transforming the original nonlinear problem to a linear one. Note that this paper concentrates on the aspect of employing ensemble approaches in the correction phase thus we continue to use polynomial-based filter to predict state. The advantages of introducing complex dynamics into the reconstruction procedure will be reported in the following paper.

## 2 Method

### 2.1 Recap of Shake-The-Box (STB)

The STB algorithm begins with an initialization phase in which the information of the first few tracks are obtained either by IPR, TomoPIV or CFD results. After the initialization phase the errors can be quite high for densely seeded flows partly due to a large portion of undetected particles, partly due to large amount of ghost particles. Therefore by introducing temporal information, STB performs quite well within several time steps of convergence phase in terms of correcting these errors by finding the missing particles as well as eliminating the ghost particles. Note that a ghost particle is discarded if its track at certain time step given by the predictor can not be found through searching around in its radius in image space. If the predictor works well, we can envisage a fast separation of ghost particles and real particles. Eventually we enter a converged phase where almost all particles have been identified and tracked. The STB scheme dealing with two snapshots functions exactly the same in the convergence phase and the converged phase. It can be interpreted as an recursive filtering scheme based on a two stage prediction-correction strategy. In this article, we concentrate on studying this filtering scheme in STB and improving it by our proposed scheme. For sake of simplicity, we consider the prior particles's

tracks are known to a large extent that corresponds to the converged phase. In the following, we highlight the optimization scheme used STB method and propose our ensemble-based (ENS) optimization method in this regard. We also show the relation between the optimization procedure used in STB and ENS approaches.

### 2.1.1 STB optimization

In STB, for particle  $p$  at time step  $k$ , we start by obtaining the prior position  $\mathbf{X}_p^k = \varphi_k(\hat{\mathbf{X}}_p^{k-1})$  by forecasting the posterior position  $\hat{\mathbf{X}}_p^{k-1}$  at time level  $k-1$  to  $k$  using predictor  $\varphi_k$ . Note the notation  $\hat{\bullet}$  denotes the posterior estimation that is either given by the previous correction stage or by other methods such as IPR or TomoPIV. In the correction stage, we search iteratively for  $\mathbf{X}_p^k$  that minimizes:

$$J(\mathbf{X}_p^k) = \sum_i \|I_{res+p,k}^i - I_{part,k}^i(\mathbf{X}_p^k)\|^2. \quad (1)$$

where  $i$  is the camera index,  $I_{part,k}^i(\mathbf{X}_p^k)$  consists of an optical transfer function (OTF) that projects the current particle  $p$  into the image space, and  $I_{res+p,k}^i$  is the augmented particle image defined by  $I_{res+p,k}^i = I_{res}^i + I_{part}^i(\mathbf{X}_p^k)$ .  $I_{res}^i$  is the residual image computed as the difference of recorded image  $I_{rec}^i$  and the projected image  $I_{proj}^i = \sum_p^P I_{part}^i(\mathbf{X}_p^k)$  where  $P$  is the total particle size.

This optimization approach used in IPR/STB belongs to a group of model-based derivative-free trust region optimization methods. Under this vein, we employ a quadratic form:  $m_k(\delta X_p)$  that locally approximates the original cost function  $J(\delta X_p)$  in terms the particle's  $X$  coordinate while keeping  $Y$  and  $Z$  coordinates invariant:

$$m_k(\delta X_p) = c + g^T \delta X_p + \frac{1}{2} \delta X_p^T G \delta X_p$$

Then with 3 sample points we are able to uniquely determine the coefficients  $c, g, G$ , which are all scalar with respect to a single particle. For each particle, we need to compute  $I_{part}^i(\mathbf{X}_p^k)$  7 times. Compare to the time dedicated to computing the image projection, the computing time used to determine the extreme of above quadratic function is trivial.

## 2.2 Ensemble methods (ENS)

Our proposed scheme focuses on improving the performance of the recursive filter discussed in previous section through the introduction of ensemble averaging in the prediction phase and the linearization of nonlinear cost function using an ensemble in the correction phase. We write below the cost function in standard DA form [9] in terms of the increment  $\delta \mathbf{X}_p = \mathbf{X}_p^k - \mathbf{X}_p^{f,k}$  considering the observation operator  $I_{part,k}^i$  defined as in the STB method:

$$J(\delta \mathbf{X}_p^k) = \frac{1}{2} \lambda \|\delta \mathbf{X}_p^k\|^2 + \frac{1}{2} \sum_i \|I_{res+p,k}^i - I_{part,k}^i(\mathbf{X}_p^{f,k} + \delta \mathbf{X}_p^k)\|^2. \quad (2)$$

Compared to (1), we have an extra term on  $\delta \mathbf{X}_p^k$  that acts as the regularization term preventing the system from overfitting the data. In (2), we introduced the regularization parameter  $\lambda$  in terms of the prior particle state. A major difficulties in finding the minimum of above problem is that the particle projection operator  $I_{part}$  is a nonlinear function of particle's 3D position  $\mathbf{X}_p$ . Using sophisticate nonlinear optimization algorithm can lead to good results with substantial computing time, yet the convergence is not guaranteed. We can linearize cost function (2) by expanding the Taylor series of function  $I_{part,k}^i$  around the prior state  $\mathbf{X}_p^{f,k}$  that leads to an alternative version:

$$J(\delta \mathbf{X}_p^k) = \frac{1}{2} \lambda \|\delta \mathbf{X}_p^k\|^2 + \frac{1}{2} \sum_i \|\mathbf{D}_{k,p}^i - \partial_{\mathbf{X}} I_{part,k}^i \delta \mathbf{X}_p^k\|^2. \quad (3)$$

where  $\partial_{\mathbf{X}} I_{part,k}^i$  is the tangent linear model  $I_{part,k}^i$ .<sup>1</sup>

<sup>1</sup>The tangent linear model  $\partial_{\mathbf{X}} I_{part,k}^i$  is defined by:  $\lim_{\beta \rightarrow 0} \frac{I_{part,k}^i(\mathbf{X}_p^k + \beta d\mathbf{X}_p^k) - I_{part,k}^i(\mathbf{X}_p^k)}{\beta} = \partial_{\mathbf{X}} I_{part,k}^i(\mathbf{X}_p^k) d\mathbf{X}_p^k$ .

We reformulate above cost functions in ensemble-based framework in which the tangent linear model  $\partial_{\mathbf{x}} I_{part,k}^i$  is approximated by an ensemble. To start, we introduce ensemble notations. Denote  $\mathbf{E}_k$  the ensemble state containing  $N$  sample state at time level  $t_k$ :

$$\mathbf{E}_k = (\mathbf{X}_{1,k}, \dots, \mathbf{X}_{l,k}, \dots, \mathbf{X}_{N,k}) \in \mathbb{R}^{P \times N}, \quad (4)$$

from which, we compute the ensemble perturbation matrix  $\mathbf{X}'_k$  through:  $\mathbf{X}'_k = \mathbf{E}_k - \bar{\mathbf{E}}_k$ , where the column of  $\bar{\mathbf{E}}_k$  contains  $N$  replicates of the ensemble mean,  $\bar{\mathbf{X}}_k = \frac{1}{N} \sum_{l=1}^N \mathbf{X}_{l,k}$ . It is also convenient to define  $\mathbf{A} = \frac{1}{\sqrt{N-1}} \mathbf{X}'$  the ensemble anomaly matrix.

The forecast stage therefore corresponds to propagate the entire ensemble state from time  $t_{k-1}$  to  $t_k$  using the predictor operator  $\varphi_k$ :

$$\mathbf{E}_k^f = \varphi_k(\hat{\mathbf{E}}_{k-1}). \quad (5)$$

We then make the assumption that  $\delta \mathbf{x}_k$  lives in the column space of  $\mathbf{A}_k^f$ :  $\delta \mathbf{x}_k^f = \mathbf{A}_k^f \mathbf{w}$ , which leads to a cost function in terms of the weighting vector  $\mathbf{w}$ :

$$J(\mathbf{w}) = \frac{1}{2} \lambda' \|\mathbf{w}\|^2 + \frac{1}{2} \sum_i \|\mathbf{D}_{k,p}^i - \partial_{\mathbf{x}} I_{part,k}^i \mathbf{A}_k^f \mathbf{w}\|^2, \quad (6)$$

which is a linear optimization problem whose solution can be easily found using gradient-based optimization scheme. The ensemble trick consists in approximating the projection of ensemble anomaly matrix  $\mathbf{A}_k^f$  using tangent linear model  $\partial_{\mathbf{x}} I_{part,k}^i$  through the difference of the projection of the entire ensemble minus ensemble mean using the nonlinear projection model  $I_{part,k}^i$ :

$$\partial_{\mathbf{x}} I_{part,k}^i \mathbf{A}_k^f \approx \frac{1}{\sqrt{N-1}} \left( I_{part,k}^i(\mathbf{E}_p^k) - I_{part,k}^i(\bar{\mathbf{E}}_p^k) \right). \quad (7)$$

We think such ensemble-based scheme would help resolving the optimization problem based on following arguments. Firstly, the reconstruction results of STB method showed that the model-based derivative-free trust region method is indeed effective. This optimization method is based on the Taylor series expansion of the cost function at the prior state. The high order terms (high than 2) is discarded. In ensemble-based method, a similar strategy is employed to linearize the cost function, however instead of using the Taylor series of the cost function, the nonlinear observation operator is expanded around the prior state. Indeed the ensemble formulation allows the approximation of the first order Taylor series (tangent linear model) through the ensemble states driven by nonlinear projection model. Secondly, the ensemble formulation of the cost function as derived from MAP (Maximum a posteriori) estimator introduces naturally a regularization term. Thirdly, the prior ensemble consists of a cloud of potential candidatures of the particle's position, contrary to only one position as the starting of the search space for STB/IPR methods. By combine different candidatures in the prior ensemble, the search space can be extended and less likely to fall into local minimum. In the meantime, the weighted averaging of different candidatures reduces the potential prediction error. This is known as ensemble averaging technique. Finally, last gradient-based optimization scheme can be employed to solve problem (6). During the whole optimization process, the computing time is also dominated by projection of different ensemble member into image space.

## 3 Experiment results

### 3.1 Synthetic particle images creation and experiment configuration

We intend to apply our proposed ensemble-based estimation scheme to particle image data and evaluate quantitatively its performance compared to the STB scheme. To this end, we have created synthetic images by projecting virtual particles driven by known Eulerian velocity field to virtual cameras. The Eulerian velocity data was obtained with a Large Eddy Simulation (LES)[10]. We have chosen the source velocity field within the domain of interest of size  $6D \times 6D \times \pi D/3$ , which is discretized into a regular mesh of size  $291 \times 291 \times 16$ , in a turbulent wake-flow past a circular cylinder at Reynolds number equal to 3900. This subdomain is cropped from the whole computing domain of size  $20D \times 20D \times \pi D$  where  $D$  is the characteristic diameter of the cylinder. The dimensionless simulation time step  $\Delta t_{sim}$  is 0.003. The initial distribution of the virtual particle are randomly generated within the subdomain, then

each particle is transported according to its velocity calculated as trilinear interpolation of the velocity vectors on the nearest 8 neighboring grids. As for the time marching scheme, we use the first-order Euler method to solve for the predicted particle position.

All virtual particles transported by the flow are projected onto four virtual cameras ( $1000 \times 1000$  pixels each) under a cross-like configuration with a rotation of  $\pm 30^\circ$  around x and y axis respectively. All cameras are pre-calibrated using both polynomial mapping model and pinhole model. In terms of the OTF parameter [11], we employed a uniform two dimensional Gaussian form resulting in a constant particle diameter of 2.5 px. The tested seeding density varies from 0.01 to 0.125 ppp.

Although particle initialization is a complex issue for both STB and ENS schemes, STB has been reported to achieve the converged state quite efficiently. All related techniques used in STB can be applied to our ENS scheme that help reaching the converged state quickly. Therefore in the current study, we employed a simplified triangulation scheme to obtain the particles' tracks of the first 9 snapshots that constitute the prior distribution of particles of prediction phase for both ENS and STB. In our case, the velocity field is characterized by fast fluctuations induced by turbulence, unlike smooth slowly varying velocity field used in other studies. Thus predicting the particle's position with polynomial-based predictor can be quite inaccurate. To improve the accuracy of the predictor, we have used a longer history of data time series (9 against 4 snapshots used in [8]). When the time interval between two consecutive snapshots  $\Delta T_{obs}$  is relatively short compared to the dimensionless simulation time step, the polynomial-based predictor remains effective. However, the prediction errors become higher with larger  $\Delta T_{obs}$  since the temporal distribution of observations cannot capture the fast spatially changing velocity. So we are inclined to lose track of more particles under such circumstances. We employed two test cases with different  $\Delta T_{obs}$  (0.1 for case I and 0.5 for case II respectively) to evaluate the performance of STB/ENS handling sparse spatiotemporal data. The mean particle displacement is a fraction of pixel for case I whilst case II lead to a larger displacement of nearly 2 pixels.

We considered that the issue of track lost particles is particularly relevant for time resolved reconstruction schemes. Because ideally, in the converged phase, only the particles entering the domain need to be reconstructed using triangulation procedure. With most particles have been identified successfully, the resulting residual image, after correction, is almost null except the peaks corresponding to the new particles. Eventually when more and more particles get lost during the tracking process, we need to employ additional triangulation procedure to re-track those lost particles. However, the triangulation procedure (or IPR) can become quite difficult when dealing with particle image of high density. So the risk of failure of STB becomes important when the number of lost particles is high compared to the successfully tracked particles. On the contrary, in the last section we have argued how ensemble-based method can overcome such limitations causing track lost particles. Still the convergences of both the 'Shaking' procedure in STB and the gradient-based optimization procedure in ENS depend on a warm start initialization. [8] proposed an additional procedure of *initial* shake to provide a good starting value for normal shake procedure. In ENS, we implemented a similar procedure to locate the warm-starting particle positions. We chose an ensemble member of 8 for ENS method.

In the following, we assess numerically both schemes by evaluating the mean positional error of detected tracks with respect to the synthetic tracks in pixel unit, the fraction of track lost particles compare to total particles as well as computational resources needed. In practice, we consider that a particle is lost if the deviation of its reconstructed location from its *true* location for current images is larger than 1 px. All track lost particles have to be relocated using triangulations so that they can be considered as re-detected. Both the image data ranging from the 10th snapshot up to the 50th snapshot and the prior particle distribution of the first 9 snapshots are passed to the proposed approach along with a STB approach for 3D particle tracks' reconstruction. The simplified triangulation procedure is applied to retrieve the positions of new particles entering the domain as well as the positions of lost particles, and a particle is removed when its position exceeds the boundaries of the domain.

## 3.2 Comparison of ENS to STB method

We plotted the temporal evolution of mean positional error of detected particles obtained by STB and ENS at different ppp levels for case I (Fig. 1a) and case II (Fig. 1b) respectively. We also plotted in figure 2a the mean positional errors of detected particles after the 50th snapshot of STB and ENS for both cases in terms of different ppp levels.

We observed that for data series of high time sampling rate (case I), both STB and ENS schemes

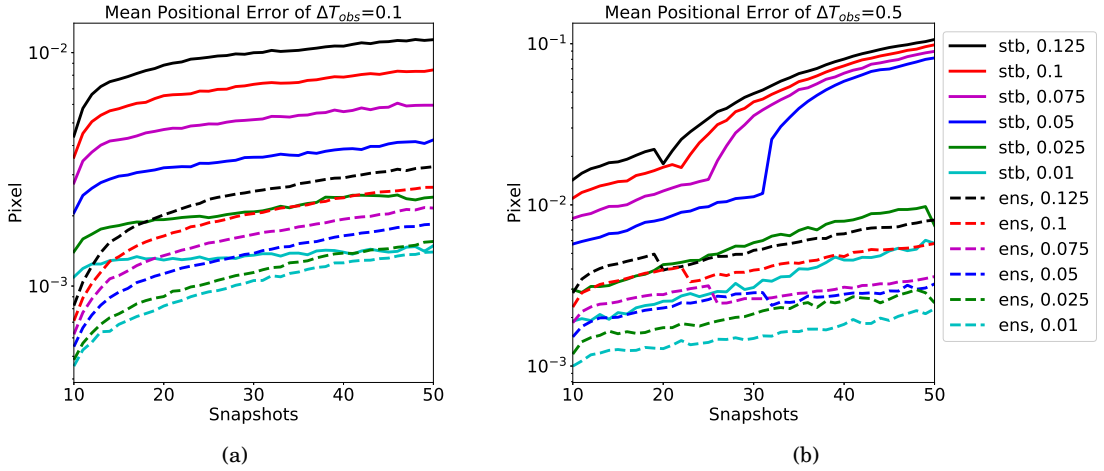


Figure 1: The temporal evolution of the mean positional errors in pixel of STB/ENS under different ppp levels (denoted by different color): (a), case I with  $\Delta T_{obs} = 0.1$ ; (b), case II with  $\Delta T_{obs} = 0.5$ .

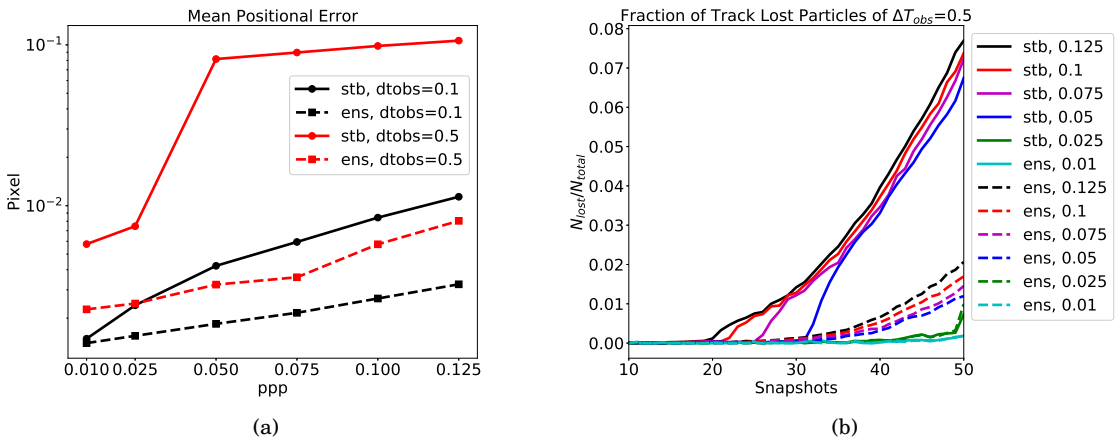


Figure 2: Mean positional errors after the 50th snapshot with respect to different ppp levels: (a), case I and case II; (b), comparison of fraction of number of track lost particles relative to number of total true particles of case II with  $\Delta T_{obs} = 0.5$ .

perform well and exhibit good temporal convergence in reconstructing the 3D particle tracks. Yet the errors of tracked particles obtained by ENS are 3-4 times lower compared to STB during the course of temporal tracking process. For both methods, the particles are almost never lost in the tracking process. The same findings can be observed with respect to sparse seeded flow ( $\text{ppp} < 0.5$ ) of case II featuring data series of low time sampling rate.

Nevertheless, when dealing with densely seeded flow ( $\text{ppp} \geq 0.5$ ) of case II, the evolution of mean positional errors obtained by STB diverged. The errors eventually reached the level of 0.1 px. Under this scenario (low time sampling rate, high ppp), the errors of ENS are nearly an order of magnitude lower compared to STB as revealed by both figures 1b and 2a. Such detrimental reconstruction quality of STB in this condition is also reflected by the fraction of track lost particles (figure 2b) that soared and reached nearly 10% for  $\text{ppp} > 0.05$ . The issue of track lost particles associated to STB indeed arise with data series of large time separation, which becomes more severe with densely seeded flow due to the overlapping particle images. ENS has produced far less track lost particles. Note that for STB, the diverged errors is not related to the exploding number of track lost particles. Because in our experimental setup, the triangulation introduced far less errors compared to the tracking process. Consequently, those triangulated particles are re-registered as detected at the end of the optimization process and they tend to pull down the error curve instead. For example, the small drop of error evolution of ENS (denoted by dashed line) in figure 1b can be explained by the small surge of number of track lost particles.

As in our experimental setup, the particle initialization is considered to be almost perfect. Therefore the inability of STB is largely associated with the failure of providing a good starting point stemming from the warm-start procedure. Note that even though the error divergence of STB can be partially avoided by carefully fine tuning parameters (search radius, number of iterations) used in the ‘initial’ shake phase, ENS is still superior. Because for one thing, ENS do not need such refining step and is able to converge with the same ‘cold’ initialization that leads STB to diverge; secondly, even with such additional fine tuning procedure, the resulting error (not shown here) is still considerably higher than that of ENS. The good performance of ENS has demonstrated its robustness especially dealing with data series of low time sampling rate. In this regard, we consider the ensemble-based method a better alternative facing the challenge of sparse spatiotemporal data.

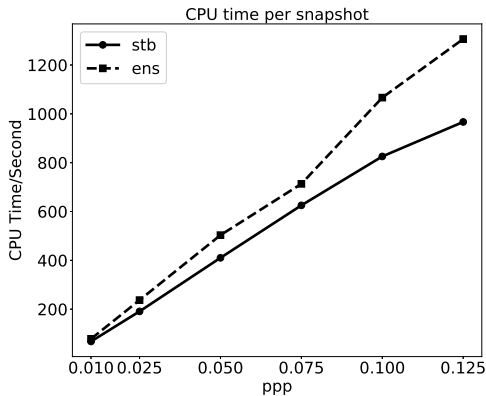


Figure 3: Comparison of sequential CPU time per snapshot by STB and ENS for case I.

We have compared the computing time and memory usage for STB and ENS with 8 members’ ensemble in figure 3. Note that all schemes evaluated here are implemented sequentially. The CPU time per snapshot of ENS is about 20% more than that of STB. However, the scalability of ENS is better since an ensemble can be parallelized efficiently. On the memory side, ENS is memory demanding and the memory required by ENS is a factor of ensemble members higher than STB under same ppp level. Nevertheless such memory usage is still affordable considering any modern computer with as much as several tens of GBs. Finally we highlighted that all results obtained by ENS method are using 8 ensemble members. If highly accurate reconstruction is desired, we can further improve the result by increasing the ensemble members at the expense of more computational cost.



## 4 Conclusions

The success of STB hinged on two factors: the image matching schemes as in IPR and a dynamical predictor exploiting the temporal coherence of the particle image data respectively. In this paper, we have proposed an ensemble-based optimal estimation scheme allowing integrating such active image matching schemes into data assimilation framework.

We have found that the ensemble method outperforms STB in terms of reconstruction accuracy under all circumstances (at least in our experiment setup). The success of ENS can be attributed to the fact that the minimization of the nonlinear data discrepancy term as in IPR method can be efficiently solved by linearizing the original problem to a linear quadratic form followed by ultra-fast gradient-based searching algorithms. More importantly, we have shown that ENS is even more effective dealing with sparse spatio-temporal data that cause STB to fail. Because the linearization through the ensemble, comprising a cloud of possible positional state of all particles, is able to explore a larger search radius than the quadratic points fitting algorithm used in STB.

Furthermore we have introduced the ensemble approach as an optimal estimation framework. This framework allows us to naturally couple the particle image based data, as well as the prior information brought in by sophisticated CFD models through a robust stochastic searching algorithm combined with the effective image matching schemes. This paves the way for many perspectives including incorporating sophisticated dynamic model and jointly estimating Lagrangian particle positions as well as Eulerian velocity fields.

## References

- [1] G E Elsinga, F Scarano, B Wieneke, and B W van Oudheusden. Tomographic particle image velocimetry. *Experiments in fluids*, 41(6):933–947, October 2006.
- [2] F Scarano. Tomographic PIV: principles and practice. *Measurement Science and Technology*, 24(1):012001–30, October 2013.
- [3] H G Maas, A Gruen, and D A Papantoniou. Particle tracking velocimetry in three-dimensional flows - Part 1. Photogrammetric determination of particle coordinates. *Experiments in fluids*, 15(2):133–146, July 1993.
- [4] N A Malik, Th Dracos, and D A Papantoniou. Particle tracking velocimetry in three-dimensional flows - Part II: Particle tracking. *Experiments in fluids*, 15(4-5):279–294, September 1993.
- [5] B Wieneke. Iterative reconstruction of volumetric particle distribution. *Measurement Science and Technology*, 24(2):024008–15, December 2012.
- [6] J FG Schneiders and F Scarano. Dense velocity reconstruction from tomographic ptv with material derivatives. *Experiments in Fluids*, 57(9):139, 2016.
- [7] S Gesemann, F Huhn, D Schanz, and A Schröder. From noisy particle tracks to velocity, acceleration and pressure fields using B-splines and penalties. In *18th international symposium on applications of laser and imaging techniques to fluid mechanics*, pages 4–7, 2016.
- [8] D Schanz, S Gesemann, and A Schröder. Shake-The-Box: Lagrangian particle tracking at high particle image densities. *Experiments in fluids*, 57(5):1–27, April 2016.
- [9] Y Yang, C Robinson, D Heitz, and E Mémin. Enhanced ensemble-based 4dvar scheme for data assimilation. *Computers and Fluids*, 115(C):201–210, July 2015.
- [10] P Parnaudeau, J Carlier, D Heitz, and E Lamballais. Experimental and numerical studies of the flow over a circular cylinder at reynolds number 3900. *Physics of Fluids*, 20(8):085101, 2008.
- [11] D Schanz, S Gesemann, A Schröder, B Wieneke, and M Novara. Non-uniform optical transfer functions in particle imaging: calibration and application to tomographic reconstruction. *Measurement Science and Technology*, 24(2):024009–15, December 2012.