Sparse analysis for mesoscale convective systems tracking

Jean-Baptiste Courbot, Vincent Duval, Bernard Legras

▶ **To cite this version:**

Jean-Baptiste Courbot, Vincent Duval, Bernard Legras. Sparse analysis for mesoscale convective systems tracking. 2019. hal-02010436v1

HAL Id: hal-02010436

https://hal.science/hal-02010436v1

Preprint submitted on 7 Feb 2019 (v1), last revised 16 Sep 2020 (v2)

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.

Sparse analysis for mesoscale convective systems tracking

Jean-Baptiste Courbot, Vincent Duval, and Bernard Legras.

***Abstract*—In this paper, we study the tracking of deformable shapes in sequences of images. Our target application is the tracking of clouds in satellite image. We propose to use a recent state-of-the-art method for off-the-grid sparse analysis to describe clouds in image as mixtures of 2D atoms. Then, we introduce an algorithm to handle the tracking with its specificities: apparition or disappearance of objects, merging, and splitting. This method provides similar numerical outputs as the recent state-of-the-art alternatives, while being more flexible, and providing additional information on, *e.g.*, cloud surface brightness.**

***Index Terms*—Shape tracking, Mesoscale Convective System tracking, sparse image analysis**

I. Introduction

A. Problem formulation

IN this paper, we study the tracking of deformable shapes in sequences of images. Our target application is the tracking of organized cloud systems, called Mesoscale Convective Systems (MCS) in infrared satellite images.

The MCS are formed when clouds generated by convective instability cluster as a common cloud system with a large upper deck of thick opaque ice clouds accompanied by precipitations over a large area [1]. Such systems are the major part of the high cloud cover in the tropical region where they can reach sizes of up to 1000 km in diameter. Under some favorable circumstances, they can eventually organize as tropical cyclones. The behavior of aggregation and of the MCS life cycle is complex and only partly understood [2], [3]. Collecting large amounts of information about MCS is a required step towards a finer understanding of their behavior. MCS are mostly observed through satellite imaging, notably from geostationary imagers that chart the full tropical cloud cover at a resolution of a few kilometers every ten or fifteen minutes. The high clouds are particularly visible in the infrared window channel, which provides a measure of the emitting surface brightness (temperature, in Kelvin degrees). As the cloud top temperature is at least 50K lower than the surrounding ground or low clouds, the high clouds can be easily isolated (see Figure 1). Due to the complexity and the huge amount of available information, there is a need for an automatic tool able to detect and keep track of MCS in time series of these images.

This problem is challenging in two aspects. First, we do not know in advance when, where, and in which quantity MCS may appear in the images. In addition to their apparition and disappearance, MCS may also split or merge. Secondly, to extract quantitative information about MCS, one must be at least able to describe in an appropriate fashion its shape and its temperature. Hence, the descriptors to extract must contain this information.

B. Previous works

1) Cloud tracking in satellite images: In the literature, several works address specifically the problem of MCS tracking in infrared satellite image. In [5], the authors propose a method based on a thresholding step followed by an area-overlap linking step to perform the tracking. The final product, used for measurements, is an ellipse fit on the contours. This method coarsely describes shapes, and does not account for local brightness temperature minima in MCS (see Fig. 1), so close MCS components are indistinguishable.

This aspect of the problem has been handled in [6], in which the authors propose an algorithm, named TOOCAN, similar to region growing techniques [7]. The method consists in iteratively dilating "seeds" from lower to higher brightness temperature, under a maximum contrast constraint. This algorithm does separate several components of MCS, and provides as in [5] a labeled map: contours are only implicitly defined, and there is no information on the pixel intensities within each labeled regions.

Alternatively, the authors of [8] propose to use a Bayesian tracking method, based on the Multiple Hy-

JBC was with Inria Paris, MOKAPLAN team and LMD, UMR 8539, PSL-ENS/CNRS/Sorbonne Université/Ecole Polytechnique, Paris, France. JBC is now with Institut de Recherche en Informatique, Mathématiques, Automatique et Signal (IRIMAS EA7499), Université de Haute-Alsace, Mulhouse, France. Contact: jean-baptiste.courbot@uha.fr

V. Duval is with Inria Paris and CEREMADE, Université Paris-Dauphine & PSL, France.

B. Legras is with LMD, UMR 8539, PSL-ENS/CNRS/Sorbonne Université/Ecole Polytechnique, Paris, France.

This work was completed within the French Idex with reference number ANR-10-IDEX-0001-02 PSL.

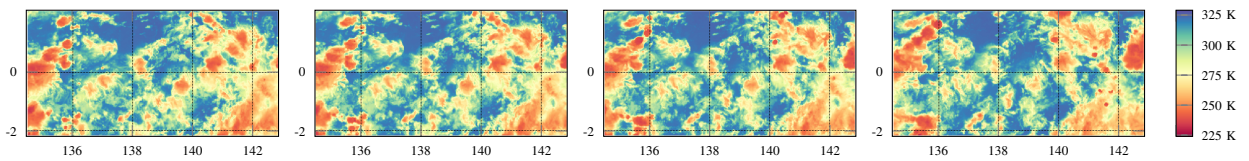


Fig. 1: Example of four consecutive infrared images provided by the Himawari satellite [4], with 2×2 km at the sub-satellite point (140.7°E, 0°N). In this figure and in the following, the colors encode brightness temperature within the same range.

pothesis Tracking (MHT) framework [9], [10]. In that paper, the authors first threshold images and fit ellipses to the resulting connected components. These ellipses are the input of the MHT method, which explicitly handles the birth, death, split and merge events. However, the output of this method consists in ellipse contour only, and does not provide information about the cloud's surface brightness.

Hence, to our knowledge, there is no method in the literature handling both intensity and shape information for MCS tracking. Our approach towards this goal is be the modeling of MCS as mixtures of 2D atoms, such as Gaussian ellipsoids (detailed below).

2) Object and shape tracking: Beside MCS tracking, there is a large corpus of literature addressing the problems of object tracking. One of the main approach in this regard is the Bayesian particle filtering methods [11] and its developments [12]. This approach relies on the sampling of a given parameter space, weighted by an appropriate likelihood/data fit term. It allows, in particular, the analysis of non-normally-distributed movements that are considered in Kalman filtering [13]. Regarding the apparition and disappearance of multiple objects in time sequences, the Multi-Bernoulli Particle Filter (MBPF) [14] and its improvements [15], [16], [17] seem relevant to our problem. In many cases, particle filters are applied in a localization context: objects to be tracked are characterized by a few parameters (*e.g.* position and size). In some works, such as [18], the parameters are used for a contour tracking framework.

Nevertheless, in the context of this work the assumption of fixed and low-dimension description does not hold, so for each mixture its cardinal must be sampled as well as its parameters. This leverages convergence, speed, and prior distribution fitting issues. Hence, sampling-based methods such as particle filtering are not well-suited to our problem.

3) Sparse recovery: On the other hand, tools from the compressed-sensing and the sparse recovery theories are specifically designed for this purpose. The topic of sparse recovery and optimization under sparse constraint is well studied (see [19] for a survey). The sparsity prior may be leveraged using a greedy approach such as the OMP algorithm [20], or in a variational approach involving a sparsity promoting functional (*e.g.* the ℓ_1 norm) as in the LASSO [21] or Basis Pursuit [22]. However, these methods are discretized: the atoms to search for are often sampled on a pre-established grid, commonly called dictionary. Choosing the dictionary for a given problem has important consequences in the result quality, as well as in the performance of the considered algorithm.

In recent years, several methods have been proposed to overcome the issues introduced by discretization. In [23], the Continuous Basis Pursuit interpolates between parameters defined on a grid so as to reach subgrid accuracy. Another approach, which follows from [24], [25], [26], is to work in a fully continuous setting, replacing the ℓ_1 norm with its continuous counterpart, the total variation of measures. Following [27], we call this method the BLASSO.

To numerically solve the BLASSO problem, several approaches have been proposed. In [25], [28], the problem is recast as a semi-definite program, whereas the ADCG solver proposed in [26], [29] relies on an alternating gradient based method which progressively adds Dirac masses. Recently, a variant of the ADCG called Sliding Frank-Wolfe (SFW) appeared in [30], which is guaranteed to converge in a finite number of steps under suitable assumptions.

C. This paper

The outline of this paper is the following. In Section II, we describe the observation model to invert, and we present how SFW can be used to analyze single mixtures in single images.

Then, we present our method to track shapes in images using sparse analysis in Section III. This method is based on the principles of SFW for individual object analysis, and extends it to manage the specificities of tracking: apparition and disappearance of objects, as well as splitting and merging.

Finally, Section IV presents the experimental results on real satellite images, as well as a numerical comparison with two state-of-the-art methods [6], [8].

In this paper, we track *objects* (MCS) represented by *mixtures*, so we use the two terms without distinction.

The mixtures contain *atoms* that are weighted *Dirac masses*, so here also the two terms will be used without distinctions. For convenience, we use in this paper the convention that high intensities in the images correspond to low surface brightness temperature, and images will be considered in their vector form. Vectors and multi-valued mathematical objects will be represented in bold.

II. Observation model and sparse shape analysis with SFW

A. Observation Model

We observe a sequence $\{\mathbf{y}_1, \dots, \mathbf{y}_T\}$ of T images, indexed by time. In our observation model, each image $\mathbf{y}_t$, $1 \leq t \leq T$, is built from an imaging operator applied to a measure:

$$\mathbf{y}_t = \mathbf{\Phi}\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t} + \boldsymbol{\epsilon_t} \tag{1}$$

$\boldsymbol{\epsilon_t}$ is an error term accounting for the model misfit. In other words, $\mathbf{y}_t$ is a mixture of several atoms. In this paper, we consider the case of 2D Gaussian ellipsoid, parametrized by position, scale, and eccentricity. Here and in the following, we consider for the measures μ a positive weighted Dirac mass sum of the form:

$$\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t} = \sum_{n=1}^{N} w_{t,n}\delta_{\boldsymbol{\theta}_{t,n}} \tag{2}$$

with the weight vector $\boldsymbol{w}_t = \{w_{t,1}, \dots, w_{t,N}\}$, such that $w_{t,n} > 0 \ \forall n \in \{1, \dots, N\}$[1]. We denote by $\mathcal{D}$ the bounded domain of interest for the parameter set, such that $\boldsymbol{\theta}_{t,n} \in \mathcal{D}$. The imaging operator $\mathbf{\Phi}$ is defined by $\mathbf{\Phi}\mu = \int_{\mathcal{D}} \phi \, \mathrm{d}\mu_{\boldsymbol{w},\boldsymbol{\theta}}$ where $\phi : \mathcal{D} \to \mathbb{R}$ is, in our case, a convolution by a 2D Gaussian ellipsoid kernel.

Gathering the Dirac masses into labeled object mixtures (the MCS), we may rewrite:

$$\mathbf{y}_t = \sum_{l=1}^{L} \mathbf{\Phi}\mu^{(l)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t} + \boldsymbol{\epsilon_t}. \tag{3}$$

where l denotes one label among the L objects appearing in the sequence.

Moreover, between $t = 1$ and $t = T$ an object l may appear (at $t_{l,\mathrm{start}}$) or disappear (at $t_{l,\mathrm{stop}}$). Accounting for these phenomena, the complete observation model is:

$$\mathbf{y}_t = \sum_{l=1}^{L} \mathbf{1}_{\{t_{l,\mathrm{start}} \leq t \leq t_{l,\mathrm{end}}\}} \mathbf{\Phi}\mu^{(l)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t} + \boldsymbol{\epsilon_t} \tag{4}$$

We also consider the following phenomena:

- *merging*: several mixtures, separated at time t, may be grouped at time $t + 1$.
- *splitting*: a mixture at time t may be split in several mixtures at time $t + 1$.
- *exchange*: atoms belonging to a label at time t may switch to another label at time $t + 1$.

The purpose of this paper is to invert the observation model (4) for the parameters of the measure, knowing only each $\mathbf{y}_t$ and $\mathbf{\Phi}$, while accounting for the split and merge phenomena. Let us notice, in particular, that for each time frame t, we do not know the number L of objects involved, nor the number of Dirac masses N in each mixture.

We seek the inversion of the observation model (4) within a continuous framework: the parameters are real-valued within the domain $\mathcal{D}$, which is a compact defined by lower and upper bounds for each parameter component.

B. Retrieving sparse shapes with Sliding Frank-Wolfe

Before handling the inversion of (4), we consider the following sub-problem: *how can we retrieve a single mixture from an image ?* This amounts to invert the simplified observation model (1) at a given time t, assuming there is only one object in the image.

To retrieve $\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ in a sparse fashion, we minimize among all non-negative measures μ the following criterion:

$$C(\mathbf{y}_t, \mu, \lambda) = \frac{1}{2} \left\| \mathbf{y}_t - \mathbf{\Phi}\mu \right\|_2^2 + \lambda\mu(\mathcal{D}) \tag{5}$$

where $\lambda > 0$ is the regularization parameter, and $\mu(\mathcal{D})$ denotes the total mass of the non-negative measure μ. In the case (2) of a sum of Dirac masses, $\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}(\mathcal{D}) = \sum_{n=1}^{N} w_{t,n}$. Such a minimization problem is challenging, since, contrary to standard ℓ_1 problems, not only the weights $\{w_{t,n}\}$ of the masses, but also their positions $\{\boldsymbol{\theta}_{t,n}\}$ in the parameter space are unknown, since they are not assumed to belong to a predefined a grid.

To optimize (5), we use the Sliding Frank Wolfe algorithm (SFW), a variant of [26], [29] proposed recently in [30]. This greedy algorithm iteratively adds new Dirac masses to the current set, then optimizes only the weights $\boldsymbol{w}_t$ in a classical LASSO setting, and then optimizes locally both $\boldsymbol{w}_t$ and $\boldsymbol{\theta}_t$ using a quasi-Newton method such as BFGS [31].

The key elements in this algorithm are the choice of new Dirac masses and the stopping criterion. Both rely on a certificate η written, at step k, as:

$$\eta_t^{(k)} \overset{\mathrm{def.}}{=} \frac{1}{\lambda}\mathbf{\Phi}^\top \left(\mathbf{y}_t - \mathbf{\Phi}\mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}} \right); \tag{6}$$

where $\boldsymbol{w}_t^{(k-1)}$ and $\boldsymbol{\theta}_t^{(k-1)}$ are estimated at the previous step. This certificate is a continuous function defined over the compact set $\mathcal{D}$. It is high-valued in the locations of $\mathbf{y}_t$ that are not well explained by $\mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}}$.

[1] Formally, in Eq. (2) N depends on the time frame t. Since there is no ambiguity when N is used, this dependence is not made explicit.

Algorithm 1 Sliding Frank-Wolfe

Input: $\mathbf{y}_t$, λ
Output: Minimizer $\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$
Initialization: $\mu_{\boldsymbol{w}_t^{(0)},\boldsymbol{\theta}_t^{(0)}} = 0$.
while $\max_{\mathcal{D}} \eta^{(k)} > 1$ (see Eq. (6)) :

1. Augment the support:
$$\boldsymbol{\theta}_t^{(k)} = \boldsymbol{\theta}_t^{(k-1)} \cup \arg\max_{\boldsymbol{\theta}\in\mathcal{D}} \eta^{(k)}$$
2. Adjust weights only (LASSO):
$$\tilde{\boldsymbol{w}}_t^{(k)} = \arg\min_{\boldsymbol{w}\in\mathbb{R}_+^k} C(\mathbf{y}_t, \mu_{\boldsymbol{w},\boldsymbol{\theta}_t^{(k)}}, \lambda)$$
3. Local variations on all parameters, using BFGS and starting at $\mu_{\tilde{\boldsymbol{w}}_t^{(k)},\boldsymbol{\theta}_t^{(k)}}$:
$$\boldsymbol{w}_t^{(k)}, \boldsymbol{\theta}_t^{(k)} = \arg\min_{\boldsymbol{w}\in\mathbb{R}_+^k, \boldsymbol{\theta}\in\mathcal{D}^k} C(\mathbf{y}_t, \mu_{\boldsymbol{w},\boldsymbol{\theta}}, \lambda)$$
4. Measure update: $\mu_{\boldsymbol{w}_t^{(k)},\boldsymbol{\theta}_t^{(k)}} = \sum_{n=1}^k w_{t,n}^{(k)} \delta_{\boldsymbol{\theta}_{t,n}^{(k)}}$

Besides, $\mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}}$ is a solution to (5) if and only if $\max_{\mathcal{D}} \eta^{(k)} \leq 1$. Indeed, that condition is equivalent to standard optimality conditions for constrained convex problems, as explained (informally) in Appendix A. Hence, this condition is a natural stopping criterion for the SFW algorithm.

The SFW procedure is described in Algorithm 1, and Fig. 2 provides a stepwise illustration of the algorithm. Details on its implementation are given in Appendix B. We refer the reader to [30] for further detail on the SFW algorithm.

III. Proposed method

In this section, we introduce our method to track several deforming shapes in image sequences, based on the sparse image decomposition presented in the previous section. Our approach requires to consider several aspects of the problem, listed below and detailed in the next subsections:

- *Initialization* refers to the estimation of the mixtures number and parameters in the initial image $\mathbf{y}_0$.
- *Propagation* is the task of recursive mixture estimation, from time $t-1$ to time t.
- *Creation* manages the apparition of mixtures at time t that did not exist at time $t-1$.
- *Merging* and *splitting* handle the interaction between mixtures over time.

A. Initialization

At this step, the problem is the retrieval of several shapes in the image $\mathbf{y}_0$. This is the conjunction of two problems:

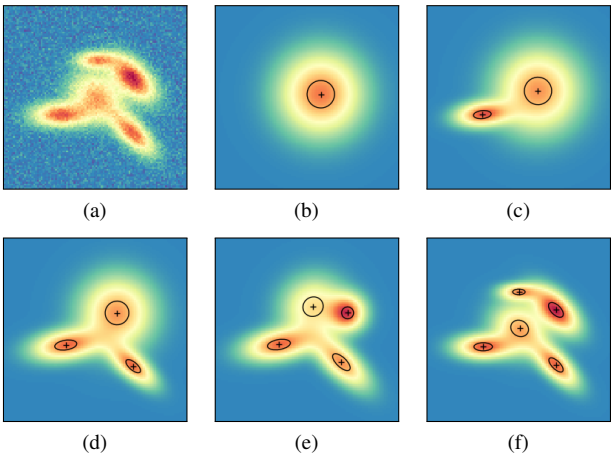


Fig. 2: Illustration of SFW steps on a synthetic image (cf. Alg. 1). (a) synthetic image $\mathbf{y}_t$ generated with white Gaussian noise, and 5 atoms using a 2D ellipsoid kernel ϕ. (b)–(d) images generated with $\Phi\mu_{\boldsymbol{w}_t^{(k)},\boldsymbol{\theta}_t^{(k)}}$ from $k=1$ to $k=5$, on which the atom are superimposed to illustrate the parameters (positions, scales, orientation).

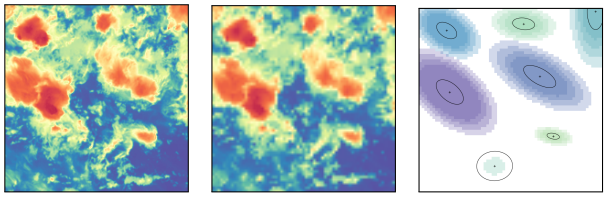
1) finding the number of objects,
2) estimating the parameter of each objects.

These problems are handled through a two-step SFW procedure. The first step consists in applying SFW to an under-resolved version of $\mathbf{y}_0$, to produce an estimation of the number of mixtures as well as a coarse estimation of their parameters. This step yields several atoms, that are for now considered as separate objects: they may be merged or split later (see Subsection III-D).

The second step processes each of the resulting atoms separately. It consists in applying SFW using the fully resolved image, in order to obtain several atoms per object. We observed that this procedure provides an accurate estimate of mixture number, while being faster than applying SFW on the fully-resolved image. The initialization procedure is summarized in Alg. 2 and its first step is illustrated in Fig. 3.

Algorithm 2 Initialization: coarse-to-fine SFW

Input: Image $\mathbf{y}_0$
Output: Estimation $\hat{\mu}_{\boldsymbol{w}_0,\boldsymbol{\theta}_0}$
Step 1.
Let $\tilde{\mathbf{y}}_0$ be the under-resolved version of $\mathbf{y}_0$.
Apply SFW (Alg. 1) to $\tilde{\mathbf{y}}_0$.
Step 2.
for each resulting Dirac mass :
Apply SFW on $\mathbf{y}_0$ starting at the coarse estimation from step 1.



(a) $\mathbf{y}_0$ (b) Under-resolved $\tilde{\mathbf{y}}_0$ (c) SFW on $\tilde{\mathbf{y}}_0$.

Fig. 3: Illustration of step 1 of the initialization step. In (c), labels are represented by color and the opacity corresponds to brightness temperature thresholds (see Fig. 6).

B. Propagation and disappearance

At this step, we handle the problem of retrieving, at time t, a mixture l that already existed at time $t-1$. This amounts to finding $\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ minimizing $C(\mathbf{y}_t, \mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}, \lambda)$, knowing already $\hat{\mu}_{\boldsymbol{w}_{t-1},\boldsymbol{\theta}_{t-1}}$. We assume that the variation in mixture number and parameters values are small between $t-1$ and t. Hence, the propagation step consists in applying SFW to $\mathbf{y}_t$, using $\hat{\mu}_{\boldsymbol{w}_{t-1},\boldsymbol{\theta}_{t-1}}$ as an initial guess. This starting measure is modified as follow:

- we need to allow for pruning, to handle decreasing number of atoms in mixtures. This is done by a LASSO optimization on the weights only, followed by a removal of the zero-valued outputs.
- we account for the average trend of mixtures in $\mathcal{D}$, as the difference of the parameter averages between $t-1$ and $t-2$. This corresponds to speed in the parameter space $\mathcal{D}$.

After this step, an object may have no Dirac mass left. In this case, the object is assumed to have disappeared, so an estimation of $t_{l,\text{end}}$ is known and the object is not proceed further. The propagation step is summarized in Alg. 3.

Algorithm 3 Propagation using SFW

Input: Estimation $\hat{\mu}_{\boldsymbol{w}_{t-1},\boldsymbol{\theta}_{t-1}}$, $\mathbf{y}_t$
Output: Estimation $\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$
if $t > 2$:
  Estimate the trend as $\hat{\boldsymbol{w}}_s = \overline{\boldsymbol{w}}_{t-1} - \overline{\boldsymbol{w}}_{t-2}$ and $\hat{\boldsymbol{\theta}}_s = \overline{\boldsymbol{\theta}}_{t-1} - \overline{\boldsymbol{\theta}}_{t-2}$.
  Initial guess: $\hat{\mu}_{\boldsymbol{w}_{t-1}+\hat{\boldsymbol{w}}_s,\boldsymbol{\theta}_{t-1}+\hat{\boldsymbol{\theta}}_s}$.
else:
  Initial guess: $\hat{\mu}_{\boldsymbol{w}_{t-1},\boldsymbol{\theta}_{t-1}}$.
Pruning, adjusting weights only (LASSO).
Run the SFW iterations.

C. Birth

We address the problem of retrieving, at a given time t, mixtures that did *not* exist at time $t-1$. The goal is to detect where new mixtures appear, and how many. This problem is close to the problem handled at initialization, excepting that some parts of the image are already explained by mixtures.

Let us consider the current residual, noted $\mathbf{r}_t$:

$$\mathbf{r}_t = \mathbf{y}_t - \boldsymbol{\Phi}\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t} \tag{7}$$

This residual is obtained with the current estimator, *i.e.* after the propagation step. Reusing the methods from the initialization step, the creation of new object at time t is done by the application of Alg. 2 to the residual $\mathbf{r}_t$.

D. Merging mixtures

When considering the merging of two mixtures, we have to consider the following cases:

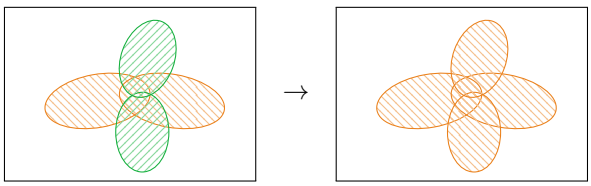
- two or more mixtures overlap entirely: they both describe the same phenomenon and must be merged under one label;
- two or more mixtures overlap partly: some Dirac masses in these mixtures describe the same phenomena, and they must be moved under the same label.

An illustration of these two cases is given in Fig. 4.

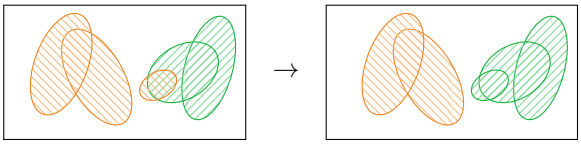
The merging procedure we propose account for the two situations. We measure correlation coefficients between mixture images $\boldsymbol{\Phi}\mu^{(l)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$, corresponding to each label l at the fixed time t.

Let l_1 be a given "reference" mixture, and l_2 an other mixture that may be merged:

- if $\boldsymbol{\Phi}\mu^{(l_1)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ and $\boldsymbol{\Phi}\mu^{(l_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ are too much correlated, they are merged.
- if the image of a Dirac mass $\mu^{(l_2,n_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ from the mixture l_2 is too much correlated to the image of



(a) Mixture merging.



(b) Atom transfer.

Fig. 4: Illustration of the two cases considered for merging. In case (a) the two mixtures are merged, while in case (b) the smaller atom is merged to the right-hand mixture and removed from the left-hand mixture.

another Dirac mass $\mu^{(l_1,n_1)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ from l_1, it is merged to l_1 and removed from l_2.

This approach requires the definition of a correlation threshold C_{thr}, above which the merging is performed. In practice, this threshold defines the tolerance to overlap in the resulting image. Alg. 4 summarizes the merging step.

Algorithm 4 Correlation-based merging

Input: Correlation threshold C_{thr}, current measure $\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$

Output: Merging of correlated mixtures.

Let $\mathcal{L}$ be the mixture list, sorted by total mass in decreasing order, and $\mathcal{R}$ be the reference mixture list, empty at start.

for $l_1 \in \mathcal{L}$: ▷ *reference mixture*

  Append l_1 to $\mathcal{R}$.

  for $l_2 \in \mathcal{L}$ with $l_2 \notin \mathcal{R}$: ▷ *other mixture*

    Step 1. Single atom merging or transfer

    if $N_2 = 1$:

      $\forall n_1 \in \{1,\dots,N_1\}$, compute:

      $c_{n_1,l_2} = \text{Corr}(\boldsymbol{\Phi}\mu^{(l_1,n_1)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}, \boldsymbol{\Phi}\mu^{(l_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t})$

      if $\exists n_1 \in \{1,\dots,N_1\}$ s.t. $c_{n_1,l_2} > C_{\text{thr}}$:

        Merge l_1 and l_2. The label of the mixture with the larger total mass is maintained.

    else:

      for $n_2 \in \{1,\dots,N_2\}$: ▷ *within* l_2

        Let $\mu^{(l_2,\setminus n_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ be $\mu^{(l_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ without the n_2 atom.

        Compute $\begin{cases} c_{n_2,l_1} = \text{Corr}(\boldsymbol{\Phi}\mu^{(l_2,n_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}, \boldsymbol{\Phi}\mu^{(l_1)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}) \\ c_{n_2,l_2} = \text{Corr}(\boldsymbol{\Phi}\mu^{(l_2,n_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}, \boldsymbol{\Phi}\mu^{(l_2,\setminus n_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}) \end{cases}$

        if $c_{n_2,l_1} > c_{n_2,l_2}$:

          Move $\mu^{(l_2,n_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ to the mixture l_1.

    Step 2. Mixture merging.

    if $\text{Corr}(\boldsymbol{\Phi}\mu^{(l_1)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}, \boldsymbol{\Phi}\mu^{(l_2)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}) > C_{\text{thr}}$:

      Merge l_1 and l_2. The label of the mixture with the larger total mass is maintained.

E. Splitting

At a given time t, a mixture may also be split into several parts. This situation arises, within a mixture, when two or more components are too much separated from one another, i.e. they do not overlap enough to represent the same object anymore.

To quantify this separation, we evaluate the correlation matrix of the mixture Dirac masses' images. This matrix is thresholded using C_{thr}, below which the Dirac masses are considered uncorrelated. The resulting binary matrix is used to build a undirected, unweighted graph Γ where nodes represent Dirac masses and edges represent correlations above the threshold. If the graph Γ has several disconnected component, then the mixture must be split along these separated components. The procedure is illustrated in Fig. 5 and reported in Alg. 5.

Let us remark that this step relies on the same principles as the merging step, so that using both in long time series allow the estimator to be balanced, *i.e.* there is no over-splitting or over-merging.

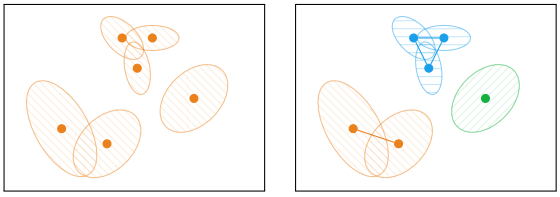


Fig. 5: Illustration of the splitting procedure. Left: a single mixture containing 6 atoms. Right: graph of correlated atoms, colored by disconnected components. The corresponding atoms groups form the new mixtures.

Algorithm 5 Splitting

Input: Correlation threshold C_{thr}, current measure $\mu_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$

Output: Splitting of non-correlated Dirac masses within mixtures.

for Each mixture $l \in \{1,\dots,L\}$:

  Compute the correlation matrix $\mathbf{C}$ with $\boldsymbol{\Phi}\mu^{(l,n)}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}, 1 \leq n \leq N$.

  Threshold this matrix with C_{thr}.

  Build the graph Γ from $\mathbf{C}$.

  if Γ has more than one connected component :

    Split l along the connected components.

F. Summing up

Before gathering the entire algorithm procedure, we need to introduce an additional local adjustment step. Its purpose is to provide further flexibility after the split and merge steps, so that newly formed mixtures can be adapted to the image. To do so, we process each mixture with the SFW algorithm, using the current mixtures values as initial guess.

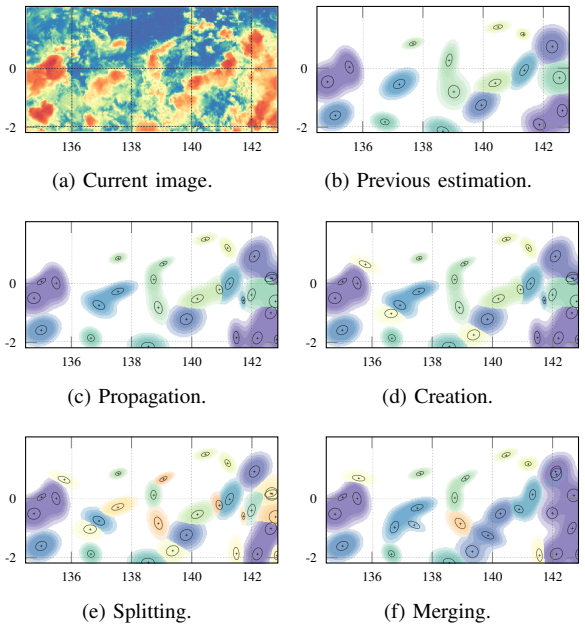
The method we propose is based on the four step described previously, and is called SAST for Sparse Analysis for Shape Tracking. The algorithm iteratively provides estimators $\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ from $t = 1$ to $t = T$, with, at each step, a recursive adaptation of the estimator. It is described in Alg. 6 and a step-by-step illustration of the algorithm is provided in Fig. 6.

IV. Results

This section presents the results of the SAST method for the tracking of MCS in infrared satellite images,

Algorithm 6 SAST algorithm

Input: Image sequence, correlation threshold C_{thr}, regularization parameter λ
Output: Estimator $\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$, $t \in \{1,\dots,T\}$
for t = 1 to t = T :
  if t = 1 :
    Initialize: Alg. 2 applied to $\mathbf{y}_0$
  else:
    Propagate: Alg. 3 with $\hat{\mu}_{\boldsymbol{w}_{t-1},\boldsymbol{\theta}_{t-1}}$
    Create: Alg. 2 applied to $\mathbf{r}_t = \mathbf{y}_t - \boldsymbol{\Phi}\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$
  Split: Alg. 5 using C_{thr}.
  Merge: Alg. 4 using C_{thr}.
  Local variations: apply SFW (Alg. 1) on the modified mixtures starting from their current values.
  Save the current value of $\hat{\mu}_{\boldsymbol{w}_t,\boldsymbol{\theta}_t}$ as a result.



(a) Current image. (b) Previous estimation.
(c) Propagation. (d) Creation.
(e) Splitting. (f) Merging.

Fig. 6: Step-by-step illustration of the SAST algorithm. (a) has the same intensity range as in Fig. 1 and (b)-(f) represents outputs by color (labels) with an opacity defined by the set (235K, 245K, 255K, 265K) of brightness temperature thresholds. Individual atoms are represented in overlay.

as well as a comparison with the two state-of-the-art methods proposed in [6] and [8].

The objective is to automatically produce informations on the MCS dynamic to help understand their behavior, which is not fully known. Hence, creating realistic synthetic data on dynamic MCS is far from trivial. Since no ground truth is available, we experiment the three methods on real Himawari [4] images, and then compare them qualitatively and numerically.

A. Experimental settings

1) SAST algorithm: The SAST algorithm requires several choices regarding its implementation. The first one is the choice of the convolution kernel, which will describe the shape of individual atoms forming mixtures. As mentioned previously, we choose to describe MCS as mixtures of 2D Gaussian, since they do not have clear contour in satellite image. The parameter of each Gaussian are the position (x, y), the big axis a, the eccentricity e and the orientation α. Hence, given the time t, a label l, and an atom index n, we have $\boldsymbol{\theta}_{t,n}^{(l)} = \{x, y, a, e, \alpha\}_{t,n}^{(l)}$. This parametrization allows to directly control ellipsoid eccentricity in order to avoid degenerate cases.

$\mathcal{D}$ contains each parameters set $\boldsymbol{\theta}_{t,n}^{(l)}$, so its bounds must be specified.We choose to set the position bounds at the image border, and α between $-\pi$ and π. In order to avoid degenerate cases, we set the big axis between 2 and 20 pixels and the eccentricity between 0 and 0.9. to avoid degenerate cases. These choices were made to provide fast, yet efficient, results on the data presented hereafter. The output of the algorithm is only marginally affected by changes in these settings.

Two parameters affects the overall behavior of the SAST algorithm:

- the regularization parameter λ sets the sparsity compromise in (5). A larger λ yields fewer atoms, so the description of the clouds will be coarser and obtained more quickly. Conversely, a smaller λ will provide a finer fit to the image.
- the correlation threshold C_{thr} sets the limit for both splitting and merging. A larger C_{thr} implies a larger number of small mixtures, while a smaller value yields larger and fewer mixtures.

These two parameters must be set in accordance with the purpose of the analysis: fine *vs.* coarse, and atomic *vs.* grouped analysis respectively. We observed that a coarser analysis is relevant when used in conjunction with a more grouped aggregation, and conversely. In consequence, we choose to study the three following cases :

A. $\lambda = 1.25$ and $C_{\text{thr}} = 0.5$,
B. $\lambda = 2$ and $C_{\text{thr}} = 0.25$,
C. $\lambda = 3$ and $C_{\text{thr}} = 0.125$.

2) State-of-the-art methods: In the literature, two recent algorithms address the problem of tracking MCS in time sequences of images. The first one, coined TOOCAN [6], tracks binary shapes based on threshold, surface, and connectivity in the spatio-temporal domain. TOOCAN produces discrete-labeled map of pixels, so there is no information regarding shape or intensity.

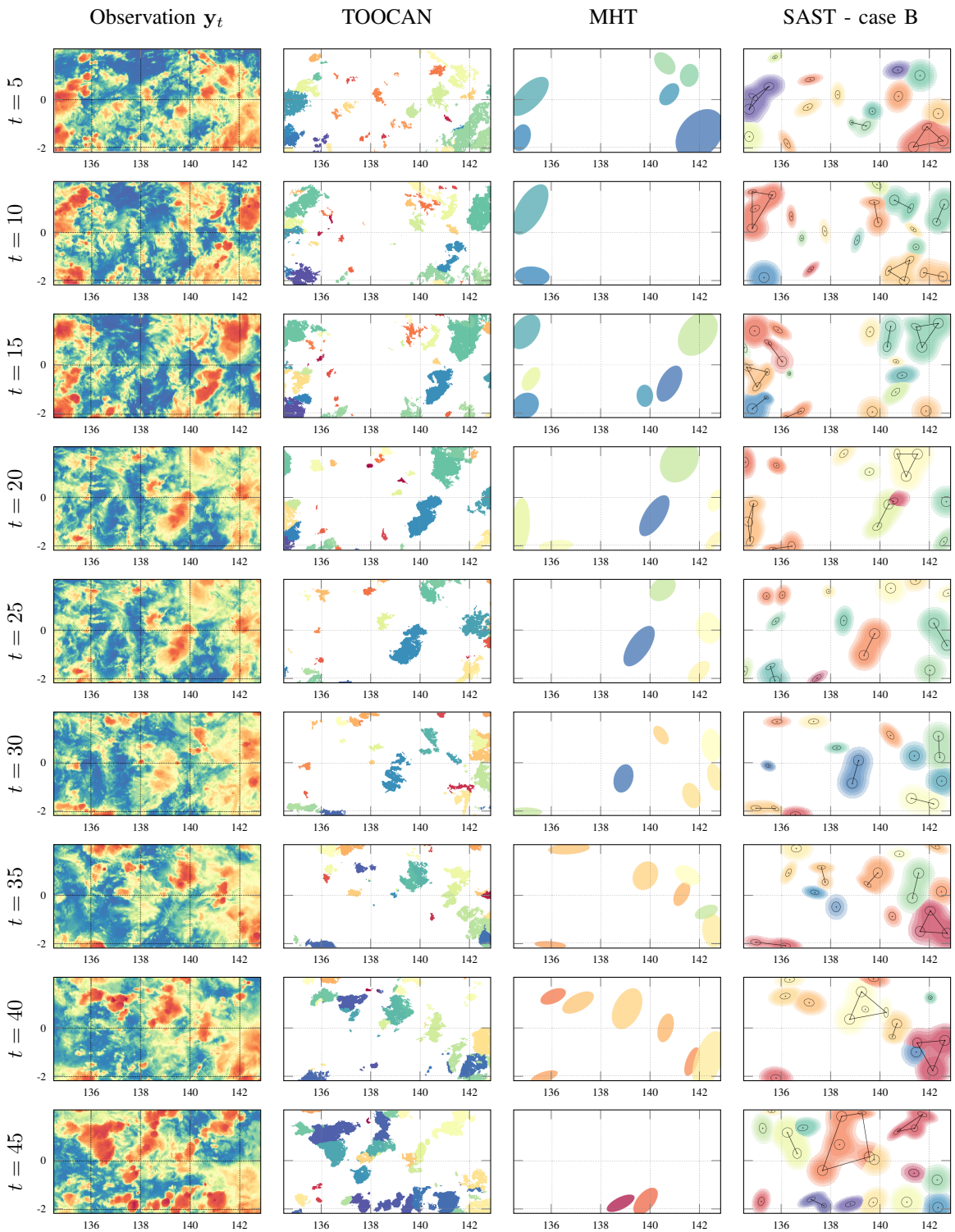


Fig. 7: First column: sequence of images from Himawari taken August 10, 2017 from 01:40 to 15:20 UTC. Second column: results with the TOOCAN algorithm [6], each color representing a label. Third column: results with the MHT method [8], represented as labeled ellipses. Fourth column: results with the SAST method in case B, using the same representation as in Fig. 6. Mixtures are in addition delimited by the convex hull of their atom's positions.

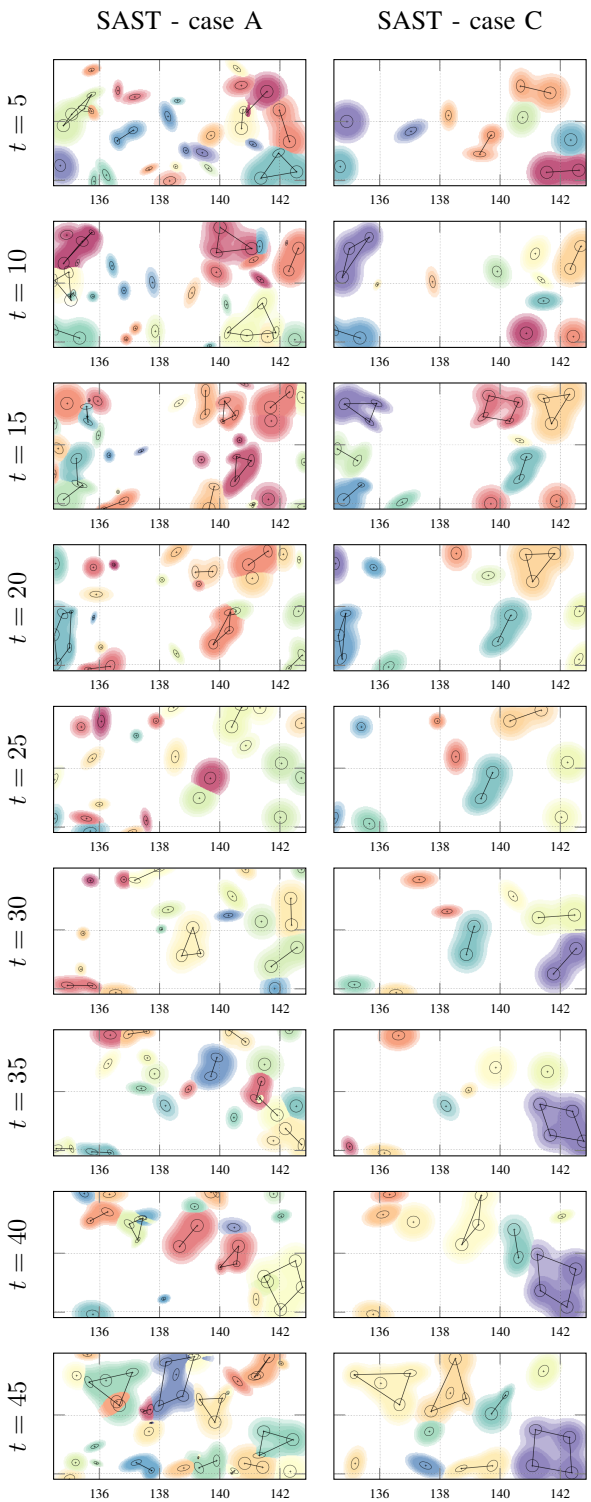


Fig. 7 (Cont.): Output of the SAST method for cases A (left) and C (right).

We used our own implementation of TOOCAN for the comparison.

The second method relies on a recursive Bayesian tracking framework, called MHT [8]. This method provides parametric estimation of ellipses, each one describing the shape of a cloud or cloud system. However, the tracking is not applied directly on images: instead, a binarized image (threshold at 233K) is produced, from which small surface regions are removed. The remaining contours are fit by ellipses, which form the observations considered by the model. Hence, the MHT method works in parametric contour space, without considering the images intensity. For the comparison, we used the implementation of MHT provided by the authors.

Here, we present comparative results on a real image sequence acquired by the Himawari satellite [4], which has a $2 \times 2\,\mathrm{km}$ Sub-Satellite Point (SSP) resolution. Since both TOOCAN and MHT were applied initially applied to MSG-1 [32] infrared images ($3 \times 3\,\mathrm{km}$ SSP resolution), the parameters of both methods were adapted to the Himawari setting.

B. Comparative results

1) Qualitative results: In Fig. 7, we illustrate the three methods on a Himawari sub-satellite image in the infrared channel at 10.4 μm covering *ca.* $800 \times 400\,\mathrm{km}^2$ over 50 time frames taken each 20 minutes. Several comments can be made regarding these results:

- the TOOCAN algorithm yields very fine details, as its analysis is performed at the pixel level. However, resulting shapes are sometimes questionable especially where the image intensity smoothly varies near the 235 K threshold value. Besides, TOOCAN may separate features that could be analysed as propagation of MCS, for instance at $t = 45$ near (136°E, 1°N).
- unlike TOOCAN, the MHT method describes coarser details in the images. While it does capture the overall displacements, many informations are missing: some objects undetected for some time period, such as the blue ellipse moving from 142°E at frame 5 to 139°E at frame 30. In consequence, sometimes large parts of the original image were missed as in frame 45.
- unlike TOOCAN and MHT, our method handles a large scope of temperatures, providing more complete object description. Despite being less precise than TOOCAN regarding the pixel-wise description, the output of SAST yields explicitly shape parameters that can be used for the physical processes analysis (beyond the scope of this work).

The three cases studied for SAST provide additional insights:

- In case A, $\lambda = 1.25$ yields the finer details: SAST fits very well the image and 342 labels are found in the sequence of frames. Because C_{thr} is high, splitting occurs frequently and there is some overlap between mixtures.
- Alternatively, case C provides a coarse estimation of the mixtures because $\lambda = 3$. On the other hand, because C_{thr} is low, the merging are frequent. In consequence, 55 labels were found, and we observe that they are well preserved across the time sequence.
- Case B is an intermediate stage between cases A and C. It provides finer details than case C, but also less overlap between mixtures than in case A. In this case, 141 mixtures were found.

More generally, we observed that for a fixed λ, diminishing C_{thr} yields larger mixtures while, when C_{thr} is fixed, mixtures become more isolated when λ is augmented. Hence C_{thr} and λ cannot be chosen independently.

2) Quantitative results: As the three methods tested in the experiments differ in the type of results, quantitatively comparing them is not trivial. The common ground between the outputs comes down to positions, scales, and labels.

Besides, the outputs differ in label number: on the sequence described above, MHT fount 32 labels, TOOCAN found 123 labels and SAST found 55 to 342 labels. In addition, the time scales of the output results vary between methods, and within each output. In consequence, the outputs cannot be paired for comparison. Here, we bypass these concerns by comparing the empirical distributions of the time-scaled values of the objects detected over 3 time frames or more.

In Fig. 8, we report the measure for relative variations for longitude and latitude of the objects, as well as their surface:

- The measurements provided by MHT seem more scattered than for other methods. This is partly due to the low number of labels (32), but also to the method itself, as the 44 remaining (existing over 3 or more frames) labels of SAST, case C induce a lower scattering.
- Regarding the position parameters, the three methods agree on the general displacement. Vertical (latitude) displacement is small and balanced, while horizontal (longitude) displacements are all similarly shifted, depicting well the east-to-west propagation of clouds visible in Fig. 7.
- The size parameters have also the same magnitude, despite the fact that they are performed on different objects.

While the methods agree on the magnitude of the phenomena, their apparent behavior differs notably. Especially, the size measurements highlight an interpretation issue regarding the life cycle of a MCS: when a new updraft rises in the vicinity of an existing MCS, should it be considered as the same MCS propagating, or as a new MCS ?

In the first case, many objects have a short lifetime and disappear in a split or merge event, so the MCS size rarely decreases. This case corresponds to the measures provided by MHT, and SAST in cases A and B.

In the second case, there is less short-live objects, but MCS do not survive by in merging of new objects, hence their size decreases and they disappear. This corresponds to the TOOCAN output and with SAST in case C. Note that in this case, the latitude displacement are more important since lifetimes are longer.

In addition, the SAST method provides more information on the MCS, as exemplified by the peak temperature measurements in the fourth line of Fig. 8. In the three case, it can be seen that at the beginning of an object's life, on average temperatures first diminish and then rise. This is in accordance with the physics of the MCS, since the updraft first rises at high altitude at low temperatures, and then slowly disaggregate while diminishing their altitude, forming a large anvil and raising its temperature [33]. Here also, the amplitude of temperature variations varies because of differences of objects' lifetime.

Summing up, the SAST method agrees with TOOCAN and MHT on the overall measurements, while providing flexibility regarding the MHT lifecycle interpretation, and while providing more data than its counterparts. By varying the parameter λ, we can analyze the images and the formation of aggregates at various scales.

3) Computation time: The implementation of SAST was made in Python 3.5, and will be made available online[2]. This implementation was not specifically optimized for speed. On a work station with 12 Intel i7-8700 3.2GHz CPUs and a 12 Gb RAM, the 400×200 px $\times$ 50 frame images depicted in Fig. 7 were processed in 20 min, 10 min, and 6 min for cases A, B and C respectively. This time is mainly used for the large number of executions of the SFW algorithm. We believe that a C/C++ implementation, coupled with GPU computation, should significantly reduce this time. For comparison, our implementation of TOOCAN using Python/Scipy took 19 min to process the same dataset and MHT took 2 min to process the same sequence because it handles parameter sets instead of image sequences.

V. Conclusion

In this paper, we propose a novel method, named SAST, for the dynamic analysis of MCS in time sequences of images. This method successfully achieves the image

[2]Repository web address: `github.com/courbot/sast`

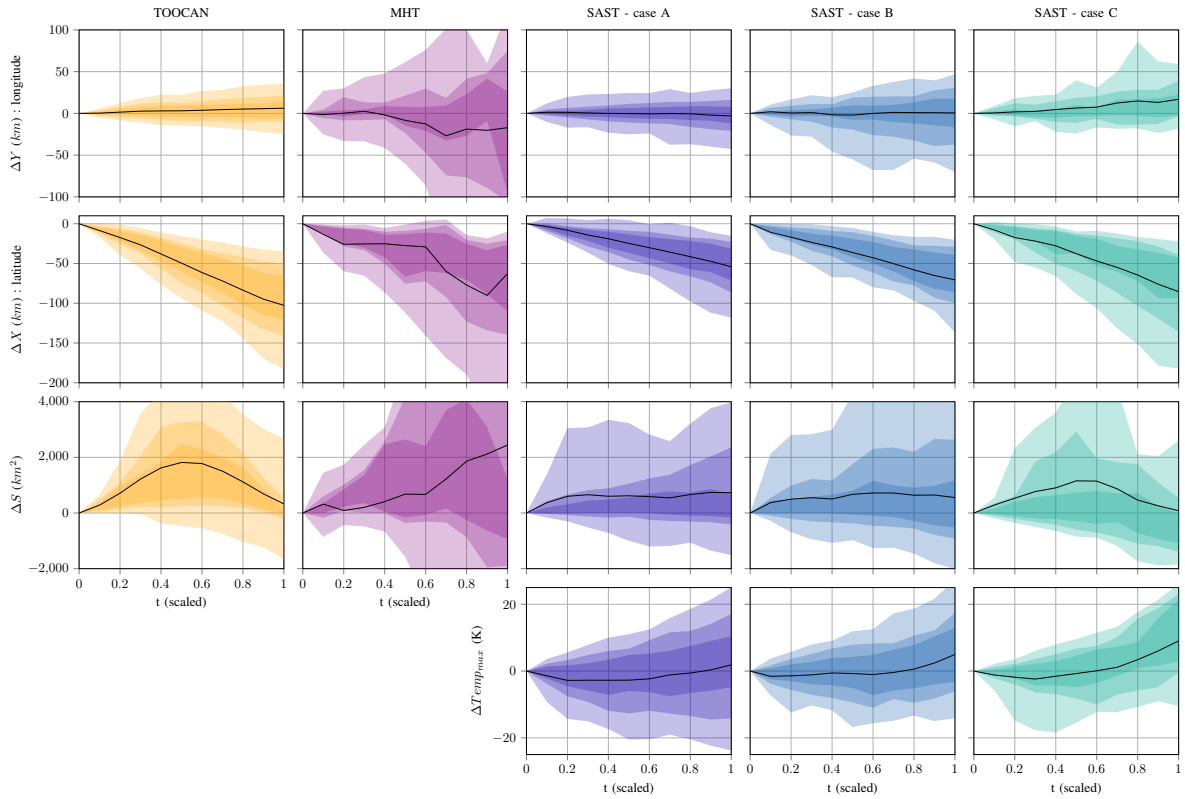


Fig. 8: Time-scaled, relative variations of positions (first and second line), surface (third line) and temperature (fourth line) for the three evaluated methods. The average measure is in black, and the three first and last deciles are represented in color. Note that neither TOOCAN nor MHT methods provide the temperature measurement.

fitting through a recent off-the-grid sparse analysis algorithm, and manages as well the apparition, disappearance, splitting and merging events. On real data, we show that the results obtained by SAST are consistent with the other state-of-the-art methods. Besides, we showed that SAST provides additional information and flexibility with respect to the existing methods.

Several perspectives stem from this work, among which the the development of multi-spectral approaches and the implementation for large-scale applications.

Acknowledgements

The authors would like to thank Alexandros Makris (Computational Vision and Robotics Laboratory, FORTH, Greece) for kindly providing the code for MHT tracking. They would also like to thank Jean-David Benamou (Inria Paris, France), François-Xavier Vialard (Université Paris-Est Marne-la-Vallée, France), Caroline Muller, Genevieve Seze (LMD, UMR 8539, Paris, France) and Pasquale Selitto (LISA, CNRS UMR 7583, Paris, France) for fruitful discussions regarding MCS tracking.

References

[1] R. A. Houze, "Mesoscale convective systems," *Reviews of Geophysics*, vol. 42, no. 4, 2004.

[2] D. Bouniol, R. Roca, T. Fiolleau, and D. E. Poan, "Macrophysical, Microphysical, and Radiative Properties of Tropical Mesoscale Convective Systems over Their Life Cycle," *Journal of Climate*, vol. 29, no. 9, pp. 3353–3371, May 2016.

[3] C. E. Holloway, A. A. Wing, S. Bony, C. Muller, H. Masunaga, T. S. L'Ecuyer, D. D. Turner, and P. Zuidema, "Observing Convective Aggregation," *Surveys in Geophysics*, Jun. 2017.

[4] K. Bessho, K. Date, M. Hayashi, A. Ikeda, T. Imai, H. Inoue, Y. Kumagai, T. Miyakawa, H. Murata, T. Ohno *et al.*, "An introduction to himawari-8/9 – japan's new-generation geostationary meteorological satellites," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 94, no. 2, pp. 151–183, 2016.

[5] V. Mathon and H. Laurent, "Life cycle of Sahelian mesoscale convective cloud systems," *Quarterly Journal of the Royal Meteorological Society*, vol. 127, no. 572, pp. 377–406, 2001.

[6] T. Fiolleau and R. Roca, "An algorithm for the detection and tracking of tropical mesoscale convective systems using infrared images from geostationary satellite," *IEEE transactions on Geoscience and Remote Sensing*, vol. 51, no. 7, pp. 4302–4315, 2013.

[7] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

[8] A. Makris and C. Prieur, "Bayesian multiple-hypothesis tracking of merging and splitting targets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 12, pp. 7684–7694, 2014.

[9] D. Reid *et al.*, "An algorithm for tracking multiple targets," *IEEE transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979.

[10] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004.

[11] R. P. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[12] ——, "PHD filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, 2007.

[13] C. K. Chui, G. Chen *et al.*, *Kalman filtering*. Springer, 2017.

[14] R. P. Mahler, *Statistical multisource-multitarget information fusion*. Artech House, Inc., 2007.

[15] B.-T. Vo, B.-N. Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 409–423, 2009.

[16] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, "The labeled multi-Bernoulli filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3246–3260, 2014.

[17] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5129–5141, 2010.

[18] N. Vaswani, Y. Rathi, A. Yezzi, and A. Tannenbaum, "Deform PF-MT: particle filter with mode tracker for tracking nonaffine contour deformations," *IEEE Transactions on Image Processing*, vol. 19, no. 4, pp. 841–857, 2010.

[19] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: algorithms and applications," *IEEE access*, vol. 3, pp. 490–530, 2015.

[20] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[21] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[22] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[23] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE transactions on signal processing*, vol. 59, no. 10, pp. 4735–4744, 2011.

[24] Y. De Castro and F. Gamboa, "Exact reconstruction using Beurling minimal extrapolation," *Journal of Mathematical Analysis and applications*, vol. 395, no. 1, pp. 336–354, 2012.

[25] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[26] K. Bredies and H. K. Pikkarainen, "Inverse problems in spaces of measures," *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 19, no. 1, pp. 190–218, 2013.

[27] J.-M. Azais, Y. De Castro, and F. Gamboa, "Spike detection from inaccurate samplings," *Applied and Computational Harmonic Analysis*, vol. 38, no. 2, pp. 177–195, 2015.

[28] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE transactions on information theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[29] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," *SIAM Journal on Optimization*, vol. 27, no. 2, pp. 616–639, 2017.

[30] Q. Denoyelle, V. Duval, G. Peyré, and E. Soubies, "The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy," Nov. 2018, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01921604

[31] D. F. Shanno, "Conditioning of quasi-Newton methods for function minimization," *Mathematics of computation*, vol. 24, no. 111, pp. 647–656, 1970.

[32] J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, "An Introduction to Meteosat Second Generation (MSG)," *Bulletin of the American Meteorological Society*, vol. 83, no. 7, pp. 977–992, Jul. 2002.

[33] Z. Luo, G. Y. Liu, and G. L. Stephens, "CloudSat adding new insight into tropical penetrating convection," *Geophysical Research Letters*, vol. 35, no. 19, pp. L19 819–L19 819, Oct. 2008.

[34] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

APPENDIX A
STOPPING CRITERION AND OPTIMALITY CONDITIONS

In this Appendix, we explain informally (ignoring the functional analysis issues raised by the infinite-dimensionality of the space of measures) how the condition $\max_{\mathcal{D}} \eta^{(k)} \leq 1$ in Alg. 1 is equivalent to $\mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}}$ minimizing (5) under non-negativity constraints.

For the sake of simplicity, we drop the dependency on $\mathbf{y}_t$ and λ, and we write that minimization problem as

$$\min_{\mu \in K} C(\mu) \tag{8}$$

where K is the closed convex set of non-negative measures, and $C(\mu) := C(\mathbf{y}_t, \mu, \lambda)$. The optimality condition for that problem is that $\nabla C(\mu)$ is normal to K, pointing "outwards", *i.e.*

$$\langle \nabla C(\mu), \nu - \mu \rangle \leq 0, \tag{9}$$

for all $\nu \in K$. When μ is of the form $\sum_{n=1}^{k-1} w_n \delta_{\theta_n}$, choosing ν of the form $\alpha \delta_\theta$ and letting $\alpha \to +\infty$, then choosing ν of the form $\mu \pm \varepsilon \delta_{\theta_n}$ for $\varepsilon > 0$ small enough, we see that (9) implies (and is in fact equivalent to)

$$\forall \theta \in \mathcal{D}, \quad \nabla C(\mu)(\theta) \leq 0, \tag{10}$$

$$\forall n \in \{1, \dots, k-1\}, \quad \nabla C(\mu)(\theta_n) = 0. \tag{11}$$

In other words, $\nabla C(\mu)$ is a non-negative continuous function which vanishes at every θ_n.

Now, from (5), one may observe that

$$\nabla C(\mu) = \boldsymbol{\Phi}^\top (\boldsymbol{\Phi}\mu - \mathbf{y}_t) + \lambda \tag{12}$$

so that $\nabla C(\mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}})(\theta) = \lambda \left(-\eta^{(k)}(\theta) + 1\right)$.

It can be checked that the optimality condition of Step 4 in Alg. 1 implies that $\eta^{(k)}(\boldsymbol{\theta}_{t,n}^{(k-1)}) = 1$ holds for every n at every iteration k. As a result, Eq. (11) always holds for $\mu = \mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}}$. Eventually, the condition $\max_{\mathcal{D}} \eta^{(k)} \leq 1$ is equivalent to (10), hence to the optimality of $\mu_{\boldsymbol{w}_t^{(k-1)},\boldsymbol{\theta}_t^{(k-1)}}$.

APPENDIX B
IMPLEMENTATION OF THE SFW ALGORITHM

In this Appendix, we provide several implementation details for the SFW algorithm detailed in Alg. 1.

Step 1: computation of $\eta^{(k)}$. This certificate is valued on the compact $\mathcal{D}$. In practice, we grid $\mathcal{D}$ and evaluate $\eta^{(k)}$ in each grid node.

Stop condition: searching for $\max_{\mathcal{D}}(\eta^{(k)})$. This values is approached, because of the previous griding This does not prevent the algorithm to perform correctly, as the sub-grid errors are compensated in the gridless step 4 (see also [26]).

Step 3: weight adjusting. At this step, we allow $\boldsymbol{w}$ to vary but not $\boldsymbol{\theta}$. This amounts to a classical LASSO problem with a fixed dictionary, and we solve it with coordinate descent.

Step 4. Local variations. As suggested in [30], we use a limited-memory bounded BFGS algorithm [34] to restrict the search space to $\mathcal{D}$.