



# **A pattern-based method for handling confidence measures while mining satellite displacement field time series. Application to Greenland ice sheet and Alpine glaciers**

Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier,  
Emmanuel Trouvé, Noel Gourmelen, Jean-Louis Mugnier

## **► To cite this version:**

Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Emmanuel Trouvé, et al.. A pattern-based method for handling confidence measures while mining satellite displacement field time series. Application to Greenland ice sheet and Alpine glaciers. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11 (11), pp.4390 - 4402. 10.1109/JS-TARS.2018.2874499 . hal-01912708

**HAL Id: hal-01912708**

**<https://hal.science/hal-01912708>**

Submitted on 6 Nov 2018

**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.

# A pattern-based method for handling confidence measures while mining satellite displacement field time series. Application to Greenland ice sheet and Alpine glaciers

Tuan Nguyen<sup>1</sup>, Nicolas Méger<sup>1</sup>, Christophe Rigotti<sup>2</sup>, Catherine Pothier<sup>3</sup>, Emmanuel Trouvé<sup>1</sup>, Noël Gourmelen<sup>4</sup>, and Jean-Louis Mugnier<sup>5</sup>

<sup>1</sup>Université Savoie Mont Blanc, Polytech Annecy-Chambéry, LISTIC, B.P. 80439, F-74944 Annecy Cedex, France

Email: {hoang-viet-tuan.nguyen | nicolas.meger | emmanuel.trouve}@univ-smb.fr

<sup>2</sup>Univ Lyon, INSA-Lyon, CNRS, INRIA, LIRIS, UMR5205, F-69621, France

Email: christophe.rigotti@insa-lyon.fr

<sup>3</sup>Univ Lyon, INSA-Lyon, CNRS, LIRIS, UMR5205, F-69621, France

Email: catherine.pothier@insa-lyon.fr

<sup>4</sup>University of Edinburgh, School of GeoSciences, Edinburgh EH8 9XP, UK

Email: ngourme2@staffmail.ed.ac.uk

<sup>5</sup>Université Savoie Mont Blanc, CNRS, ISTerre, F-73376 Le Bourget-du-Lac, France

Email: jean-louis.mugnier@univ-smb.fr

## Abstract

For more than 40 years, Earth observation satellites have been regularly providing images of glaciers that can be used to derive surface displacement fields and study their dynamics. In the context of global warming, the analysis of *Displacement Field Time Series (DFTS)* can provide useful information. Efficient data mining techniques are thus required to extract meaningful displacement evolutions from such large and complex datasets. In this paper, a pattern-based data mining approach which handles confidence measures is proposed to analyze DFTS. In order to focus on the most reliable measurements, a displacement evolution reliability measure is defined. It is aimed at assessing the quality of each evolution and pruning the search space. Experiments on two different DFTS (annual displacement fields derived from optical data over Greenland ice sheet and 11-day displacement fields derived from SAR data over Alpine glaciers) show the potential of the proposed approach.

**Keywords** Satellite image time series (SITS), displacement field time series (DFTS), confidence measure, data mining, glacier dynamics, climate change.

# 1 Introduction

Studying glacier dynamics is of primary interest to assess the local impact of climate change and derive their contribution to the sea level rise. One effective way to characterize glacier evolutions consists in analyzing *Displacement Field Time Series (DFTS)* obtained from *Satellite Image Time Series (SITS)* (e.g., [2, 19, 5, 9]). When it comes to automatically gain insights into a DFTS, data mining techniques are to be considered. A DFTS is indeed a large and complex spatiotemporal vectorial data volume that cannot be explored manually. As reported in [10] and [18], sequential pattern mining techniques can be successfully applied to SITS. They can also be used to mine DFTS, as evidenced in [17], allowing to discover interesting displacement evolutions over time and space.

Nevertheless, in [17], confidence measures were not taken into account. These measures express the quality of the measured displacements and to which extent DFTS data can be trusted. They are computed either while building DFTS, e.g., by considering correlation peaks, full widths at half maximum, or signal to noise ratio ([23, 3]) if amplitude correlation is performed, or also once a DFTS has been obtained, by assessing the spatial and/or temporal distribution of the displacement vectors (e.g., [25, 4]). Confidence measures are available for each data point, i.e., they are available for each location and for each pair of acquisitions. They therefore represent datasets that are as large and complex as DFTS data themselves.

To our knowledge, no automatic method for discovering displacement evolutions from DFTS while handling confidence measures is available. In the state-of-the-art approaches such as [25], [4] or [20], displacement evolutions analysis is performed as follows: after having estimated the data quality by computing confidence measures, low confidence data points are filtered out using the same threshold for the whole dataset. A few transects in the areas of interest are then chosen manually. In turn, displacement profiles are drawn along these transects, at different dates, and are overlaid on the same graph. The displacement evolutions and the possible impact of remaining quality fluctuations are finally assessed by visual inspection. Such an approach has been for example applied in [25] to exhibit the decadal slowdown of the Greenland Ice Sheet and in [20] to characterize the evolution of glacier surges in Karakoram region.

In this paper, the proposed method aims to automatically extract displacement evolution profiles as sequential patterns that are built on data points whose average quality is acceptable. When compared with 1D transect-based approaches, the advantage of our proposal is that it completes them by providing a 2D view of the evolution profiles for large areas in the form of maps. Sequential patterns can indeed automatically extract any kind of displacement evolutions covering sufficiently large and homogeneous areas, without being limited to transects. Experiments on two datasets of very different quality show that the method confirms the evolutions reported along transects in [25] and [7], and provides the spatial extent of the corresponding phenomena.

An alternative strategy, based on a combination of state-of-the-art techniques, would be to filter out low confidence data points (e.g., [25, 4, 20]) and mine sequential patterns as done in [17] to analyze the displacement evolutions of the remaining data. When compared with such a filter-based approach, the benefit of our proposal is that it captures evolutions accounting for larger parts of the dataset. It indeed averages and checks confidence measures at the evolution pattern occurrence level

instead of removing all low confidence data points. This is achieved by extending the notion of evolution pattern occurrences, which is a fruitful approach in data mining as demonstrated by the works about probabilistic databases (e.g., [11, 13]). The extension proposed in this paper does not consider data distributions as done in such probabilistic settings since it is aimed at handling a displacement quality measure such as the confidence one.

In summary, this paper presents an approach for extracting reliable displacement evolution patterns from DFTS data by exploiting confidence measures at the pattern occurrence level. Its main contributions are: (1) the definition of a reliability measure for displacement evolution patterns, (2) a pruning technique to reduce the search space, and (3) the application of the proposed approach to the monitoring of glacier displacement field time series. Experimental results are presented for high quality annual displacement fields built from a Landsat medium resolution optical SITS and low quality 11-day displacement fields computed from a TerraSAR-X high resolution radar SITS. These DFTS respectively cover the Greenland ice sheet and Alpine glaciers.

This paper is structured as follows: Section 2 presents the proposed method along with the assessment framework and the parameter setting, while Section 3 reports both qualitative and quantitative experimental results. Finally, Section 4 draws conclusions and gives future work directions. Preliminary ideas regarding this contribution were presented in [14].

## 2 Methodology

A DFTS computed from a SITS and defined for a set of time stamps  $T = \{t_1, \dots, t_n\}$  gives, for each  $t \in T$  and for each pixel location  $(x, y)$ , the displacement vector observed between date  $t$  and another date at location  $(x, y)$ . Such a vector can also be considered as a velocity if the original SITS has a regular temporal sampling or if the displacements are divided by elapsed times. In the case of glaciers, this holds only if ablation and accumulation effects are neglected or if horizontal displacements are measured, which is the case in our experiments. Consequently, in this paper, the proposed method is described by focusing on velocities though it can process either displacements or velocities indistinguishably. It is based on the following workflow:

1. a *symbolic displacement field time series* is built and confidence measures are computed,
2. *reliable Grouped Frequent Sequential patterns (GFS-patterns)* are extracted from the symbolic displacement field time series,
3. the most promising reliable GFS-patterns and their corresponding *Spatio Temporal Localization maps (STL-maps)* are selected and ranked.

The structure of this section follows this workflow depicted in Figure 1. In addition, Section 2.4 presents the assessment framework of the proposed method and Section 2.5 details the parameter setting.

### 2.1 Data preparation

Once a DFTS is available, a first optional step is standardization. Then, the DFTS is encoded into symbols to obtain a symbolic displacement field time series. In



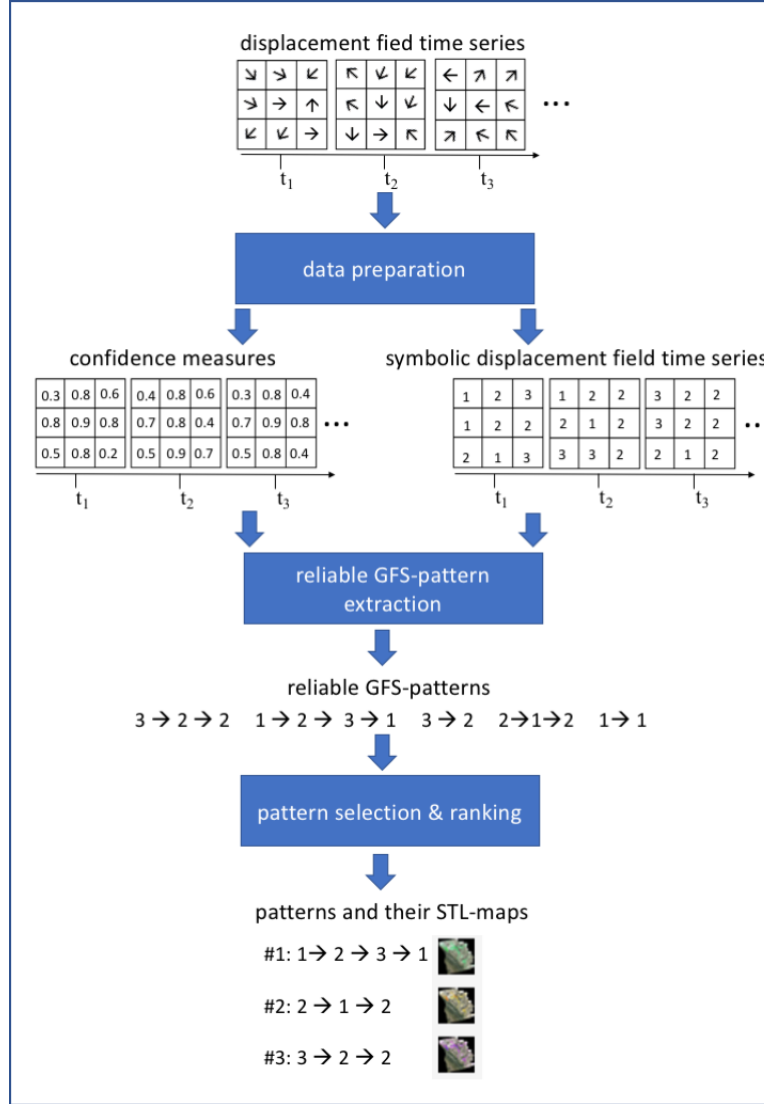


Figure 1: From displacement field time series to reliable GFS-patterns: the workflow.

addition, if no confidence measures originating from the DFTS computation are available, then they are established from the DFTS itself.

### 2.1.1 DFTS standardization

This step is optional and the remaining steps could be applied to a non-standardized DFTS. Nevertheless, it is recommended to standardize the displacement fields to make them comparable. This is done for example in [25] or in [4] where velocities are standardized by computing *median differential velocities*: velocities are centered with respect to their median value before being normalized.

If standardized velocities are not provided beforehand, we propose in this paper to compute robust standardized estimates by exploiting the local stationarity and homogeneity of DFTS data. Let  $\vec{v}(x, y, t)$  be the velocity vector observed between date  $t$  and another date at location  $(x, y)$ . The initial DFTS is reduced by performing a low-pass filtering and sub-sampling based on a spatial tiling. Let  $\Omega_{i,j,t}$  be the list of the values of  $\|\vec{v}(x, y, t)\|$  contained in a tiling window  $w_{i,j,t}$ , and  $\Omega_{i,j}$  be the concatenation of the lists  $\Omega_{i,j,1}, \Omega_{i,j,2}, \dots, \Omega_{i,j,n}$  (i.e., all values over time in windows of indices  $(i, j)$ ). Let  $MAD_{i,j}$  denotes the Median Absolute Deviation at location  $(i, j)$  and defined simply as:

$$MAD_{i,j} = \text{median}_{z \in \Omega_{i,j}} \{|z - \text{median}(\Omega_{i,j})|\} \quad (1)$$

Then, the values describing the displacement fields are the median differential velocities at each timestamp  $t$ , and they are obtained for each window  $w_{i,j,t}$  as follows:

$$mdv(i, j, t) = \frac{\text{median}(\Omega_{i,j,t}) - \text{median}(\Omega_{i,j})}{MAD_{i,j}} \quad (2)$$

The values are centered and normalized by using the median and the  $MAD$  instead of a mean and a standard deviation in order to make this standardization robust to the presence of outliers. The resulting measure captures the evolution of the median of the velocity values over one tiling window  $w_{i,j,t}$  compared to the median value over all the windows at the same location  $(i, j)$  for  $t \in T$ .

### 2.1.2 Symbolic DFTS generation

The method presented in this paper does not focus on precise variations of displacement values but rather aims to exhibit their evolutions over time and space at a more abstract level. Therefore, the next step consists in encoding the magnitude of the displacements using symbols of a discrete domain computed by equal frequency bucketing (percentiles). In the experiments reported in Section 3, three symbols are used to represent the magnitude of velocities: 1=low, 2=medium and 3=high. This leads to a *symbolic displacement field time series* containing for each location  $(x, y)$  an *evolution sequence*  $seq(x, y) = \langle (t_1, \alpha_1), \dots, (t_i, \alpha_i), \dots, (t_n, \alpha_n) \rangle$  where each  $t_i$  is a timestamp and the associated  $\alpha_i$  is the symbol encoding the magnitude of the velocity at date  $t_i$ . Notice that other features of the DFTS could be encoded instead of the velocity magnitudes, as for instance displacement directions.

### 2.1.3 Confidence measure computation

The proposed method requires each displacement to be characterized by a confidence value expressing to which extent it can be trusted. If no confidence measures are available beforehand, then they can be obtained either by exploiting redundant observations or the flow direction stationarity.

*Redundant observations.* Let  $\vec{V}(x, y, t)$  be a set of velocity estimates computed from different pairs of images. For the displacement observed at location  $(x, y)$  and date  $t$ , this set can be used to calculate both the velocity  $\vec{v}(x, y, t)$  and its

corresponding confidence  $\rho(x, y, t)$ . In this paper, one of the confidence measure that is used is the *velocity vector coherence* introduced in [4] and defined as follows:

$$\rho(x, y, t) = \begin{cases} \frac{\|\sum_{\vec{v} \in \vec{\mathcal{V}}(x, y, t)} \vec{v}\|}{\sum_{\vec{v} \in \vec{\mathcal{V}}(x, y, t)} \|\vec{v}\|}, & \text{if } \sum_{\vec{v} \in \vec{\mathcal{V}}(x, y, t)} \|\vec{v}\| \neq 0 \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

It behaves as the coherence in SAR interferometry: it varies in the interval  $[0, 1]$ . It is equal to 1 if all vector estimates point in the same direction, and tends to 0 if these vector estimates have uniform random directions.

*Flow direction stationarity.* If no redundant observations are available, then the stationarity of the flow direction can be used to derive confidence measures reflecting the temporal directional coherence of the 2D velocity vectors. For example, in [22] or [8], the assumption of a displacement flow in the direction of the maximum downhill slope was considered to study glaciers. This implies that, for a given location, the glacier displacement direction at different dates must be similar.

In this case, for each location  $(x, y)$  and for each timestamp  $t$ , the confidence measure of the velocity  $\rho_{disp}$  is defined as:

$$\rho_{disp}(x, y, t) = \begin{cases} \cos(\widehat{\vec{u}_{x, y, t}, \bar{u}_{x, y}}), & \text{if } \cos(\widehat{\vec{u}_{x, y, t}, \bar{u}_{x, y}}) \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where  $\vec{u}_{x, y, t}$  is the unit velocity vector in location  $(x, y)$  at timestamp  $t$  and  $\bar{u}_{x, y} = \sum_{t=1}^n \vec{u}_{x, y, t}$  reflects the global velocity direction over the sequence located in  $(x, y)$ . The cosine operator is used to measure the angular difference between the global velocity direction and the velocity vector at timestamp  $t$ . The larger is the angle, the smaller is the confidence. In order to get robust estimates, and as previously explained in Section 2.1.1, the original size of the acquisitions is reduced by tiling each velocity field with windows noted  $w_{i, j, t}$ , where  $t$  is the velocity field time stamp and  $(i, j)$  is a sub-sampled location. The final measure  $\rho(i, j, t)$  is then chosen to be the median of the confidence  $\rho_{disp}$  over the window  $w_{i, j, t}$ .

## 2.2 Reliable GFS-pattern extraction

In this section, the concept of *Grouped Frequent Sequential pattern (GFS-pattern)* is recalled before being extended to the concept of *reliable GFS-pattern* for which an efficient extraction technique is proposed.

### 2.2.1 GFS-patterns

As introduced in [17], the *Grouped Frequent Sequential patterns (GFS-patterns)* of [10] can be adapted to analyze symbolic displacement time series. Let us recall their corresponding definitions in this context. A *sequential pattern*  $\beta$  is here a pattern of the form  $\beta_1 \rightarrow \beta_2 \rightarrow \dots \rightarrow \beta_m$  where  $\beta_1 \dots \beta_m$  are any symbols used for the encoding. A location  $(x, y)$  is termed *covered* by  $\beta$  if somewhere in  $seq(x, y)$  we find symbol  $\beta_1$  and then some time later  $\beta_2$  and so on, before finally observing

symbol  $\beta_m$ . More formally, let  $seq(x, y) = \langle (t_1, \alpha_1), (t_2, \alpha_2), \dots, (t_n, \alpha_n) \rangle$  be an evolution sequence, then location  $(x, y)$  is covered by  $\beta$  if there exist  $i_1 < i_2 < \dots < i_m$  so that  $\beta_1 = \alpha_{i_1}, \beta_2 = \alpha_{i_2}, \dots, \beta_m = \alpha_{i_m}$ . In this case,  $o = \langle t_{i_1}, t_{i_2}, \dots, t_{i_m} \rangle$  is called an *occurrence* of  $\beta$  in  $seq(x, y)$ . In addition,  $t_{i_k}$ , defined for all  $k \in \{1, \dots, m\}$ , is called the  $k^{th}$  occurrence date of  $o$ .

The GFS-patterns [10] are the sequential patterns that are *frequent* and *grouped*. The term *frequent* means that a pattern covers at least  $\sigma$  locations  $(x, y)$  in the studied area (i.e., it covers at least a minimum surface). A pattern  $\beta$  is said to occur in a *grouped* way if, on average, a location covered by  $\beta$  is surrounded by at least  $\kappa$  other locations in its 8-neighborhood that are also covered by  $\beta$ . This second constraint is used to ensure that the occurrences of  $\beta$  are not spatially too isolated. The parameters  $\sigma$  and  $\kappa$  are respectively called the *surface threshold* and the *grouping threshold*.

To visualize the location in space and time of a pattern  $\beta$ , a *Spatio Temporal Localization map (STL-map)* is drawn. The STL-maps of  $\beta$  is simply an image where a pixel at location  $(x, y)$  is black if  $(x, y)$  is not covered by  $\beta$ , and is otherwise displayed with a color indicating the ending date of the first minimal occurrence of  $\beta$  in  $seq(x, y)$ . An occurrence  $o = \langle t_{i_1}, t_{i_2}, \dots, t_{i_m} \rangle$  of pattern  $\beta$  is minimal if  $\beta$  does not occur in any proper subinterval  $[t, t'] \subset [t_{i_1}, t_{i_m}]$ . The first minimal occurrence of  $\beta$  in a sequence  $seq(x, y)$  is the one with the earliest occurrence dates. More formally, a minimal occurrence  $o = \langle t_{i_1}, t_{i_2}, \dots, t_{i_m} \rangle$  is the first minimal occurrence if for all minimal occurrences  $o' = \langle t'_{i_1}, t'_{i_2}, \dots, t'_{i_m} \rangle$ , we have  $t_{i_k} \leq t'_{i_k}, \forall k \in \{1, \dots, m\}$ .

## 2.2.2 Reliable GFS-patterns

Starting from GFS-patterns, our proposal is to handle confidence measures at the pattern occurrence level to focus on reliable patterns without having to threshold the whole dataset at the data point level. Let  $\rho(x, y, t)$  denote the confidence associated to location  $(x, y)$  for timestamp  $t$ . The method does not focus on a particular measure of confidence, the intuition is simply that the higher is the confidence  $\rho(x, y, t)$ , the more confident one can be in the values of the displacement field at date  $t$  for location  $(x, y)$ .

The notion of reliability is now considered at different scales. From  $\rho(x, y, t)$  we define  $\rho_{occ}(x, y, o)$  the reliability measure of an occurrence  $o = \langle t_{i_1}, t_{i_2}, \dots, t_{i_m} \rangle$  in a sequence  $seq(x, y)$  as  $\rho_{occ}(x, y, o) = \min\{\rho(x, y, t) | t \text{ in tuple } o\}$ . This measure adopts a conservative point of view, associating to an occurrence  $o$  the minimal confidence of its constituting elements.

At the sequence level, the reliability of a pattern  $\beta$  in a sequence  $seq(x, y)$  aims to reflect to which extent  $seq(x, y)$  contains a good quality occurrence of this pattern. If  $\beta$  covers location  $(x, y)$ , then this measure is defined as  $\rho_{pat}(x, y, \beta) = \max_{o \in \mathcal{O}} \{\rho_{occ}(x, y, o)\}$  where  $\mathcal{O}$  is the set of occurrences of  $\beta$  in  $seq(x, y)$ . Consequently, instead of relying on the standard sequential pattern [1] and GFS-pattern [10] schemes that only check whether a pattern occurs or not in a sequence, the notion of occurrence is here extended to select the most reliable one.

The reliability measure  $\rho$  retained for a pattern in a dataset is then simply the mean of its reliability over all the locations covered by the pattern:

$$\rho(\beta) = \frac{\sum_{(x, y) \in cov(\beta)} \rho_{pat}(x, y, \beta)}{|cov(\beta)|} \quad (5)$$

where  $cov(\beta)$  is the set of locations  $(x, y)$  covered by  $\beta$  and  $|cov(\beta)|$  is the size of this set. Low reliability occurrence measures are thus compensated by high ones. A pattern is then said to be *reliable* if its reliability measure in the dataset is greater or equal to a reliability threshold  $\gamma$ , i.e.,  $\rho(\beta) \geq \gamma$ .

This constraint allows to select patterns built on velocity values that reach on average a given level of reliability. In addition, it can also lead to a more efficient extraction of the reliable GFS-patterns when used as an active constraint during the search for these patterns. This can be done as follows, by defining an upper bound of  $\rho(\beta)$ . Let pattern  $\beta$  be called a *subpattern* of a pattern  $\beta'$ , denoted  $\beta \prec \beta'$ , if  $\beta$  can be obtained by removing one or more symbols from  $\beta'$ . The function  $\rho(\beta)$  is not *anti-monotonic*, in the sense that  $\beta \prec \beta'$  does not imply  $\rho(\beta) \geq \rho(\beta')$ . Let us define  $\tilde{\rho}(\beta)$  as

$$\tilde{\rho}(\beta) = \frac{\sum_{(x,y) \in cov(\beta)} \rho_{pat}(x, y, \beta)}{\sigma} \quad (6)$$

where  $\sigma$  is the surface threshold. The function  $\tilde{\rho}(\beta)$  is anti-monotonic because  $\beta \prec \beta'$  implies that  $cov(\beta) \supseteq cov(\beta')$  and thus  $\tilde{\rho}(\beta) \geq \tilde{\rho}(\beta')$ . Moreover, since  $\sigma$  is the minimal possible value for the surface covered by a GFS-pattern then  $\sigma \leq |cov(\beta)|$  and so for all  $\beta$  we have  $\tilde{\rho}(\beta) \geq \rho(\beta)$  providing thus an upper bound of the reliability of  $\beta$ .

During the search for the reliable GFS-patterns these properties allow to reduce the search space as follows. If we find a pattern  $\beta$  such that  $\tilde{\rho}(\beta)$  is strictly lesser than  $\gamma$  the minimum reliability threshold, then this implies that for all  $\beta'$  such that  $\beta \prec \beta'$  we have  $\gamma > \tilde{\rho}(\beta) \geq \tilde{\rho}(\beta') \geq \rho(\beta')$ . Thus  $\beta'$  can be excluded from the search space since its reliability cannot reach the  $\gamma$  threshold. This potential reduction of the number of candidate patterns, using anti-monotonicity, can be incorporated in many pattern mining approaches as the general pattern-growth technique for sequences [16] on which is based our implementation. It should be noticed that, in addition to this pruning, the finding of the occurrence of a pattern  $\beta$  having the highest reliability in a sequence (needed to compute  $\rho_{pat}(x, y, \beta)$  for each location) can be done efficiently using dynamic programming, as it is the case for similar problems like sequence alignment.

## 2.3 Pattern selection and ranking

Finally, though surface, connectivity and reliability constraints are applied to focus on the most interesting sequential patterns, reliable GFS-patterns can still be numerous. Thus, in order to get the most detailed evolutions, the maximal reliable GFS-patterns are selected. These maximal patterns are the ones that are not a subpattern of any other reliable GFS-pattern. End-users are then guided towards the most promising maximal patterns using a randomization-based ranking procedure. This method was introduced in [12]. It is aimed at identifying the STL-maps/patterns that are either hardly or fully destroyed by randomization. In other words, the STL-maps obtained for original DFTS are compared with the ones obtained for randomized DFTS using a normalized mutual information measure. The latter is used to produce a ranking ranging from the most expected STL-maps/patterns to the most unexpected ones, with respect to the symbol distributions in the symbolic DFTS. The  $K$  most expected and the  $K$  most unexpected reliable GFS-patterns are finally retained, with  $K$  a user defined parameter.

## 2.4 Assessment of the proposed method

The results of data mining techniques are assessed according to their nature [24]. When considering classification, standard Quality Assessment Measures (QAM) such as accuracy, precision, or recall measures are computed. Other global models of the data such as clustering are for example assessed using entropy, Sums of Squared Errors (SSE) or silhouette coefficients. The assessment of patterns such as reliable GFS-patterns is based on quality assessment measures that are different. Extracted models are indeed not global ones but are local dependencies describing parts of the dataset.

The surface measure used in this paper corresponds to the *support* measure, which is the standard QAM for frequent sequential patterns, as originally defined in [1]. It gives the number of sequences in which a pattern occurs. Sequential patterns whose support measure is higher than a user-defined threshold are said to be frequent. Frequent sequential patterns are further assessed using their support measures to focus either on the less or on the most frequent ones. This measure makes sense within the original scheme. For example, each sequence could be a purchasing sequence holding for a certain customer. Consequently, the support measure of frequent sequential pattern gives the number of customers sharing this purchasing pattern, which is a very direct and simple indicator.

Back to reliable GFS-pattern, if we were to mimic what is done for frequent sequential patterns, then they would have been to be assessed using the surface, the connectivity, the reliability and the normalized mutual information measures. These numerous measures must be balanced by taking into account the final application, which is not as straightforward as for standard sequential patterns. A more direct and still objective assessment is to measure, on average, to which extent extracted patterns describe the dataset in terms of data points, as we are interested in both the spatial and the temporal dimensions. Let  $\beta = \beta_1 \rightarrow \beta_2 \rightarrow \dots \rightarrow \beta_m$  be a frequent grouped sequential pattern. If a location  $(x, y)$  is covered by  $\beta$ , the most reliable occurrence of  $\beta$  in  $seq(x, y)$  is composed of  $m$  elements. Thus, the *Data Point cover* of  $\beta$  with these reliable occurrences over all locations is  $DP_{cover}(\beta) = cover(\beta) * m$ . Let  $R$  be the set of the reliable GFS-patterns obtained as a result of the proposed method. The *Mean Data Point cover* of  $R$  is defined as follows:

$$MDP_{cover}(R) = \frac{\sum_{\beta \in R} DP_{cover}(\beta)}{|R|} \quad (7)$$

This measure is exploited in this paper to compare the proposed method with the closest competing method. The latter is built on the available state-of-the-art approaches [25, 4, 20, 17]. It extracts and selects the same number of GFS-patterns from the same symbolic DFTS whose low confidence data points have been filtered out in a first place. As for the proposed method, the maximal GFS-patterns are selected and ranked using a randomization-based procedure to build a final set of GFS-patterns containing as many patterns as obtained with the proposed approach. For the sake of clarity, in the remaining of this paper, this competing method is termed *filter-based GFS-pattern extraction*. It does not require parameter  $\gamma$ , but needs the same other parameters and an additional one,  $\rho_{filter}$ , the confidence threshold used to filter out low confidence data points.

## 2.5 Parameter setting

The parameter setting is performed as follows. First, the grouping threshold is set to  $\kappa = 5$ . This is a standard setting for exhibiting meaningful geophysical phenomena (see [10] and [17]). It ensures that if an evolution is retained, then, when it occurs at a location  $(x, y)$ , it also occurs, on average, in at least 5 locations of the 8-neighborhood of  $(x, y)$ . Evolutions occurring in isolated locations only are thus discarded.

The second parameter is  $\sigma$ , the surface threshold. It is set so as to retrieve the largest number of maximal GFS-patterns, i.e., to obtain, in some sense, the richest description.

With regard to the reliability threshold  $\gamma$ , the more stringent it is (i.e., closer to 1), the lower is the number of retrieved patterns. Its value is chosen so as to reach a trade-off between the number of patterns and their quality (reliability). To find such a compromise, a very simple objective function to maximize is used: the product of the threshold  $\gamma$  and the number of patterns obtained for this threshold. It is noted  $o(\gamma) = \gamma \times p$  with  $p$  the number of reliable maximal GFS-patterns that are extracted.

The parameters for the filter-based GFS-pattern extraction (see Section 2.4) are chosen so as to consider the closest setting,  $\rho_{filter}$  is set to the value of  $\gamma$  and the same values as for the proposed method are used for  $\kappa$  and  $\sigma$ . Finally,  $K$  is set to 20, which is a standard setting [12], to get the 20 most expected and the 20 most unexpected patterns.

## 3 Experiments

### 3.1 Greenland ice sheet

#### 3.1.1 Dataset

For this experiment, we directly use the DFTS that was prepared and used in [25] to exhibit a decadal slowdown in the ablation zone of the western Greenland Ice Sheet. It was built from a series of Landsat images (Landsat 5, 7 and 8) spanning three decades, from 1985 to 2013. It contains annual median differential velocities that were spatially aggregated and sub-sampled to get reliable estimates. The final resolution is  $240 \times 240$  m and a zone of  $458 \times 500$  pixels is described for 20 timestamps. The first seven fields correspond to years 1985 to 1999 (one field every two-year interval) and the thirteen other fields correspond to years 2000 to 2013 (one field every one-year interval).

The symbolic DFTS is built by encoding mean differential velocity values using three symbols (1, 2 and 3) and an equal frequency bucketing. For each location  $(x, y)$ , this leads to an evolution sequence  $seq(x, y) = \langle (t_1, \alpha_1), \dots, (t_i, \alpha_i), \dots, (t_n, \alpha_n) \rangle$  with  $t_i \in T$  and  $\alpha_i \in \{1, 2, 3\}$ , where symbol 1 denotes a slow velocity, 2 a velocity close to the median value, and 3 a higher velocity.

Finally, each median differential velocity value, and thus each symbol, is associated to a confidence measure that was computed and used in [25]. It is given by the velocity vector coherence computed according to Equation 3 by aggregating several observations of the one-year displacements (see Section 2.1.3). Most locations in the selected region suffer from important loss of confidence as shown in figures 2a, 2b

and 2c, that give, for each location  $(x, y)$ , the minimum, maximum and mean of the confidence measure  $\rho(x, y)$  over the series (according to the color scale of Figure 2d).

### 3.1.2 Parameter values

The parameter setting is performed according to Section 2.5. Consequently:

- The grouping threshold  $\kappa$  is set to 5.
- The surface threshold  $\sigma$  is set to 7.5% according to Figure 3a. The latter presents the numbers of maximal patterns output when varying the minimum surface threshold (expressed as a percentage of the  $458 \times 500$  pixels of the area).
- The reliability threshold  $\gamma$  is set to  $\gamma = 0.85$  according to Figure 3b. The latter presents the product of threshold  $\gamma$  and the number of patterns obtained versus  $\gamma$ , i.e.  $o(\gamma) = \gamma \times p$ .
- The minimum confidence threshold  $\rho_{filter}$  used by the filter-based method is set to  $\gamma = 0.85$ .
- The number of the most expected (respectively unexpected) patterns  $K$  is set to 20, i.e. 40 patterns are selected.

### 3.1.3 Quantitative results

The proposed method extracts, selects and ranks 375 maximal reliable GFS-patterns before outputting  $R$ , the 40 patterns result set. Its mean cover is  $MDP_{cover}(R) = 201,981.6$  data points.

The filter-based GFS-pattern extraction method also extracts, selects and ranks 375 maximal GFS-patterns that differ from the ones obtained with the proposed method. Its 40 patterns result set  $R'$  covers fewer data points in average:  $MDP_{cover}(R') = 188,418.6$  data points.

This variation of the quality assessment measure  $MDP_{cover}$  shows that the proposed method extracts patterns that, on average, account for 7.2% more data points than what can be achieved with filter-based GFS-patterns. More details about the cover distribution are given in Figure 3c by the histogram of the number of patterns with respect to the cover values. It shows that reliable GFS-patterns (in blue) tend to cover more data points than filter-based GFS-patterns (in orange).

Concerning the resource consumption, the whole process implemented in C and Python (from data discretization to STL-maps selection) requires 813 seconds using a single core of an Intel Xeon 3.5GHz running Linux (Ubuntu). The maximal memory consumption is 311 MB.



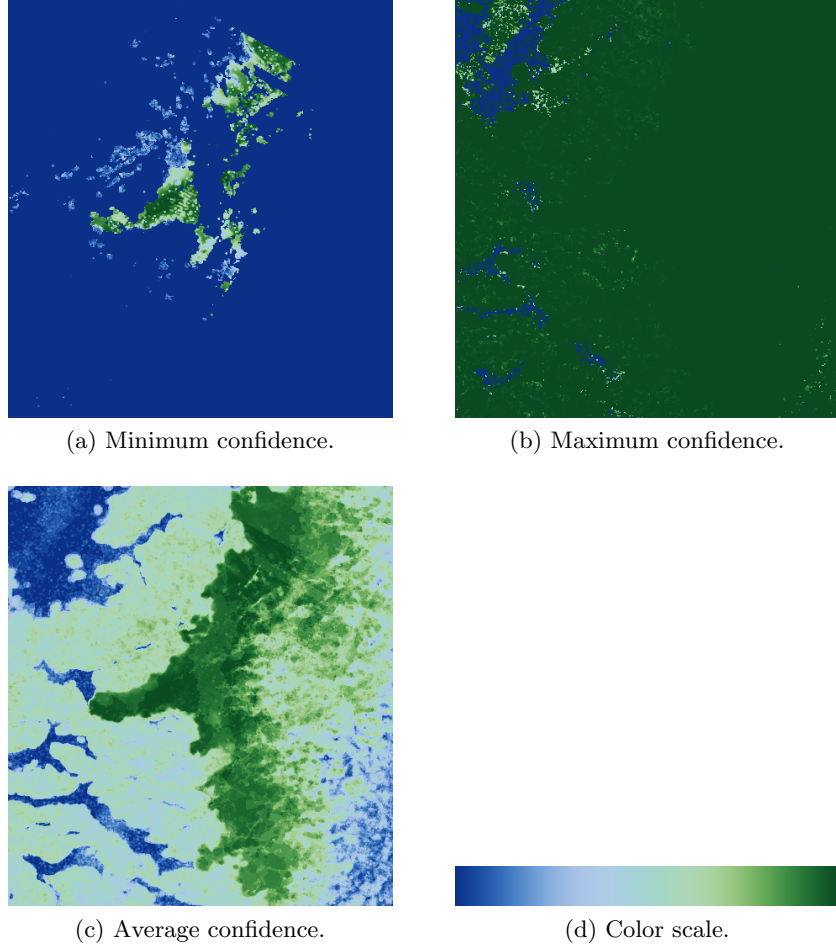


Figure 2: Confidence of each location in Greenland over the series: (a) minimum, (b) maximum, (c) average. (d) Confidence color scale: from 0 (dark blue) to 1 (dark green).

### 3.1.4 Qualitative results

The qualitative results are discussed by considering the STL-maps of the extracted patterns. These maps reflect different evolutions over various regions at different times, and three among the best ranked ones are presented in figures 4a, 5a and 5b drawn on a grayscale background image. The locations of the main relevant glaciers are indicated by labels (1), (2) and (3) Figure 4b. The first two maps are complementary evidences of the regional slowdown reported in [25]. The one given in Figure 4a corresponds to a deceleration pattern:  $3 \rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1$ . The map shows with colored dots all the locations where this pattern is found in the data. For each of these locations, the color indicates the date at which the last 1 of the pattern is found (according to the color scale of Figure 6). The highlighted part in the middle is the Nordenskjöld glacier (1), for

which a slowdown along one longitudinal transect was exhibited in [25]. The map shows that this deceleration pattern also occurs in the northern part, close to the Sarqardliup sermia (3), and in the South in the area of the Polonia glacier (2). Notice that according to the color scale the end of the pattern is observed earlier in the middle-left part (mostly blue) than in the northern and southern parts (magenta). Another displacement type, different from this progressive slowdown, is depicted in complementary areas in Figure 5a. Here, the associated pattern  $3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 1$  underlines a long stable period at a (locally) high speed level before a sudden deceleration. A dedicated search in the dataset shows that the sequence of the symbols 2 in the pattern of Figure 4a ends by 2006-2007, while for the less gradual slowdown of Figure 5a the end of the sequence of the symbols 3 occurs earlier (by 2003-2005). At that time, the velocity was already at level 2 for the pattern of Figure 4a. This means that, even if the two patterns express the same global motion, i.e., a deceleration, they correspond to different evolutions exhibited by the method. Such a detailed analysis can still be performed when the underlying regions are adjacent. For example, in the Nordenskjöld glacier area (1), the method isolates the two different evolutions: the progressive slowdown given by the pattern of Figure 4a in the center of the Nordenskjöld glacier, and the sudden deceleration reported for the other parts of the glacier by the pattern of Figure 5a.

A third map is given Figure 5b. It corresponds to pattern  $1 \rightarrow 1 \rightarrow 3 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1$  that could suggest a singular speedup within a sequence of low speed level. It highlights in blue the area (3) of the Sarqardliup sermia and Alangordliup sermia. Again a specific verification in the dataset shows that there is a local maximum at level 3 over 1997-1999 in the North of area (3) and over 1990-1992 in its southern part.

These three maps of coherent patterns given in figures 4 and 5 are spatially consistent with the mean of the velocity confidence measures depicted in Figure 2c. However, it should be noticed that a simple preprocessing strategy based on a selection of the regions having a high mean for the confidence measure over the series (locations in green, Figure 2c) would not have been sufficient. Indeed, it would not have prevented us from building poor quality pattern occurrences in these regions, because as shown in Figure 2a, poor data velocity confidence (in blue) can be observed almost everywhere in the area. Moreover, as evidenced in Figure 2b, for nearly all locations, there exist some dates showing good confidence measures (in green). This potentially allows reliable occurrences of patterns to be built even in areas where the mean of the confidence measure is low. These observations advocate for the handling of the confidence at the occurrence level, as for instance in the reliability measure-based method proposed in this paper: it allows to extract meaningful velocity evolution maps covering areas that would have been filtered out if, for example, we were to rely on a filter-based extraction.

## 3.2 Alpine glaciers

### 3.2.1 Dataset

The proposed approach has also been applied to explore short time displacement evolutions over Alpine glaciers in the Mont Blanc massif. The DFTS is built using 26 TerraSAR-X images (ascending track), with a pixel resolution of about  $2 \text{ m} \times 2 \text{ m}$ . The series was acquired between 05/31/2009 and 09/25/2011. It covers

the French part of the Mont Blanc massif and contains its four main glaciers: Argentière, Mer de Glace, Bossons, and Taconnaz. It is composed of two distinct periods, one in 2009 from May 31 to October 21 and another in 2011 from May 5 to September 25. The two periods contain 13 images each (one image every 11 days). Glacier displacements can be more easily observed during these periods of the year compared to other ones because there is less snow.

The original size of the images is  $10484 \times 9560$  pixels. First, we describe how this data have been processed to compute the displacement fields at full resolution, and then have been reduced to a size of  $3494 \times 3186$  to get robust estimates. *Synthetic Aperture Radar (SAR)* images contain amplitude and phase information, which both can be used to estimate different kinds of displacements. For Alpine glaciers, rapid surface changes are likely to reduce the phase coherence and amplitude information is more adapted for calculating glacier displacements (e.g., [23, 22, 6]). Moreover, instead of having only displacements in the *line of sight* with phase-based methods, amplitude-based methods give us displacement vectors expressed in *azimuth* and *range* directions. Thus, to compute the displacement fields the amplitude cross-correlation method implemented in the *EFIDIR Tools*<sup>1</sup> is used. After having applied a glacier mask from the *Randolph Glacier Inventory*<sup>2</sup>, the cross-correlation is performed for every consecutive image pair, with a correlation window size of 65 pixels and a search window size of 105 pixels. A total of 25 displacement fields expressed in *m/day* is obtained.

Median differential velocities are then computed by applying Equation 2 as well as the corresponding low-pass filtering and sub-sampling using a  $3 \times 3$  tiling windows. Consequently, the initial velocity field size of  $10484 \times 9560$  is reduced to  $3494 \times 3186$ , which, in addition to supplying robust estimates, will lower resource consumption when it comes to extracting patterns. The symbolic DFTS is then built from the median differential velocities values by quantization, using an equal frequency bucketing and symbols 1, 2 and 3 (denoting again low, medium and high values).

From the initial DFTS that is available at full resolution, and following the flow coherence strategy presented in Section 2.1.3, the confidence of each remaining data point, noted  $\rho_{disp}$ , is then obtained according to Equation 4. Finally, as explained in Section 2.1.3, the confidence  $\rho(i, j, t)$  associated with median differential velocity values is computed over each tiling window. For each location  $(i, j)$ , the minimum, maximum, and average values over time of  $\rho(i, j, t)$  are given in Figure 7. It can be noticed that the minimum is close to 0 for nearly the whole area, while the maximum reaches a value of 1 only in limited regions. This is very different from the Greenland series. Even though the confidence measure is not formally the same as in the previous series, from a qualitative point of view, the Mont Blanc series offers a weaker confidence. These lower confidence values are consistent with the fact that for the Greenland series several pairs of images were combined to produce one displacement field [25] whereas a single image pair is used for each displacement field for the Mont Blanc series. Lower confidence values also come from the SAR imagery which suffers from the presence of speckle affecting the amplitude images as a multiplicative noise. On the TerraSAR-X time series, the speckle is decorrelated on the fast changing glacier surface (the InSAR coherence is close to 0) and reduces the performances of the cross-correlation methods that are used to

<sup>1</sup><http://efidir.poleterresolide.fr/index.php/efidir-tools>

<sup>2</sup><https://www.glims.org/RGI/>

compute DFTS. Consequently, confidence levels rarely exceeds 0.5, whereas they are often higher than 0.8 with optical data.

### 3.2.2 Parameter values

The parameter setting is performed according to Section 2.5:

- The grouping threshold  $\kappa$  is set to 5.
- The surface threshold  $\sigma$  is set to 4% according to Figure 8a.
- The reliability threshold  $\gamma$  is set to  $\gamma = 0.22$  according to Figure 8b. This value is lower than the one used for the Greenland series, which is consistent since, as shown by Figure 7, lower confidences are observed for this Mont Blanc massif series.
- The minimum confidence threshold  $\rho_{filter}$  used by the filter-based method is set to  $\gamma = 0.22$ .
- The number of the most expected (respectively unexpected) patterns  $K$  is to 20.

### 3.2.3 Quantitative results

On this DFTS, the proposed method extracts, selects and ranks 5625 maximal reliable GFS-patterns. The set  $R$  of the resulting 40 patterns has a mean cover  $MDP_{cover}(R)$  equal to 5,231,810.6 data points.

The filter-based GFS-pattern extraction method extracts, selects and ranks 6484 maximal GFS-patterns. It outputs a set  $R'$  of 40 patterns whose mean cover is  $MDP_{cover}(R') = 3,410,626.7$  data points. The proposed method extracts patterns that, on average, account for 53.4% more data points. This improvement is by far greater than the one observed for the Greenland series and directly comes from the fact the Mont Blanc massif series is a low confidence one. In other words, the lower confidence measures are, the more effective the handling of confidence measure at the occurrence level is, which allows to explore parts of the dataset that would have been filtered out by a filter-based GFS-pattern extraction. Figure 8c confirms this behavior: all reliable GFS-patterns (in blue) cover more data points than filter-based GFS-patterns (in orange).

For this series, the whole process takes 119894 seconds (about 33 hours and 18 minutes) with a maximal memory consumption of 7470 MB (same software/hardware configuration as for the Greenland study).

### 3.2.4 Qualitative results

Extracted reliable GFS-patterns and their STL-maps reflect different evolutions over the displacement series, and one among the best ranked maps is shown in Figure 9a. The locations of the relevant glaciers are given Figure 9b. The map of Figure 9a corresponds to the pattern  $3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow$

$2 \rightarrow 2$ . It exhibits two slowdowns: one from level 3 to level 1 and then from level 3 to level 2. The colors on Figure 9a indicate the dates of occurrence of the last element of the pattern (the last symbol 2) and according to the color scale drawn Figure 11, these dates are by the end of the series (late 2011).

Let us consider the first slowdown captured by the pattern (i.e., the  $3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1$  part). The beginning of the phenomenon is in early summer 2009 as depicted Figure 10a that gives the dates of occurrence of the first symbol 3 of the pattern. Figure 10b shows the dates of occurrence of the last symbol 1 and it can be observed that they correspond to the end of the first half of the series (late summer/fall of year 2009). The second part of the pattern (i.e.,  $\rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$ ) reveals that the slowdown was repeated over the second half of the series (year 2011). Therefore, our pattern based method ranks among the best maps the one related to an annual cycle of velocity variation, a well-known glaciological process for temperate glaciers (e.g., [26]). This annual cycle phenomenon was already reported on three transects of the Mont Blanc massif in [7] and [19] where a slowdown was measured in 2009: on Taconnaz (1) and Bossons (2) glaciers from their heads to their terminuses, and from Géant glacier (3) to the Mer de Glace (4) terminus. In addition to these previous analyses, the STL-map underlines the 2D spatial extent of the phenomenon, in particular for the slowdown reported in [7, page 168] 2000 m from the heads of Taconnaz and Bossons, and in the area of the seracs of the Géant (locations given Figure 9b as labels b, d and 3'). For the Bossons glacier such fluctuations are observed till  $\approx 3000$  m and suggest that cold-based glacier zone is restricted to higher altitudes.

In this series of larger size and lower confidence than the Greenland one, the handling of the confidence at the occurrence level using a reliability measure still turns out to be effective to perform a global analysis. The method allows to search for regularities among the displacement data of all main glaciers of the French part of the Mont Blanc massif. Finding such common patterns over space and time is an interesting feature of the method. It shows for instance (Figure 9a) that the same slowdown pattern that holds for the Taconnaz and Bossons glaciers is also strongly shared by the Rognons glacier (label 7, Figure 9b), the velocity of which has never been studied up to now.

## 4 Conclusion

With the quick development of Earth Observation Satellites and the necessity of research concerning climate change, DFTS data are increasingly computed and processed to assess glaciers displacement evolutions. In order to extract meaningful displacement evolution patterns, confidence measures inherent to this kind of data need to be dealt with. In this paper, a notion of pattern reliability exploiting confidence measures at the pattern occurrence level has been introduced to extract displacement evolutions patterns having reliability values that exceed a user-defined threshold. We also proposed a method to automatically set this threshold with the aim of ensuring that extracted patterns are reliable enough and bring a good description of the dataset. In addition, a pruning technique using this constraint has been also implemented to reduce the search space and therefore decrease resource consumption.

Our method has been applied to two DFTS of very different quality and covering glaciers of different sizes: an optical-based one covering Greenland ice sheet

glaciers and a radar-based one covering Alpine glaciers. The results showed that the proposed method extracts patterns accounting for larger parts of the dataset when compared to a method based on state-of-the-art techniques. Meaningful patterns were extracted from the two types of satellite data. Different maps were produced to localize these displacement evolutions in space and time. These evolutions and maps are consistent with the glaciological knowledge concerning these zones and extends it by providing 2D maps of the phenomena. Indeed, 1D profiles along transects can be found in the literature while full 2D maps containing temporal information are provided for the first time thanks to the proposed method.

Future work includes extending the scope of confidence measures. Indeed, displacement measures are first quantized into symbols. Afterwards, confidence measures, that basically characterize displacement vectors, are directly mapped to these symbols. Such a raw mapping could be refined by considering quantization effects. For example, the degrees of membership to symbols of each displacement measure should be computed and used. Another information could also be handled: the confidence in the calculus of each displacement vector derived from the characteristics of the similarity function. Such a *broader/wider* confidence measure would be more representative of the uncertainty and could be directly handled by the proposed reliability measure without any modification of the extraction algorithm.

## Acknowledgments

Funding for this project was provided by a grant from la Région Auvergne-Rhône-Alpes (Tuan Nguyen’s grant). The work was also supported by a University Savoie Mont Blanc AAP Montagne grant - project BG *Big data pour la surveillance des Glaciers* and benefited from the PHOENIX ANR-15-CE23-0012 grant of the French National Agency of Research. Catherine Pothier and Christophe Rigotti are members of LabEx IMU (ANR-10-LABX-0088) that provided complementary support. Thanks to the German Aerospace Agency (DLR) for the TerraSAR-X images (project MTH0232).

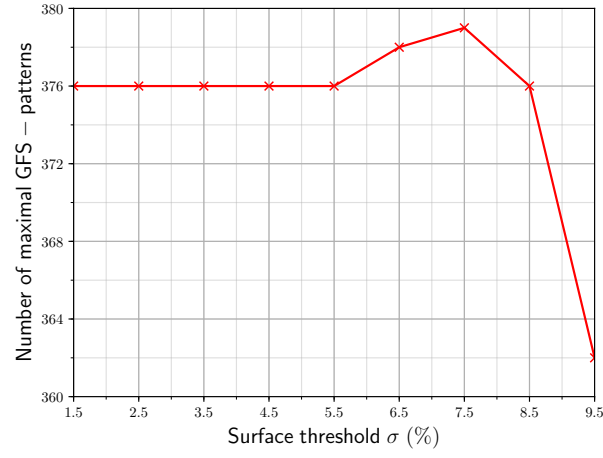
## References

- [1] Rakesh Agrawal and Ramakrishnan Srikant. Mining sequential patterns. In *Proceedings of the Eleventh International Conference on Data Engineering, ICDE '95*, pages 3–14, Washington, DC, USA, 1995. IEEE Computer Society.
- [2] V. Akbari, A. P. Doulgeris, and T. Eltoft. Monitoring glacier changes using multitemporal multipolarization SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6):3729–3741, June 2014.
- [3] Misganu Debell-Gilo and Andreas Kb. Locally adaptive template sizes for matching repeat images of mass movements. In *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, pages 4281–4284. IEEE, 2011.
- [4] Amaury Dehecq, Noel Gourmelen, and Emmanuel Trouvé. Deriving large-scale glacier velocities from a complete satellite archive: Application to the Pamir-Karakoram-Himalaya. *Remote Sensing of Environment*, 162:55–66, 2015.

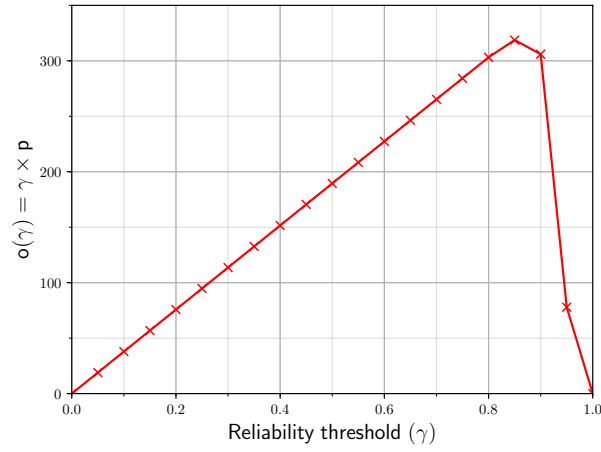
- [5] Mark Fahnestock, Ted Scambos, Twila Moon, Alex Gardner, Terry Haran, and Marin Klinger. Rapid large-area mapping of ice flow using Landsat 8. *Remote Sensing of Environment*, 185:84 – 94, 2016.
- [6] R. Fallourd, O. Harant, E. Trouvé, J. M. Nicolas, M. Gay, A. Walpersdorf, J. L. Mugnier, J. Serafini, D. Rosu, L. Bombrun, G. Vasile, N. Cotte, F. Vernier, F. Tupin, L. Moreau, and P. Bolon. Monitoring Temperate Glacier Displacement by Multi-Temporal TerraSAR-X Images and Continuous GPS Measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2):372–386, June 2011.
- [7] Renaud Fallourd. *Suivi des glaciers alpins par combinaison d’informations hétérogènes: images SAR Haute Résolution et mesures terrain*. PhD thesis, Université de Grenoble, 2012.
- [8] Olivier Harant, Lionel Bombrun, Gabriel Vasile, Laurent Ferro-Famil, and Michel Gay. Displacement Estimation by Maximum-Likelihood Texture Tracking. *IEEE Journal of Selected Topics in Signal Processing*, 5(3):398–407, June 2011.
- [9] Frédéric Herman, Brian Anderson, and Sébastien Leprince. Mountain glacier velocity variation during a retreat/advance cycle quantified using sub-pixel analysis of ASTER images. *Journal of Glaciology*, 57(202):197–207, 2011.
- [10] Andreea Julea, Nicolas Méger, Philippe Bolon, Christophe Rigotti, Marie-Pierre Doin, Cécile Lasserre, Emmanuel Trouvé, and Vasile N. Lazarescu. Unsupervised Spatiotemporal Mining of Satellite Image Time Series Using Grouped Frequent Sequential Patterns. *IEEE Transactions on Geoscience and Remote Sensing*, 49(4):1417–1430, April 2011.
- [11] Carson Kai-Sang Leung. Mining uncertain data. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4):316–329, July 2011.
- [12] Nicolas Méger, Christophe Rigotti, and Catherine Pothier. Swap Randomization of Bases of Sequences for Mining Satellite Image Times Series. In *Machine Learning and Knowledge Discovery in Databases*, pages 190–205. Springer, 2015.
- [13] Muhammad Muzammal and Rajeev Raman. Mining sequential patterns from probabilistic databases. *Knowledge and Information Systems*, 44(2):325–358, Aug 2015.
- [14] Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Emmanuel Trouvé, and Noël Gourmelen. Handling coherence measures of displacement field time series: application to greenland ice sheet glaciers. In *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)*. IEEE, to be published.
- [15] NunaGIS. Asiaq’s maps: Topographic map 1:500000 and 1:250000 (accessed 2017-07-05).
- [16] J. Pei, J. Han, and W. Wang. Constraint-based sequential pattern mining: the pattern-growth methods. *Journal of Intelligent Information Systems*, 28(2):133–160, 2007.

- [17] Y. Pericault, C. Pothier, N. Méger, C. Rigotti, F. Vernier, H. T. Pham, and E. Trouvé. A swap randomization approach for mining motion field time series over the Argentiere glacier. In *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (Multi-Temp)*, pages 1–4, July 2015.
- [18] Yoann Pitarch, Dino Ienco, Elodie Vintrou, Agnès Bégué, Anne Laurent, Pascal Poncelet, Michel Sala, and Maguelonne Teisseire. Spatio-temporal data classification through multidimensional sequential patterns: Application to crop mapping in complex landscape. *Engineering Applications of Artificial Intelligence*, 37:91–102, January 2015.
- [19] Fanny Ponton, Emmanuel Trouvé, Michel Gay, Andrea Walpersdorf, Renaud Fallourd, Jean-Marie Nicolas, Flavien Vernier, and Jean-Louis Mugnier. Observation of the Argentire Glacier Flow Variability from 2009 to 2011 by TerraSAR-X and GPS Displacement Measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(8):3274–3284, August 2014.
- [20] D. J. Quincey, N. F. Glasser, S. J. Cook, and A. Luckman. Heterogeneity in Karakoram glacier surges. *J. Geophys. Res. Earth Surf.*, 120:1288–1300, 2015.
- [21] R. Rosenau, M. Scheinert, and R. Dietrich. A processing system to monitor Greenland outlet glacier velocity variations at decadal and seasonal time scales utilizing the Landsat imagery. *Remote Sensing of Environment*, 169:1–19, November 2015.
- [22] D Scherler, S Leprince, and M Strecker. Glacier-surface velocities in alpine terrain from optical satellite imagery - Accuracy improvement and quality assessment. *Remote Sensing of Environment*, 112(10):3806–3819, October 2008.
- [23] Tazio Strozzi, Alexei Kouraev, Andreas Wiesmann, Urs Wegmller, Aleksey Sharov, and Charles Werner. Estimation of Arctic glacier motion with satellite L-band SAR data. *Remote Sensing of Environment*, 112(3):636–645, March 2008.
- [24] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.
- [25] Andrew J. Tedstone, Peter W. Nienow, Noel Gourmelen, Amaury Dehecq, Daniel Goldberg, and Edward Hanna. Decadal slowdown of a land-terminating sector of the Greenland Ice Sheet despite warming. *Nature*, 526(7575):692–695, October 2015.
- [26] C. Vincent and L. Moreau. Sliding velocity fluctuations and subglacial hydrology over the last two decades on argentièrè glacier, mont blanc area. *Journal of Glaciology*, pages 1–11, 2016.

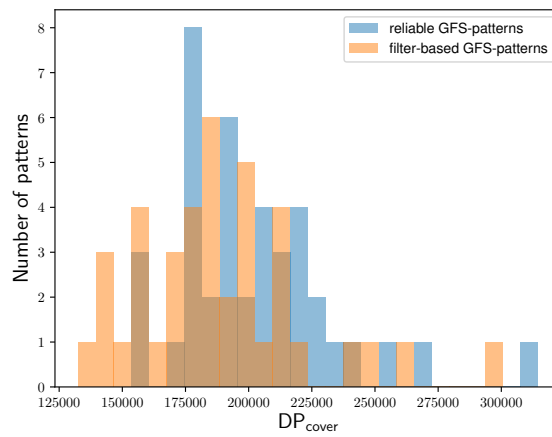




(a) Surface threshold setting.

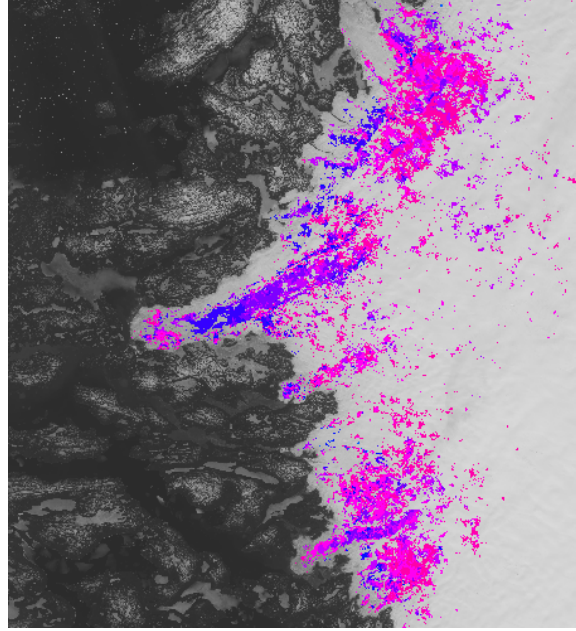


(b) Reliability threshold setting.

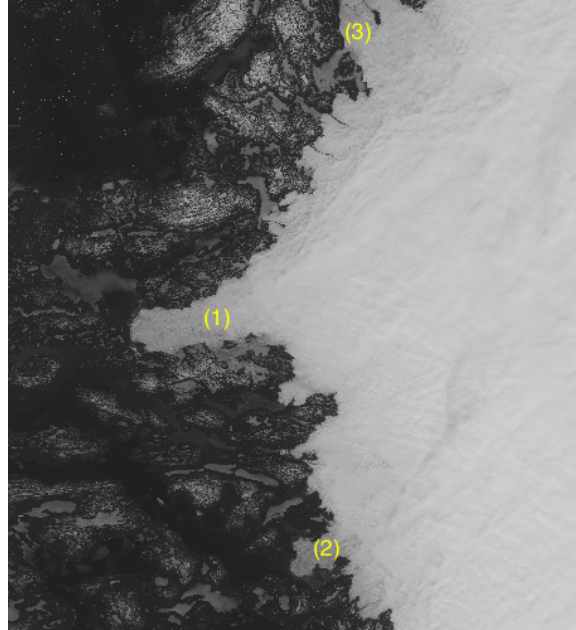


(c) Number of extracted patterns vs. individual  $DP_{cover}$  measures.

Figure 3: (a) Surface threshold setting. (b) Reliability threshold setting. (c) Number of the extracted patterns vs. individual  $DP_{cover}$  measures.

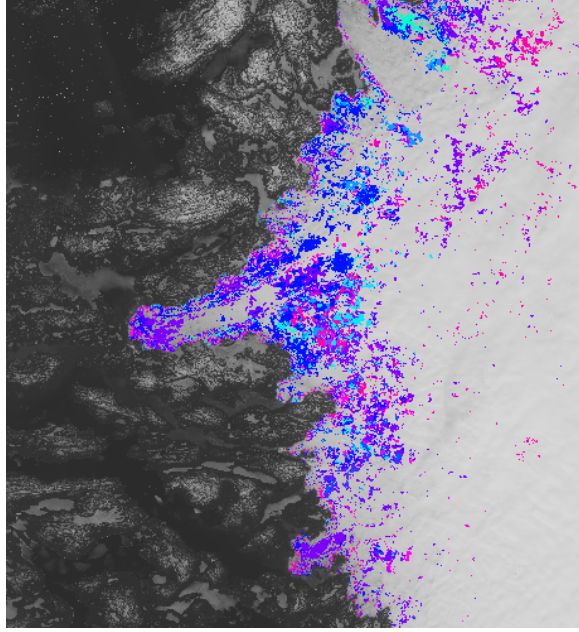


(a)  $3 \rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1$ .

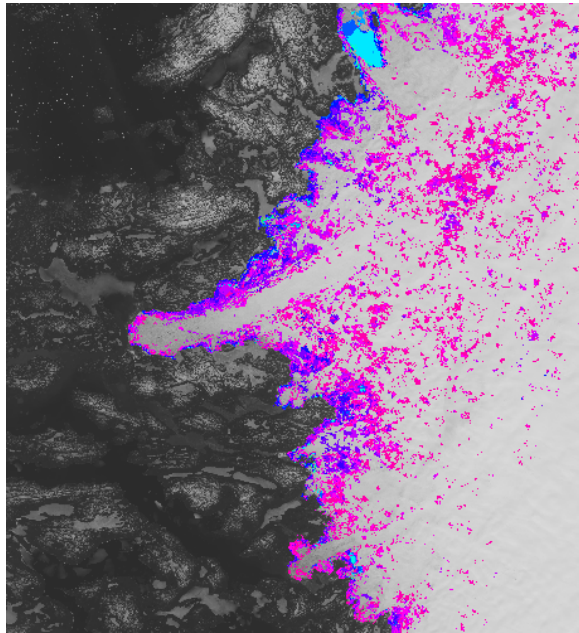


(b) Glacier locations.

Figure 4: (a) STL-maps for pattern  $3 \rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1$ . (b) Three of the main glaciers in the study area of the Greenland Ice Sheet: (1) Nordenskjöld glacier, (2) Polonia glacier and (3) Sarqardliup sermia and Alangordliup sermia (names from [21] and [15]) .



(a)  $3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 1$ .

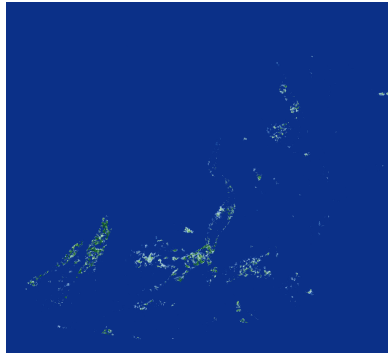


(b)  $1 \rightarrow 1 \rightarrow 3 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1$ .

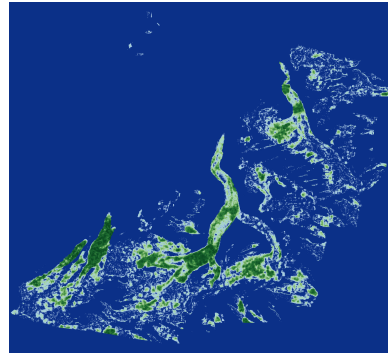
Figure 5: STL-maps for the patterns (a)  $3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 1$  and (b)  $1 \rightarrow 1 \rightarrow 3 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1$ .



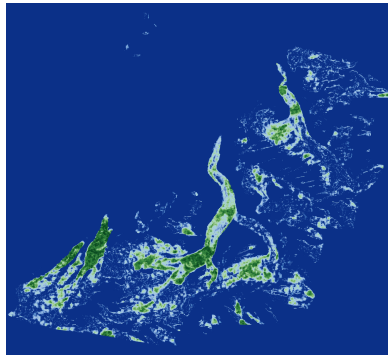
Figure 6: STL-map color scale: from 1985 in red to 2013 in magenta decomposed in 20 timestamps.



(a) Minimum confidence.



(b) Maximum confidence.

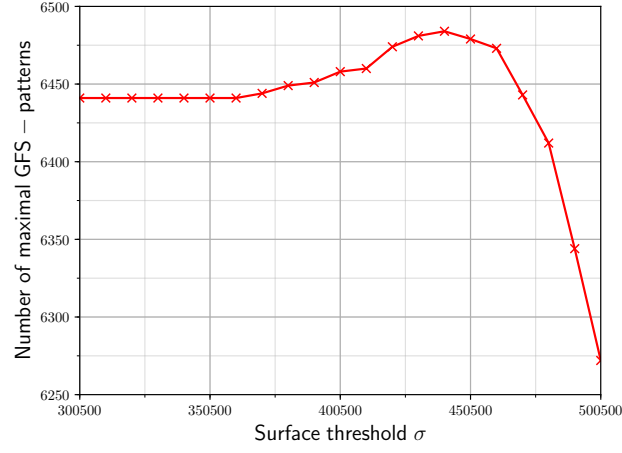


(c) Average confidence.

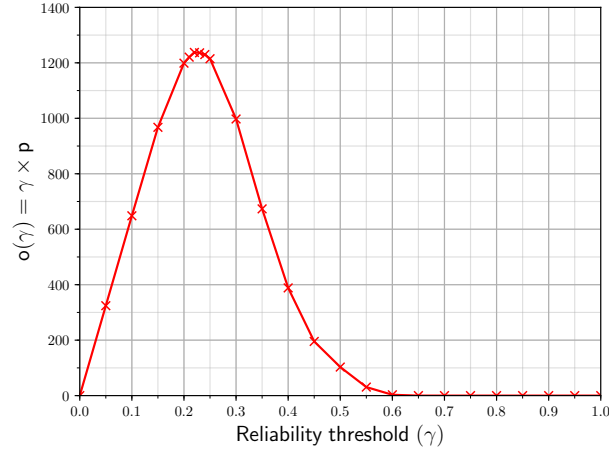


(d) Color scale.

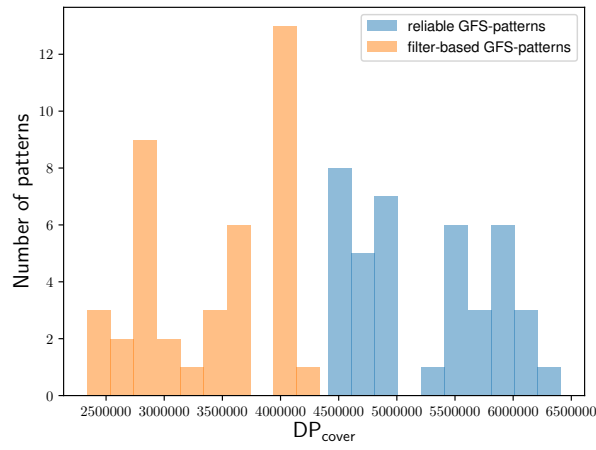
Figure 7: Confidence of each location over the series: (a) minimum, (b) maximum, (c) average. (d) Confidence color scale: from 0 (dark blue) to 1 (dark green).



(a) Surface threshold setting.

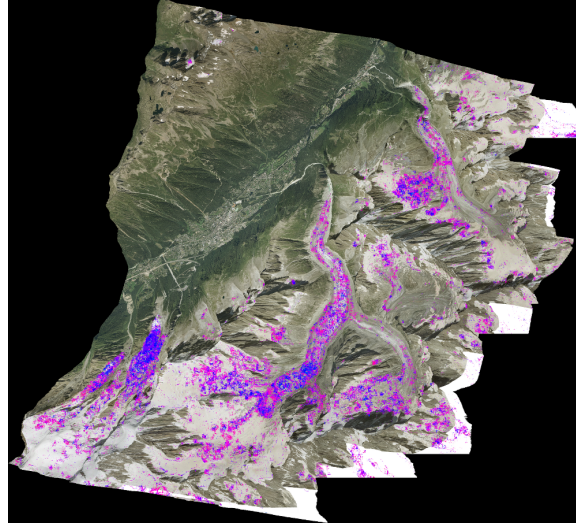


(b) Reliability threshold setting.

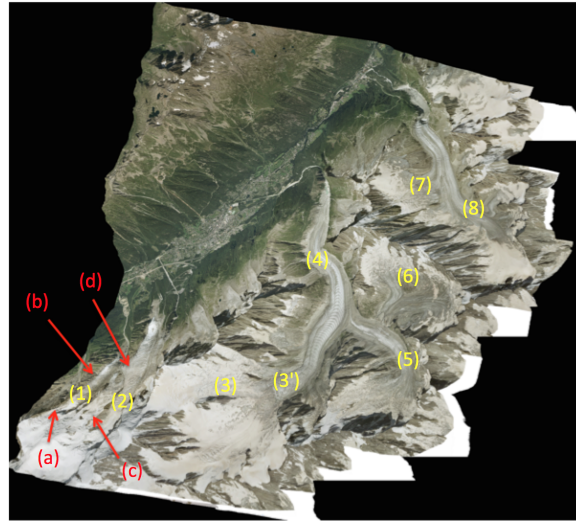


(c) Number of the extracted patterns vs. individual  $DP_{cover}$  measures.

Figure 8: (a) Surface threshold setting. (b) Reliability threshold setting. (c) Number of the extracted patterns vs. individual  $DP_{cover}$  measures.



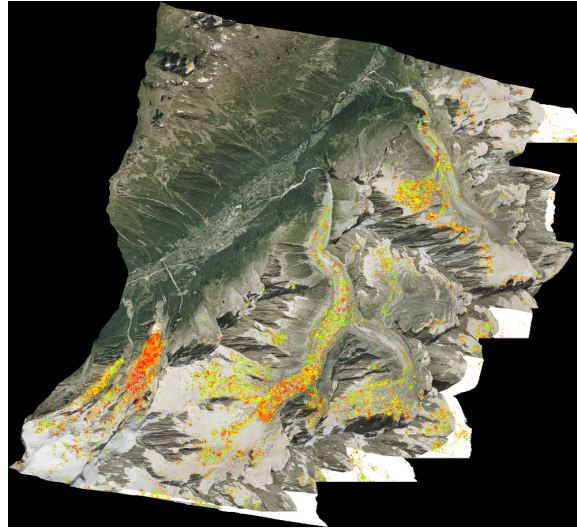
(a)  $3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$ .



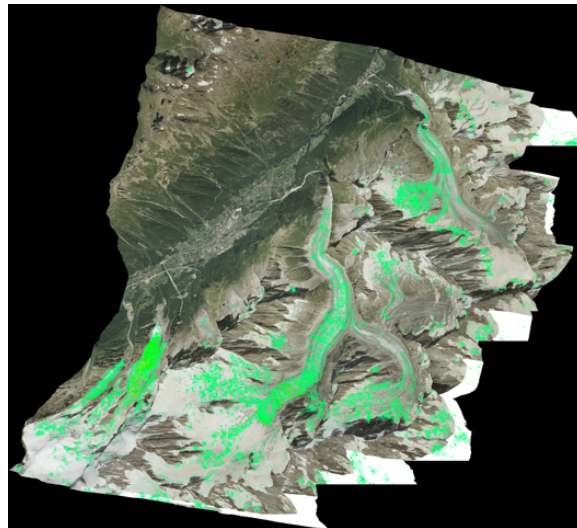
(b) Glacier locations.

Figure 9: (a) STL-maps of pattern  $3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$  on RGB composite background projected in radar geometry. (b) Glacier locations in radar geometry: (1) Tacconnaz, (2) Bossons, (3) glacier of the Géant, (3') seracs of the Géant, (4) Mer de Glace, (5) Leschaux, (6) Talèfre, (7) Rognons glacier and (8) Argentière. In red: (a) head of Tacconnaz, (b) 2000 m from head of Tacconnaz, (c) head of Bossons and (d) 2000 m from head of Bossons.





(a)



(b)

Figure 10: Occurrence dates of elements of pattern  $3 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$ : (a) occurrence dates of the first symbol 3, (b) occurrence dates of the last symbol 1.



Figure 11: STL-map color scale: red to green corresponding to May 2009 to October 2009, blue to magenta corresponding to May 2011 to September 2011.