A Cyclone Identification Algorithm with Persistent Homology and Merge-Tree

Masaru Inatsu¹, Hayato Kato², Yuta Katsuyama³, Yasuaki Hiraoka⁴, and Ippei Ohbayashi⁵
¹Faculty of Science, Hokkaido University, Sapporo, Japan
²Graduate School of Environmental Science, Hokkaido University, Sapporo, Japan
³Graduate School of Science, Hokkaido University, Sapporo, Japan
⁴Advanced Institute for Materials Research, Tohoku University, Sendai, Japan

Abstract

This paper addresses a cyclone identification algorithm with the super-level set filtration of the persistent homology together with the merge-tree reconstruction of data. Based on the information of peaks and saddles of the scalar field, the newly developed algorithm divides the analysis area into several homology classes, each of which satisfies the peak-to-saddle difference larger than a criterion that should be set in advance. Applied to the 850-hPa relative vorticity in the western North Pacific at 1200 UTC on 2 March 2013, 3 homology classes were found with the criterion of 100 × 10⁻⁸ s⁻¹ and 17 homology classes were found with the criterion of 50 × 10⁻⁸ s⁻¹. The merge-tree restructuring clarified the neighbour relation among homology classes. The result suggests that the weak criterion detected too much homology classes, some of which are small peaks inside of a single cyclone. The climatology feature density provides the Pacific storm track with the strict criterion. Finally, a possible way to extend toward cyclone tracking with the persistent homology is discussed.

(Citation: Inatsu, M., H. Kato, Y. Katsuyama, Y. Hiraoka, and I. Ohbayashi, 2017: A cyclone identification algorithm with persistent homology and merge-tree. SOLA, 13, 214–218, doi: 10.2151/sola.2017-039.)

1. Introduction

Extratropical cyclones play an important role in atmospheric dynamics through the transport of momentum, heat, and moisture from low to high latitudes. They often develop rapidly in the strong baroclinic zone and bring many natural hazards such as gale damage, river flood, landslide, and ship wreck. Projecting extratropical cyclones onto a two-dimensional metric of (potential) vorticity, sea-level pressure (SLP), wind speed, clouds, or precipitation, a single cyclone in a snapshot is identified as a connected component with its local extremum as the centre. Based on meteorological data with a discrete temporal interval during a period, therefore, the cyclone tracking is practically with some way realised by relating one cyclone in a time frame to another in the next time frame. A set of cyclone identification and cyclone tracking provide not only the preferable storm passage, say storm tracks (Chang et al. 2002), but also the life time, the growth rate, the translation velocity, and the preferable genesis and lysis points, all of which enable us to diagnose the climatology, the interannual variability, and a possible change in the near future (Neu et al. 2013; Ulbrich et al. 2013).

A careful subjective analysis based on daily weather charts may be one way to make such diagnostics, but it is neither realistic for a massive amount of simulated or reanalysis data, nor excluding some subjective choices by an analyst. An objective, automatic algorithm is therefore needed for the diagnostics, but some choices in metrics, and criteria of intensity, life time, and translation are possible in an accurate identification and tracking of three-dimensional moving, deforming objects simultaneously residing therein from data with rather insufficient temporal resolution. Therefore, tens of algorithms have been advocated for a couple of decades. A pioneering work in 1990s simply used minima of SLPs as cyclone centres and tracked them by nearest neighbouring (Murray and Simmons 1991; Sinclair 1994; Sinclair 1997). A sophisticated algorithm was then developed with an idea of connected component labelling (CCL) in identification and minimising a cost function in tracking (Hoskins and Hodges 2002). The application of contour dynamics was also provided (Wernli and Sprenger 2007). However, in any algorithms provided as above, there are many tuning parameters to be fit with a subjective intuition amongst synoptics. To resolve the problem, Inatsu (2009) extended the CCL technique and proposed the neighbour enclosed-area tracking (NEAT) algorithm, in which cyclone tracking is performed for the overlapping CCL images in adjacent time frames. This algorithm has a choice of metrics with only 4 parameters: CCL boundary level, CCL and overlapping areas, and life time. They recommended the use of relative vorticity later (Inatsu and Amada 2013; Satake et al. 2013).

On the other hand, applied mathematicians have recently developed some techniques to characterise multi-scale topological features such as connected component, ring, and cavity, embedded in finite-number data. Persistent homology (PH) is one of them, in which topological features in multi-dimensional space can be estimated from scattered points or estimated from gridded scalar field. For the former type of datasets, by sweeping a parameter of the radius of a uniformly inflated ball from a point, the PH identifies a topological feature robustly observed in a wide range of the parameter. For the latter type of datasets, similarly, by gradually descending from the maximum of data, called the superlevel set filtration¹, or by ascending from the minimum of data, called the sublevel set filtration, the PH identifies a topological feature above or below the level of the scalar field (Edelsbrunner and Harer 2010). The previous studies succeeded in capturing the material structure that had not been systematically recognised with a conventional technique in material science (Hiraoka et al. 2017; Saadatfar et al. 2017) with the former type of data. The cyclone identification discussed in this paper is the latter type, because it can be mathematically reworded as an optimal search of isolated features around local extremas of SLP or vorticity data.

The paper aims to develop the cyclone identification method by applying the PH to the gridded dataset of relative vorticity in Northern Hemisphere (NH) winter. Similar to the NEAT algorithm (Inatsu 2009) and Hodges’ technique (Hodges 1994), the PH can also identify connected components in a two-dimensional image as cyclones. However, different from the previous studies, the PH is able to provide the robustness of connected components and the neighbour relationship among them by focusing on not only extrema, i.e., peaks, but also saddles of the scalar field². This is based on a fundamental idea of topology, not geometry, concentrating properties of space that are preserved under continuous deformations. This paper focuses only on connected components in PH, say 0-th persistent homology, and is posed as an interim

¹ This terminology is used in topology as a sequence of geometric objects that are monotonically increasing with a parameter.
² This is usually called lifetime of homology classes in topology, but we avoid using this term not to confound it with the lifetime of an atmospheric cyclone.
2. Data and method

The data used are the 6-hourly JRA55 reanalysis dataset archived by the Japan Meteorological Agency (Kobayashi et al. 2015). The input to the algorithm is 850-hPa relative vorticity; this is estimated from a spatial finite difference of zonal and meridional winds and is not a standard product from the reanalysis. The horizontal grid mesh is 1.25° in longitude by 1.25° in latitude. The analysis is limited to the domain of 100°E–180° by 20°N–70°N in NH winter between November and the following March from 2006/07 to 2015/16.

The PH of the super-level filtrations is applied to the data for every time frame. Here the super-level set filtration is explained with hypothetical 5 × 5 data in Fig. 1. In this example, the level is swept from the maximum of data, 51, down to the minimum 0. The PH algorithm identifies simply-connected areas in the data that satisfy the quantity exceeding the level. First the level is set to 51, and a simply-connected area appears. We chose the Neumann neighbourhood in spatial connection searching only toward up/down and left/right directions, but if it is replaced with the Moore neighbourhood adding with 4 oblique directions, all the results that we show will not change substantially (not shown). When the threshold level goes down, other areas emerge in the right side and in the bottom left corner. In the level down to 22, these two areas are connected. The left-corner area then has a peak of 37 and a saddle of 22. Moreover, this area is unified with the area around the upper left corner at the level of 19. The PH algorithm identifies the right-side area with a peak at 45 and a saddle at 19. In this super-level set filtration, we retain a homology class that contains the information of a peak and saddle values and their positions for all connected components identified in the PH. As the cyclone can be defined as the connected component characterised by the higher relative vorticity than its surroundings, we pick up connected components that are persistently found in the level sweeping in the PH. This process corresponds to ignoring homology classes close to the diagonal in the persistence diagram. Here, in the persistence diagram, each homology class is represented on the 2-dimensional plane as a pair of birth (peak) and death (saddle) threshold values at which the corresponding class appears and disappears, respectively. The criterion is however arbitrary and we check its sensitivity to results in this study.

Homology classes extracted in the above process is reorganised by merge-tree structure as in Fig. 2. Each branch corresponds to a homology class from peak to saddle. When two homology classes are coalesced at a level, a homology class with a higher peak absorbs another homology class with a lower peak. The tree structure hence shows a hierarchy of homology classes. In the example of Fig. 1, a homology class with the peak of 51 initially appeared at the left side is the most robust homology class in the data. This reconstruction of the connected components to the merge-tree structure clarifies the neighbourhood among the homology classes detected in the superlevel set filtration.

3. Results

We explain the cyclone identification by the PH with an example of 850-hPa relative vorticity, at 1200 UTC on 2 March 2013, which involved an isolated, distinct peak in northeast of Japan. The super-level set filtration of the PH identifies a cyclone system as a homology class by searching the domain with the peak-to-saddle difference larger than a criterion. When we set the criterion to 100 × 10⁻⁶ s⁻¹, there are 3 homology classes from this two-dimensional image (Fig. 3a). The PH successfully extracted the highest peak labelled 1, and the homology class covers not only a cyclone system but all surrounded areas that do not have any clear peaks. We found another peak around the Aleutian Islands, which is much weaker than the peak labelled 1. The PH can detect it, because of a large difference between peak and saddle straddling homology classes #1 and #2. The third homology class has a very weak peak on the eastern boundary of the analysis domain, but it is an independent homology class in light of the peak-to-saddle difference larger than 100 × 10⁻⁶ s⁻¹. The saddle of homology class #3 crosses the homology class #2. Therefore we can summarise the information on peaks and saddles to the merge-tree graph (Fig. 3b).

By descending the criterion, the number of homology classes rapidly increases (Fig. 3c). When we set the peak-to-saddle criterion at 50 × 10⁻⁶ s⁻¹, there are 17 homology classes in the analysis domain. The highest peak is also labelled 1 as the case of 100 × 10⁻⁶ s⁻¹ criterion, but this case relaxes to detect a less distinct peak like ones labelled 2 and 3 (Fig. 3d). These are definitely local maxima of relative vorticity, but they should be grouped as a single comma-shaped cyclone system according to synopticians’
intuition. In the merge-tree structure (Fig. 3e), these independent homology classes have the common branch with merging at a relatively higher value of saddle points. This weaker criterion detects very weak peaks such as #11, which is not generally identified as a cyclone. Moreover, the effect of boundary is so crucial that the homology classes #6, 8, 9, and 17 have the peak on the boundary of the analysis domain.

Compiling the peak position of homology classes, we can draw the graph of the feature density that means how often the peak comes out in the images throughout the analysis period. In the $100 \times 10^{-6}$ s$^{-1}$ criterion, the climatological feature density is large in the western North Pacific basin (Fig. 4a). The axis of large density extends from eastern Japan to south of Alaska. This corresponds to the Pacific storm track region, and is consistent with other Lagrangian tracking studies (Hoskins and Hodges 2002; Murray and Simmonds 1991) and day-to-week variations (Blackmon 1976; Chang et al. 2002). The weaker criterion of $50 \times 10^{-6}$ s$^{-1}$ also provides the high feature density along the Pacific storm track region. In contrast with the strict criterion, it emphasises small cyclone centres that potentially grow under the baroclinic environment. It is also interesting that the southern coast of Japan is characterised by relatively large density probably due to many small-scale eddies or an unclear peak inside of cold fronts, which can be linked with a preferable cyclogenesis location above the Kuroshio Current (Xie et al. 2002).

As have seen above, the cyclone identification by the PH is quite sensitive to the criterion of peak-to-saddle difference. Figure 4c displays the distribution of the number of connected components in the analysis domain for every $10 \times 10^{-6}$ s$^{-1}$ bin of peak-to-saddle difference criterion, based on the cyclone identifications during 5 cold months in 2006/07 to 2015/16. With the criterion at $150 \times 10^{-6}$ s$^{-1}$, about a half of cases falls into the detection of a single homology class. The PH detects only a few homology classes, even with the criterion down to $100 \times 10^{-6}$ s$^{-1}$. Lowering the criterion more, the detecting number of connected components rapidly increases; the most frequent number is 5 for $80 \times 10^{-6}$ s$^{-1}$, 10 for $50 \times 10^{-6}$ s$^{-1}$, and 21 for $30 \times 10^{-6}$ s$^{-1}$. For the stable cyclone identification, such a high sensitivity to the criterion is not desirable. On the other hand, a single homology class over the whole domain is a trivial solution of cyclone identification. Together with the problem that a lower criterion exaggerates small-scale eddies along the coast (cf. Fig. 3d), one of the optimal criteria that we recommend is $100 \times 10^{-6}$ s$^{-1}$ for the cyclone identification in the extratropics, though a different optimal value could be generally selected for a different case.

4. Concluding remarks

We have performed the cyclone identification as an application of the PH with an aid of the merge-tree structuring of the results. The method separates the analysis domain into a finite number of homology classes of which peak-to-saddle difference exceeds a criterion. In the analysis with an example of 850-hPa unfiltered relative vorticity at 1200 UTC on 2 March 2013, the PH successfully identified 3 homology classes with the criterion at $100 \times 10^{-6}$ s$^{-1}$ and 17 ones with $50 \times 10^{-6}$ s$^{-1}$. According to the cyclone identification throughout the analysis period, the number of connected components were highly sensitive to the criterion especially for that lower than $100 \times 10^{-6}$ s$^{-1}$. In order to maintain consistency with the synopticians’ intuition, the criterion at $100 \times 10^{-6}$ s$^{-1}$ is recommended for cyclone identification in the extratropics. Compared with many conventional methods, with which the same task as this paper can be achieved, this can be obtained with the application of PH originally in this paper, indicating the influential area of cyclones with a peak-to-saddle difference of ho-
mology classes. Moreover, the merge tree clarified the association amongst cyclone systems. These are the strength of this method, because any neighbour point tracking algorithms that have been mostly used since now only identify the cyclone centre and do not provide such additional information, which may be used in the cyclone tracking that we will implement in the forthcoming study.

We will stop the analysis with the PH and merge-tree structuring only applied to cyclone identification, and this method can possibly extend to cyclone tracking. Cyclone tracking is substantially the temporal connection procedure that relates a cyclone identified in a time frame to another cyclone identified in the next time frame. In context of this study, this procedure solves a corresponding relation of homology classes between adjacent time frames. In the remaining part of this paper, we will propose an idea to implement cyclone tracking with an example of 1200 UTC on 2 March 2013 (Fig. 3) and its next time frame at 6 hours later (Fig. 5) with an approach based on the application of neighbour point tracking.
Simply applying a nearest-neighbour search for a set of peak positions between adjacent time frames as a conventional technique, however, the tracking of a typical cyclone-front system embedded with multiple subsystems, say local extrema, often failed. For example, the nearest-neighbour method would not achieve the tracking of a developing cyclone from 1200 to 1800 UTC, if the peak point of the homology class #1 in a case of a criterion of $100 \times 10^6 \text{s}^{-1}$ jumped possibly longer than a reasonable critical distance that should be set in advance (Figs. 3a and 5a). This can be checked by the weak criterion of $50 \times 10^6 \text{s}^{-1}$, in which the PH identified the corresponding homology classes #1, #2, and #3 at 1200 UTC (Fig. 3d) and #1 and #2 at 1800 UTC (Fig. 5d). Hence the jumping of the local maximum is attributed to a peak switching in a double-peak data with one peak as high as another. This means that a homology class #1 is dominant at 1200 UTC (Fig. 3d) and then a homology class #3 becomes dominant at 1800 UTC (relabelled #1 in Fig. 5d). Though these peaks are changeable in time, the common saddle is quite robust because of a large saddle-to-saddle difference (See that the saddle between #1 and #3 is far from that between #10 and #1 in Fig. 3e; Similarly in Fig. 5e). The tree structure gives us a hint to overcome the problem in a simple nearest-neighbor searching. If we utilised the peak of the homology classes with a $50 \times 10^6 \text{s}^{-1}$ criterion as a potential peak in the next time frame, a nearest-neighbour searching could successfully grab the jumping of the peak in a homology class in the PH with a $100 \times 10^6 \text{s}^{-1}$ criterion. The saddle-to-saddle difference would be an additional parameter to group the branches of merger-tree.

Acknowledgments

The Japanese MaeT group encouraged our collaborating works. This study was partly supported by Grants-in-Aid for Scientific Research 26130201 and Social Implementation Project for Climate Change Adaptation Technology, funded by the Ministry of Education, Culture, Sports, Science, and Technology of Japan. Figures were drawn using Grid Analysis and Display System. Y. H. acknowledges the support by JST CREST Mathematics (15656429).

Edited by: K. Ito

References

Blackmon, M. L., 1976: A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. J. Atmos. Sci., 33, 1607–1623.

Chang, E. K. M.; S. Lee, and K. L. Swanson, 2002: Storm track dynamics. J. Climate, 15, 2163–2183.

Edelsbrunner, H., and J. Harer, 2010: Computational topology: An introduction. Am. Math. Soc., 241 pp.

Hiraoka, Y., T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishiura, 2016: Hierarchical structures of amorphous solids characterized by persistent homology. Proc. Natl. Acad. Sci., 113, 7035–7040.

Hodges, K. I., 1994: A general-method for tracking analysis and its application to meteorological data. Mon. Wea. Rev., 122, 2573–2586.

Hoskins, B. J., and K. I. Hodges, 2002: New perspectives on the Northern Hemisphere winter storm tracks. J. Atmos. Sci., 59, 1041–1061.

Inatsu, M., 2009: The neighbor enclosed area tracking algorithm for extratropical wintertime cyclones. Atmos. Sci. Lett., 10, 267–272.

Inatsu, M., and S. Amada, 2013: Dynamics and geometry of extratropical cyclones in the upper troposphere by a neighbor enclosed area tracking algorithm. J. Climate, 26, 8641–8653.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. J. Meteor. Soc. Japan, 93, 5–48.

Murray, R. J., and I. Simmonds, 1991: A numerical scheme for tracking cyclone centers from digital data. Part I: Development and operation of the scheme. Aust. Meteor. Mag., 39, 155–166.

Neu, U., M. G. Akperov, N. Bellenbaum, R. Benestad, R. Blender, R. Caballero, A. Coccozza, H. F. Dacre, Y. Feng, K. Fraedrich, J. Griefer, S. Gulev, J. Hanley, T. Hewson, M. Inatsu, K. Keay, S. F. Kew, I. Kindem, G. C. Leckebusch, M. L. R. Liberato, P. Lionello, I. I. Mokhov, J. G. Pinto, C. C. Raible, M. Reale, I. Rudeva, M. Schuster, I. Simmonds, M. Sinclair, M. Sprenger, N. D. Tilinina, I. F. Trigo, S. Ulbrich, U. Ulbrich, X. L. Wang, and H. Wernli, 2013: IMILAST – a community effort to intercompare extratropical cyclone detection and tracking algorithms. Bull. Amer. Meteor. Soc., 94, 529–547.

Saadatfar, M., H. Takeuchi, V. Robins, N. Francois, and Y. Hiraoka, 2017: Pore configuration landscape of granular crystallization. Nature Commun., 8, 13082.

Satake, Y., M. Inatsu, M. Mori, and A. Hasegawa, 2013: Tropical cyclone tracking using a neighbor enclosed area tracking algorithm. Mon. Wea. Rev., 141, 3539–3555.

Sinclair, M. R., 1994: An objective cyclone climatology for the Southern Hemisphere. Mon. Wea. Rev., 122, 2239–2256.

Sinclair, M. R., J. A. Renwick, and J. W. Kidson, 1997: Low-frequency variability of Southern Hemisphere sea level pressure and weather system activity. Mon. Wea. Rev., 125, 2531–2543.

Ulbrich, U., G. C. Leckebusch, J. Griefer, M. Schuster, M. Akperov, M. Y. Bardin, Y. Feng, S. Gulev, M. Inatsu, K. Keay, S. F. Kew, M. L. R. Liberato, P. Lionello, I. I. Mokhov, U. Neu, J. G. Pinto, C. C. Raible, M. Reale, I. Rudeva, I. Simmonds, N. D. Tilinina, I. F. Trigo, S. Ulbrich, X. L. Wang, H. Wernli and the IMILAST team, 2013: Are greenhouse gas signals of Northern Hemisphere winter extra-tropical cyclone activity dependent on the identification and tracking methodology? J. Geophys. Res. Atmos., 118, 11,399–11,416.

Xie, S.-P., J. Han, T. Tanimoto, W. T. Liu, H. Tokinaga, and H. Xu, 2002: Bathymetric effect on the winter sea surface temperature and climate of the Yellow and East China Seas. Geophys. Res. Lett., 29, 2228.

Manuscript received 1 September 2017, accepted 17 October 2017 SOLA: https://www.jstage.jst.go.jp/browse/sola/