Ilkogretim Online - Elementary Education Online, Year; Vol 20 (Issue 4): pp. 1754-1764
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.04.198

Topology and Image Analysis

AJIETHKUMAR U.S., RESEARCH SCHOLAR, DEPARTMENT OF MATHEMATICS, KARPAGAM ACADEMY OF HIGHER EDUCATION, COIMBATORE, TAMIL NADU, INDIA– 641021, ajiethkumar@gmail.com
Dr.K.KALIDASS, DEPARTMENT OF MATHEMATICS, KARPAGAMACADEMY OF HIGHER EDUCATION, COIMBATORE, TAMIL NADU, INDIA – 641021, dassmaths@gmail.com
Dr.K.BHUVANESWARI, ASSOCIATE PROFESSOR AND HEAD, DEPARTMENT OF MATHEMATICS, MOTHER TERESA WOMEN'S UNIVERSITY, KODAIKANAL, TAMIL NADU, INDIA – 624101, drkbmaths@gmail.com

Abstract: - A wide-scope of mathematical approaches is being used to the greater value in the intelligent data analysis. The basic mathematical concept about the shapes is Topology. This article shows how topology is appropriately reformulated to be valuable in the activities of intelligent data analysis.

Keywords: Topology, data analysis, Image Analysis

I. INTRODUCTION

An ultimate expressive study of shapes is termed as Topology. Bringing these ideas into the realistic practice is fairly challenging, because the conventional topology is an inexhaustible -precision concept, where as actual information stands equally finite with in range as well as quantified in the dimension of time and space. The arena of computer-based topology grew from this encounter [3,7].

Among the doctrines,in the concept of variable-resolution topology the data characteristics (like the quantity of components, holes, as well as the respective dimensions with varied precisions) are analyzed, and from the restrictive characteristics of those curvatures the topology is construed. This framework [17,18,19] is superlative for intelligent data analysis.

From the work of Cantor, connectedness as well as components in a set of data is assessed.

Definition-1:
A pair of points is defined as 'ϵ-connected'provided $\exists$a ϵ-chainlinkingthat, and$\forall$ points in aϵ-connected set is associated by a ϵ-chain.
Note-1:
The fundamental quantities [18] used to achieve this work are: $C(\epsilon)$ = | ϵ-connected components |,$D(\epsilon)$=Max(diameter), and $I(\epsilon)$ = | ϵ-singleton points |.

From [17,19], all the above mentioned three dimensions for an array concerningϵ values can be computed as well as the topological characteristics of the fundamental set S on the basis of their restrictive characteristics were deduced. The characteristics of C and D are evident when S is connected.

When $\epsilon \to \infty$, each and every point in Swere ϵ-connected, also $C(\epsilon) = 1$ then $D(\epsilon)$ is the largest diameter of S until $\epsilon \to$ the largest spacing between any two points in S, where $C(\epsilon)$ =2creatingtwo subsets as well as $D(\epsilon)$ narrows toward the largest of their diameters etc.

If$\epsilon \to$ the minimum spacing between any two points in S, then each point is a ϵ-connected module, $I(\epsilon) = C(\epsilon)$, and $D(\epsilon) \to 0$.

For any disconnected fractal S, the characteristics are almost the same but C and D reveal a staircase pattern with varying ϵ due to the data gaps' scaling.
For instance, though$\epsilon \to$ the maximum gap size in the center one-third region of a Cantor set, $C(\epsilon)$ doubles also $D(\epsilon)$ reduces by means of one-third, and so on repeating the scaling when $\epsilon \to$next-smallest gap size. The results of fractal dimensional relationships are derived, detailed, and demonstrated, and debated [18].

Pixelation also cause staircase effects as it quantifies spacing distances between any two points resulting in confusion that way the fundamental data set has severed fractal geometry.

Propositions of Discrete Geometry is used for the computer implementation of calculating these connectedness –The MST Fig.1(a) and (NNG) the Nearest Neighbor Graph.

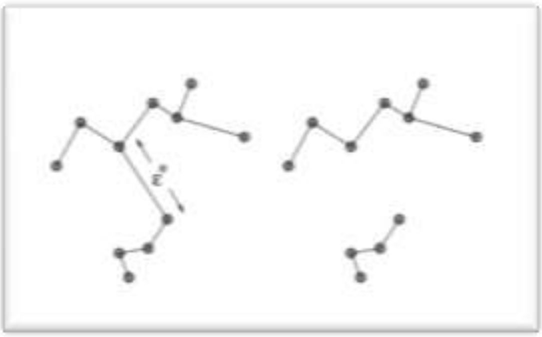
For MST creation, Prim's Algorithm[5] is applied, start with some nearest nbd point in the set, add the immediate point, and repeat till all points fall under the tree.

A directional graph with an edge from neighbors $x_A \rightarrow x_B$, provided$\{x_B: |x_B - x_A| < \delta\}$ is a NNG. Towards creating it, start through the MST, preserve the fleeting edge originating from every point. Mutually these algorithms can be easily executed in $R^{d.}$

The computational complexity in MST is commonly $O(N^2)$,where as in the plane, it is $O(N \log N)$, whereby N =| data points |.

To calculate C and I, just count the boundaries.
$C(\epsilon)> 1 +|$ MST boundaries$| >\epsilon$
$I(\epsilon) =|$ NNG limits $|>\epsilon$.

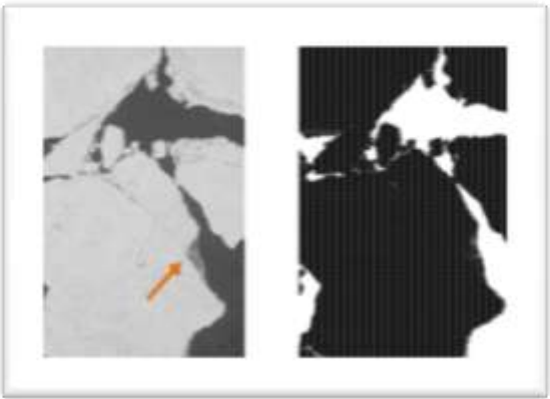
Fig.1. Computing Connectedness



(a) (b)

(a) The Minimal Spanning Tree (MST) connects the nearest neighbors in a dataset with its boundaries

(b) This set includes two ϵ-linked components if $\epsilon \lesssim \epsilon^*$



(c) (d)

(c) An in-flightrepresentation of a piece of Ocean; the orange pointer implies a bay abounding small icebergs

(d) The MST of the iceberg pixels in (c)

Note-2:
- Count NNG edges with multiplicity.
- The reason for multiplicity is because $x_A$becomes x_B's closest neighbor does not imply the reverse is true (i.e., provided some point $x_C \neq x_B$and $x_C > x_A$).
- The NNG and MST are needed to be constructed only once.
- All of the I and C details for variousϵ values are in their edge dimensions.

Remove the edges $>\epsilon$to spot the individual ϵ-elements. Normal computational geometry is used on ϵ-connected components to discover its diameter $D(\epsilon)$.

In intelligent data analysis, the trees and the statistics set in their edges are very valuable. Vertical jumps in $C(\epsilon)$ occurs at ϵvalues corresponding to gaps' sizes in a dataset. Diameter $D(\epsilon)$ is of clear value for relating the objects' sizes, as well as $I(\epsilon)$ used to sort out the noise.

In Fig.1(c), the MST can structure out as a white and black patch of Arcticocean. The tiny bay pointed by the arrow is abounding small ice floes in the dominant wind appears as grey. In Fig.1(d), the MST appears white, is small pouches of ice and water. All these quantities were mesh well through the sea-ice image MST capturing exactly with the number of floes resolved with instrument of precisions 1m, 5m, 10m, etc.

TheChapter 2 illustrations extend these ideas. MSTs can consolidate points into structures, creating applicability in clustering errands in pattern recognition [2, 8].

MST's branching structure can also be used in various possibilities like identifying orbit kinds indynamical methods[24] or on the way todetect discontinuities in the bubble-chamber tracks[25].

Chapter 3 covers the applications ofMST include Clustering, coherent-structure extraction,transverse 'hairs' creation on trees, noise addition and filtering out points by pruning the associated edges.

Through homology, using algebraic groupsassociation of an object in every dimension, other vital topological propertieslike number, shape, and sizeof the gaps in an object can bebrandedarithmetically.

Definition-2:

If every homology groupH_n provides various connected components at every order, n, and rank, k, then the Betti number, $b_n = k$ is represented as

$H_n(X) \cong \mathbb{Z}^k$, then $b_n(X)= k$

Note-3:

Geometric indications ofthe addedBetti numbers depend upon the dimensions of the space, the subject belongs to.

- A wide array of holes is represented by b_1 in two dimensions.
- A numerous open-ended tunnels are represented by $b_1$in 3D whereas enclosed voids in 3D are given by b_2.

The object's homology groups can be definedbydistinct geometricalillustration[14]:

- In 2D, subsets'orienteering.
- In 3D or higher dimensions, a simplicial complex.

Computational practice of the holesusesEdelsbrunner'sα-shape algorithm [10] thatcalculates the Bettinumbers on the basis oftopologicaltriangulations of data at various resolutions. Primarily, thepresumption flattens data by creating the α-nbdthat is the union of balls$B_\alpha(\epsilon)$ or $B(\epsilon; r)$centered at each data pointϵ.

When $\alpha \rightarrow \infty$, α-nbd is auniquely connected B. Since α reduces, theα-nbddiminishesandresolves betteroutline details in the data.

When$\alpha \rightarrow 0$, $B_\alpha(\emptyset)$fitswithin the data devoid ofdata points creating gap in α-nbd[6].

A α-nbd of any data set can have numerous holes thatcan be computed by varying α, and infer the topological properties [16,17].

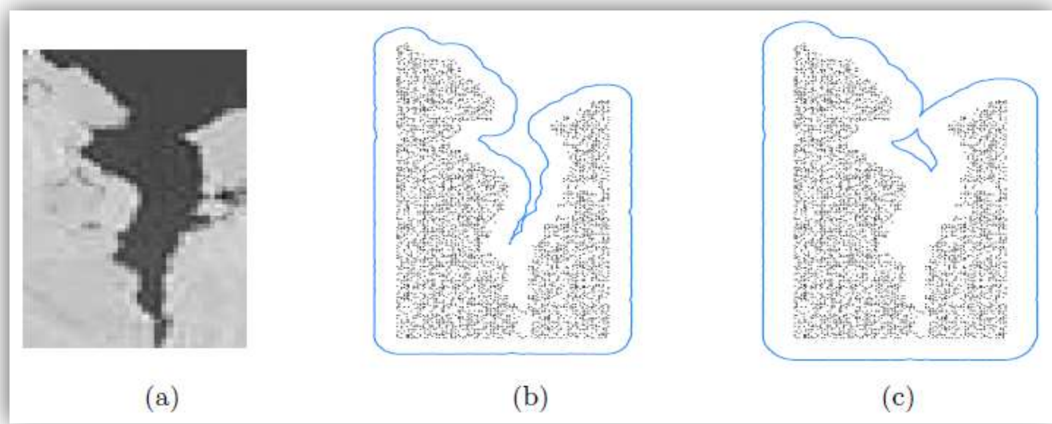
The Fig.2 shows thata set with no holes can end up having holes in the α-nbds caused by the geometry of the set. This fallacy can be resolved with a well-definedset-mapping of inside of its α-nbd.

Definition-3:

For all $\epsilon < \alpha$,$b_k(\epsilon, \alpha)$ is the holes' count in the α-nbdsuch that$\exists$holes in the ϵ-nbd. These are Persistent BettiNumbers (PBN).

[9, 19] defines PBN for sequences ofcomplexes and for α-shapes, incrementally providing better approximations.

Fig.2. Holes Calculation



α-nbds can findif (a) includes a bay, also if true, how extensive its opening is
(b) as well as (c) shows α-nbds of the icebergs in (a) forsome of the α's.

Note-4:
Anon-persistent holehindersperforming a spot-onvital topology diagnosisgiving geometric information onSet embedded in the plane. This is vital in coherent structureextraction.
In Fig.2,

- $b_1(\alpha) = 0$ if α <half the breadth of the bay's mouth,
- $b_1(\alpha) = 0$ if α > the principal radius of the bay's interior, and
- $b_1(\alpha) = 1$ if α ball fits within the bay, and does not fit inthe bay'sopening.

Note-5:
The 'α' value for the fake hole is exactly the half-the-width of bay'sopening.
If thehole was in the iceberg instead ofbay, then $b_1(\alpha)$ is a staircase function

- $b_1(\alpha) = 0$ when α > the maximum hole radius, and
- $b_1(\alpha) = 1$ when α < the maximum hole radius.

This coherent structurecharacterizingtechnique is pivot in intelligent data analysis of allgaps and holes of different shapes like channels,tunnels,ponds, peninsulas, etc.

II. EXTRACTIONOFCOHERENT STRUCTURE

In this chapter, three examples are used which demonstrate theadjustable-resolution topology approaches to discover coherent architectures in aerial footagesof glaciered ocean with ice floes, melt pond, and openwater leads (Fig.3).

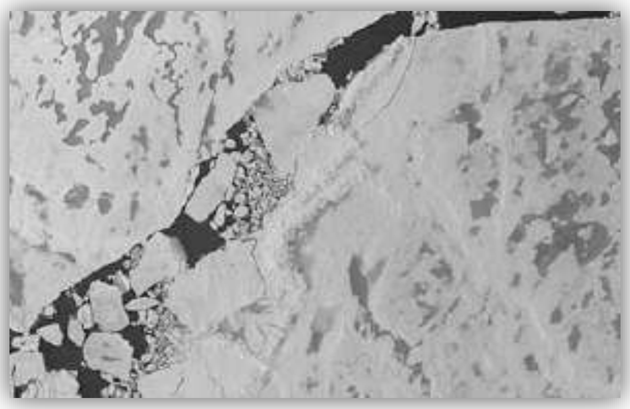


Fig.3. *The arctic glacier is composed of ice floes, melt ponds and open water leadsrespectively shown as white, grey and black.*

One of the application areas of topology is climatology, which studies the seasonal development of the coherent structuresof Arcticocean. Hence to process large number of images with terabytes of total data, automating topology-supported solutioning methods are enhanced.

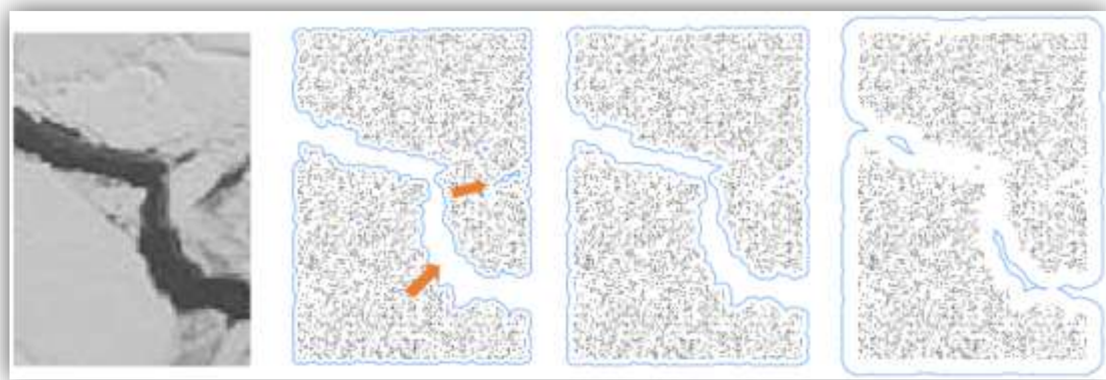
The three examples and contents of this chapterexplains

- discovering open water leadsthroughoutan icepack,
- differentiating parts of variousfrost and water concentrations, also
- learningthe distribution of the size and number of the melt ponds in ocean glaciers.

Identifying the size, position, and shape of ice floes and open water leads in the polar zone help the ships infinding a way through anintricateglacier part of the ocean. Theadjustable-resolution computational topology approaches from Chapter-1 arehelpful in resolving the issue.

Considering an n-Meter-wide (nMw) ship which must navigatethepatch from the right towards the left is revealed in Fig.4(a).

Fig.4. *Theα-shapes used to discover a way through theice pack.*



(a) (b) (c) (d)

The three α-nbds concerning icebergmeant by white pixelswithin (a) are (b), (c), and (d).

Resolving this situationby applying α shape procedure, with $\alpha = n/2$, involves assessment of the ice holes, hencebrink the data as well as shedthe water (dark pixels).

Fig.4(b)identifies all holes ≥nMw, resulting intwo holes highlighted with orange arrow. Then using computational geometric methods, determine the shortest hole that reaches from one end to the other end.

Fig.4(c) and (d) show, for $\alpha > (n/2)M$, thechannel shape as well as the holes relationship in that α-nbd, consecutivelydetermining somewater regions $\geq \alpha M$from ice.

Holistically, α-shapes evaluate the water and icemix in allregions. It also determines

- the distribution ofregions'sizes
- discover and illustrate regions of diverseice concentration.
- holesthat are water leads either in the water or the ice pixels.

Fig.5(a) depicts when the ice → solid, the waterchannel is narrow and determined as a minute α-hole in those ice pixels.

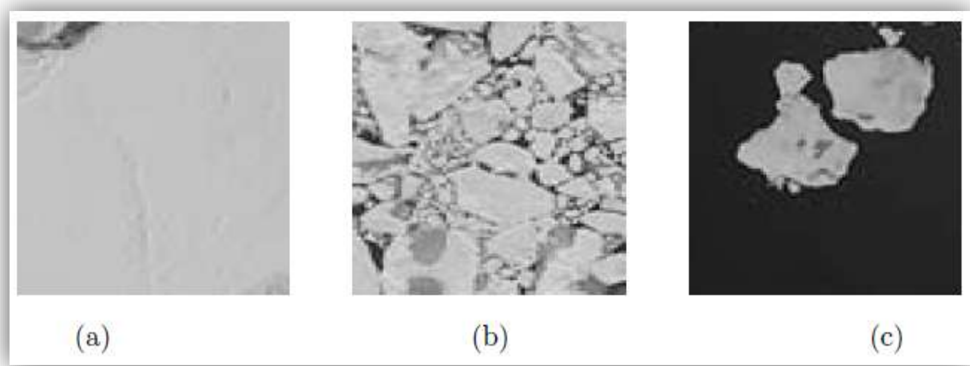
Fig.5(c) depicts when open waterarea is large and only a few ice-floes, α shape study of pixels representing water identifieshuge holes for a considerable number of α values.

Note-6:

For agroup of lesser α values, those α-nbds fills the gaps in the ice-floes resulting infake holes.

Fig.5(b) depicts the ice-floesdistribution witha wide α variety where a minute change identifies newholes in the water pixels of the ice-floes.

Fig.5.*α shape study to differentiate diverse degree of morphology of the water and ice.*



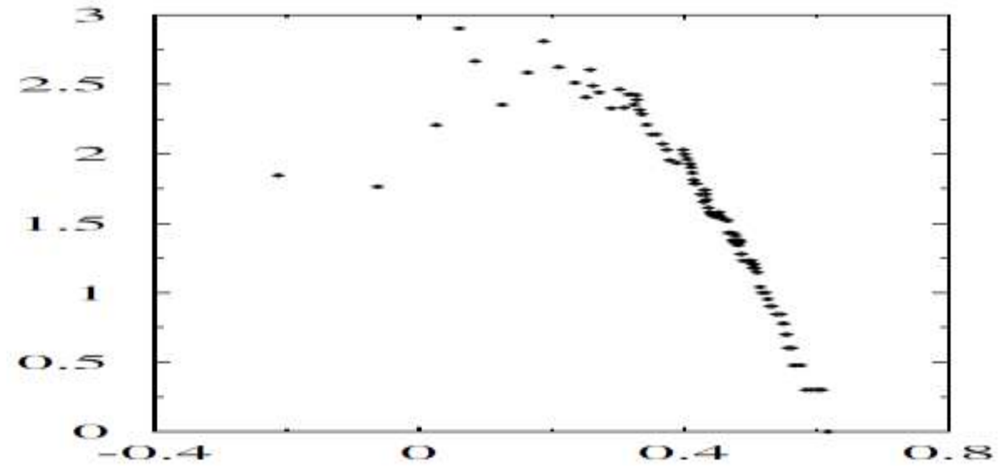
(a) mostly-solidified ice (b) waterwith different sizediceberg(c) water with a few largeicebergs

The temporal transformation of aforementioned albedo[1]of sea ice is apivotal factor that governs the power-law [15] distribution of the melt ponds' size.

α-shapes help in calculatingthe distributions by thresholding the datato mark the 'non-pond' ice or 'non-pond' water points black, and detecting holes in that data.

Fig.6.*The Computational topology as well as melt ponds*

log(Number of Holes) →



log(α) →

(a) The pond's lower-right portion's (of Fig.3) α-shape study

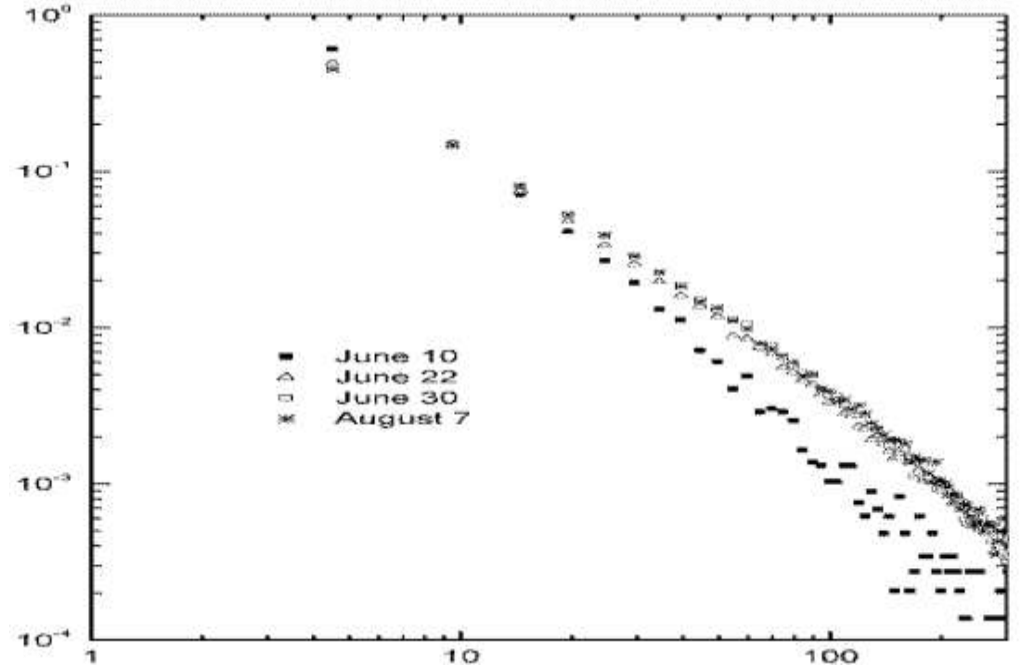
Fig.6(a) shows the results that substantiates thepower-law distribution hypothesis absolutely.
For the initial region <log α = 0.3, changes toresolve headlands as well asbays in that melt-ponds, and graphicaldesign is impartially undeviating.

Albedo is a measure of an extent of light which hits a surface being reflected even though it is not absorbed. White color has maximum Albedo, whereas darker colors have less albedo.

Fig.6(b) replicates theresultantgraph [15] for thecontrast.

Note-7:

- The power-law gradients in Fig.6(a) as well as (b) differsas the horizontal axis for plots are pond radius and pond area respectively.
- When the additional power of 2 is considered, the power-law slopes of Fig.6(a) and (b) match with a deviation < 8% because not all ponds are flawlessly circular.



Probability →

Pond Area (m^2) →

Fig.6.*(b) Corresponding figure from [15]*

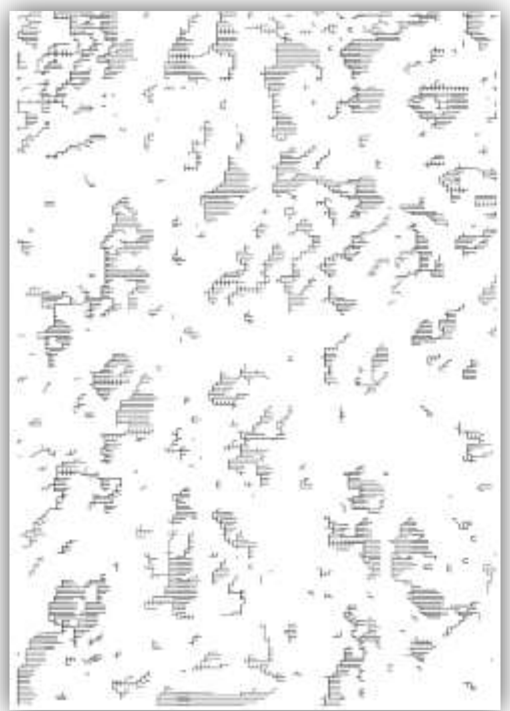


Fig.6.*(c) Componentsconnected in thedata of the pond*

Calculating the Pond's MST as shown inFig.6(c) demonstrates similar curves followed by an excerpt, the connected component parts for their areas as well as diameters.

To handle the vagueness, and identify spatial and temporal patterns of real coherent structures, the concepts of multiple-resolution topology, spatial statistics and artificial intelligence are syndicated.

Note-8:

- Computational topology generates cumulative measurements on the full dataset. Using spatial statistics alongside, with topospatial analysis sub-regions with different measurements are distinguished.
- Artificial intelligence (AI) approaches help in recognizing coherent facilities with a wide range of representations and techniques harmonized with topospatial study.

III. FILTERING

In reality, the data habituallyincludessome form of noise.

- Traditional linear filters can remove some of the noiseexcluding all due to ahectic behavior.
- Fourier-based filters can remove all the signals in a fewnoise-bandconsiderably [23].

Topology-based filtersare better in removing almost all noises of all bands. This filter treats noise as isolatedpoints, and these pointsare identified by variable-resolutiontopology. Fig.7 establishes thefundamental ideas.

Fig.7. The impacts of noise on data topology

(a) (b) (c)

(a) If adatasetis anundisputedillustration in dynamical systems where every point is perturbedby means ofa small amount (0.01probabilities) of noiseadded.
(b)The MST evidently shows the noise points.
(c) The MST reveals the identical data set upontreating witha topological strainingthatsnips off the 'furs' in the MST in (b).

The spanning tree fetchesashiftin the direction as well as in the edge-lengthof those noisepoints in the orbit.

- If |noise|>>the spacing between any two noisepoints,

then theconnectingMSTedges> original edge lengths.

- If |noise|<<the spacing between any two noisepoints,

thentheconnectingMST edges are shorter, and the noise points are unclear.

This kind of scale separation happens when two processes work on the data. So variable-resolution topologycan be used to find and eliminatethe noise points with the following steps:

i. Meticulously look forI(ϵ), i.e., first and second peaks, in the distribution of edge-length of MST
ii. The longest edge of MST of the noiseless data ∈I(ϵ) breakpoint, where almost all noisepoints, and very few non-noise pointsare ϵ-singleton.
iii. Trimeach and every MSTedge which exceed that range byeliminating the points upon their limits as inFig.7(c)

3.1. EXPERIMENTAL DATA ANALYSIS AND METRICS

The filter based on topology removed 1068out of the 1090 noise points (= 98.0% success) also300 out of 15712 non-noisy points (which is 1.9% false positive).

- Those resultswere better than some noise-reduction or linear filter techniques [1].
- Increasing the pruning lengthdecreases the false-positive rate.

- These success and false-positive rates can vary to some extent for various types as well asquantities of noise;howeverpersist close to the100% as well as 0%[20] respectively.
- This filter will not perturb the aggressiveLyapunovproponent [20].

Note-9:

- This method just removes noise points withoutinferring where each point should be and move it in that direction. This issue can be sorted out by using arithmetic average of the points on both sides of the main edge linkingtheinaccessible point to the remainingorbit.
- In rasterized descriptions noise does not move points insteadshades individual pixels.
- The MST is more effective in detecting distortion type noise[4].

IV. Conclusion

The mutable-resolution topology framework hasunmatchedvista for an intellectual data analysis (automateddiscovery and illustrationof orderly structures). Fig.6(b) was plotted by means of processing manually the 1000 images, that required roughly 25 man-days. Fig.6(a) was plotted through automation in a fewminutes time. From the experimental analysis, the topological strainingrecognizedalmost 100% of the noisepoints and with a very-lowfalse-positive rate. Few other topology-based filtering approaches [13] utilize algebraic topology to hypothesize a coarsely-grainedillustration of data as a filter to reduce the noise issue. Instead of algebraic topology, geometric topology is used to remove noise points to get the finer-grained results.

Only a very few papers came out on straining the orderly frames' topology in the data [12, 21], and are restricted to 2D or 3D. Many of the current commercial off-the-shelf tools that deal with geographic data systems[22] as well ascomputer digital image processing [4] are designed with an uncomplicated topological process like connectedness or adjacency which could work merely with 2D latticed data, also notan unreliable resolution. The variable-resolution topologyframework is algorithmically generic and computationally precise, and it works in all dimensions that include fractal dimensions as well.

References

1. H. Abarbanel. Analysis of Observed Chaotic Data. Springer, 1995.
2. D. Ballard and C. Brown. Computer Vision. Prentice-Hall, 1982.
3. M. Bern and D. Eppstein. Emerging challenges in computational topology, 1999.
4. E. Bradley, V. Robins, and N. Rooney. Topology and pattern recognition, 2004.
5. T. Cormen, C. Leiserson, R. Rivest, and C. Stein. Introduction to Algorithms. The MIT Press, 2001. pp 570–573.
6. C. Delfinado and H. Edelsbrunner. An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. Computer Aided Geometric Design, 12:771–784, 1995.
7. T. Dey, H. Edelsbrunner, and S. Guha. Computational topology. In B. Chazelle, J. Goodman, and R. Pollack, editors, Advances in Discrete and Computational Geometry. American Math. Society, Princeton, NJ, 1999.
8. R. Duda and P. Hart. Pattern Classification. Wiley, New York, 1973.
9. H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. In IEEE Symposium on Foundations of Computer Science, pages 454–463, 2000.
10. H. Edelsbrunner and E. M¨ucke. Three-dimensional alpha shapes. ACM Transactions on Graphics, 13:43–72, 1994.
11. U. Fayyad, D. Haussler, and P. Stolorz. KDD for science data analysis: Issues and examples. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 1996.
12. H. Li. Identification of coherent structure in turbulent shear flow with wavelet correlation analysis. Transactions of the ASME, 120:778–785, 1998.
13. K. Mischaikow, M. Mrozek, J. Reiss, and A. Szymczak. Construction of symbolic dynamics from experimental time series. Physical Review Letters, 82:1144–1147, 1999.
14. J. Munkres. Elements of Algebraic Topology. Benjamin Cummings, 1984.
15. D. Perovich, W. Tucker, and K. Ligett. Aerial observations of the evolution of ice surface conditions during summer. Journal of Geophysical Research: Oceans, 10.1029/2000JC000449, 2002.
16. V. Robins. Towards computing homology from finite approximations. Topology Proceedings, 24, 1999.
17. V. Robins. Computational Topology at Multiple Resolutions. PhD thesis, University of Colorado, June 2000.

18. V. Robins, J. Meiss, and E. Bradley. Computing connectedness: An exercise in computational topology. Nonlinearity, 11:913–922, 1998.
19. V. Robins, J. Meiss, and E. Bradley. Computing connectedness: Disconnectedness and discreteness. Physica D, 139:276–300, 2000.
20. V. Robins, N. Rooney, and E. Bradley. Topology-based signal separation. Chaos,4:881–892, 1993.
21. P. Saha and A. Rosenfeld. The digital topology of sets of convex voxels. Graphical Models, 62:343–352, 2000.
22. S. Shekhar, M. Coyle, B. Goyal, D. Liu, and S. Sarkar. Data models in geographic information systems. Communications of the ACM, 40:103–111, 1997.
23. J. Theiler and S. Eubank. Don't bleach chaotic data. Chaos, 3:771–782, 1993.
24. K. Yip. KAM: A System for Intelligently Guiding Numerical Experimentation by Computer. Artificial Intelligence Series. MIT Press, 1991.
25. C. Zahn. Graph-theoretical methods for detecting and describing Gestalt clusters. IEEE Transactions on Computers, C-20:68–86, 1971.