Peer Reviewed Paper openaccess Special Issue on Chemometrics in Hyperspectral Imaging

When remote sensing meets topological data analysis

Ludovic Duponchel

LASIR CNRS UMR 8516, Université Lille 1, Sciences et Technologies, 59655 Villeneuve d'Ascq Cedex, France.
E-mail: ludovic.duponchel@univ-lille.fr ORCID: https://orcid.org/0000-0002-7206-4498

Hyperspectral remote sensing plays an increasingly important role in many scientific domains and everyday life problems. Indeed, this imaging concept ends up in applications as varied as catching tax-evaders red-handed by locating new construction and building alterations, searching for aircraft and saving lives after fatal crashes, detecting oil spills for marine life and environmental preservation, spying on enemies with reconnaissance satellites, watching algae grow as an indicator of environmental health, forecasting weather to warn about natural disasters and much more. From an instrumental point of view, we can say that the actual spectrometers have rather good characteristics, even if we can always increase spatial resolution and spectral range. In order to extract ever more information from such experiments and develop new applications, we must, therefore, propose multivariate data analysis tools able to capture the shape of data sets and their specific features. Nevertheless, actual methods often impose a data model which implicitly defines the geometry of the data set. The aim of the paper is thus to introduce the concept of topological data analysis in the framework of remote sensing, making no assumptions about the global shape of the data set, but also allowing the capture of its local features.

Keywords: remote sensing, hyperspectral imaging, topological data analysis, clustering, data models

Introduction

Remote sensing already has a rather long history and started around 1950.[1] It is obvious that instrumental developments have played and continue to play a crucial role in its ascendancy, but spectral data analysis has not been overlooked either. Moreover, we can say that the first domain always takes advantage of the progress made in the second one and vice versa. Many multivariate analysis methods have been and are being developed in order to explore remote sensing data sets. However, the main challenge of hyperspectral imaging (HSI) remains image classification, which is the process of grouping pixels into spectrally similar clusters. This task is all the more difficult due to the fact that we have to manage data cubes often exceeding several hundred thousand pixels. Considering all the data analysis methods proposed in the literature, many different models are used to try to capture the intrinsic geometric nature of a data set.[2] The problem with many methods is that they implicitly define the geometry of the hyperspectral data set in the spectral space. As a consequence, such data models often over specify the global nature of a data set and give us a biased vision. It is in this sense that we have to propose new models without assumptions about the global shape of the data, and which are more focused on its local nature. Thus, we propose in this paper to explore the concept of topological data analysis (TDA) in the framework of remote sensing because it shares these characteristics.

Correspondence
Ludovic Duponchel (ludovic.duponchel@univ-lille.fr)

Received: 10 October 2017
Revised: 15 January 2018
Accepted: 31 January 2018
Publication: 6 February 2018
doi: 10.1255/jsi.2018.a1
ISSN: 2040-4565

Citation
L. Duponchel, "When remote sensing meets topological data analysis", *J. Spectral Imaging* **7,** a1 (2018). https://doi.org/10.1255/jsi.2018.a1

In mathematics, topologists usually study the shape of abstract objects. However, they discovered about ten years ago that topology could be used to explore large and complex data sets. It was the beginning of topological data analysis.[3,4] Many scientific domains have already taken advantage of TDA, including biology and, more specifically, genomics,[5–7] disease understanding[8–10] and neuroscience.[11–13] To a lesser extent, other fields have also benefited from the potential of the method, such as analytical chemistry,[14] physical chemistry,[15] material science[16] or, even more surprising, for the analysis of NBA basketball players' characteristics or the voting behaviour of the members of the US House of Representatives.[17] The diversity of both domains and data structures already suggest nice properties of TDA. In the first section of the paper, we will introduce topological data analysis. More precisely, we will explain how to obtain a topological network from a remote sensing data set and, finally, how to generate classification maps from it. The Results and discussion section will allow us to observe the behaviour of TDA for the analysis of a large, remote sensing data set.

Materials and methods

Image data set

The data set we have selected in this work is well-known in the remote sensing community. It corresponds to the hyperspectral image of the Washington DC Mall area in the USA acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor on 23 August 1995.[2,18,19] HYDICE is a push broom aircraft sensor system providing spectral information from 400 nm to 2500 nm with an approximate resolution of 10 nm. However, because of strong water absorption and noise, three spectral subdomains (1364.8–1407.2 nm, 1814.48–1926.31 nm and 2481.4–2506.0 nm) have been removed prior to data exploration, reducing the total number of spectral variables to 191. This data cube consists of 1280 × 307 pixels with a spatial resolution of 2.8 m. It is often used as a workbench data set considering its original form,[20–26] i.e. all the available pixels or just a cropped version of it.[26–41] In our case, the whole data set of almost 400,000 pixels has been used to show the good scalability of the proposed concept. In the literature, the data set is also used for purposes such as supervised classification,[18,19,21,25,27,29,30,33,34] clustering[20,23,24,26,31,32,36,38,39] and, to a lesser extent, in signal unmixing.[28,35,40] We will explore it, in our own turn, with topological data analysis in order to generate clusters highlighting different materials in the scene. Six to seven classes (buildings, trees, grass, water, roads, trays and shadow) are usually considered for this image. Ground truth data are available for this image, but the Google Maps website could also be used even if changes may have occurred since 1995. The image HYDICE can be downloaded from Purdue University's website (https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html).

Using topological data analysis in remote sensing

In topological data analysis, we generate a topological network which represents the shape of the explored data set. In this section, we will discover the way we construct such a network and how to obtain classification maps from it. Figure 1 gives us a global view of the concept. First, each pixel of the image is observed through what we call a "lens". In fact, all functions that produce a value from a pixel can be a lens. It can be selected from different domains such as statistics, geometry, chemometrics and much more. A lens value is then obtained for each pixel. Given the lens value scale, we divide it into overlapping subsets. In this way we partition the data set. We then apply a cluster analysis to each pixel's subset. The single linkage algorithm[42] is often used for this task. We are thus beginning to construct the topological network. Indeed, each cluster from each pixel's subset is represented by a node. Nodes are also connected by edges when they share at least one pixel. A hot colour node indicates a high number of pixels in the considered cluster. In a final step, the network is split into different groups of pixels according to the node's density and/or particular features of the network shape. Classification maps are then generated. For this work, the Ayasdi software platform (ayasdi.com, Ayasdi Inc., Menlo Park, CA, USA) was used for TDA. Final classification maps were generated with an additional Python script in connection with the Ayasdi Python SDK.

Results and discussion

The first task of TDA is to generate a topological network which represents the shape of the explored data set. Figure 2A gives its first representation using the

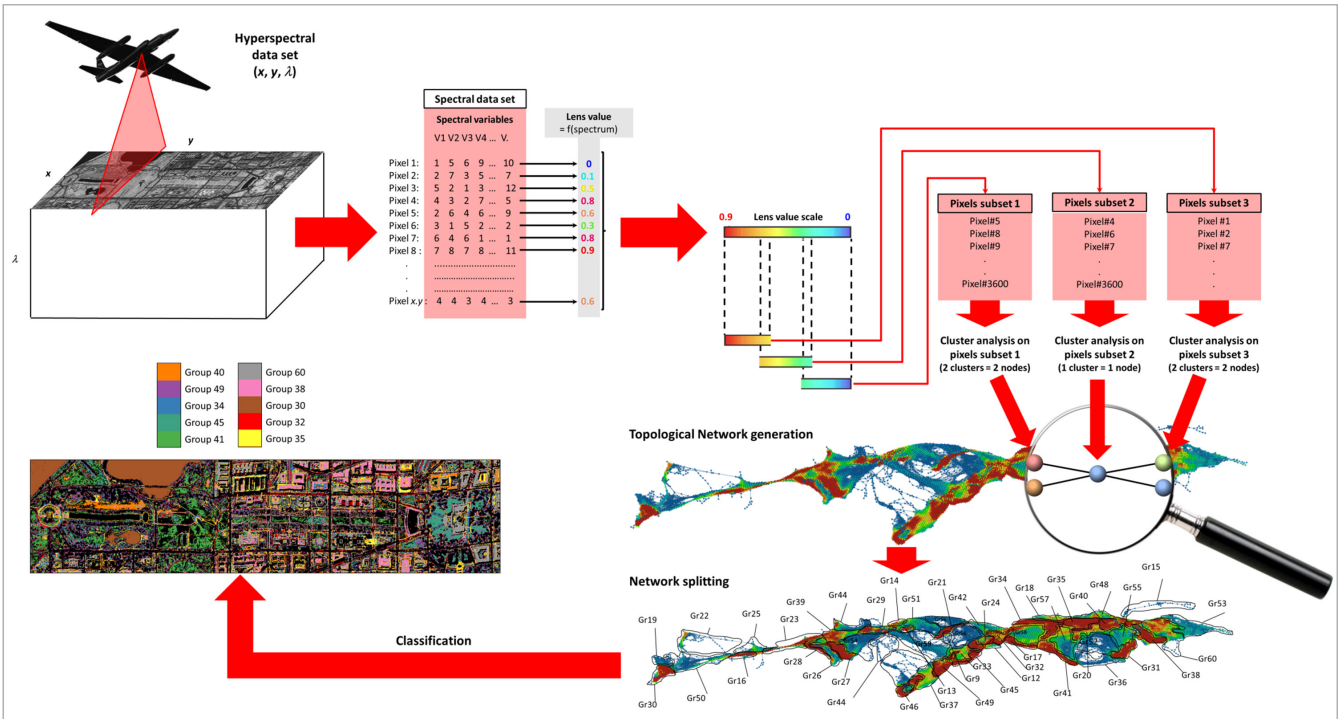

Figure 1. Topological data analysis in the framework of remote sensing.

so-called variance normalised Euclidian distance and the "Neighbourhood" lens. The variance normalised Euclidean (VNE) distance between two points takes into account that each column in the data set could have significantly different variance. Thus, the distance between two pixels (i.e. spectra) **x** and **y** is given by Equation 1:

$$VNE(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{v_i}} \qquad (1)$$

with x_i the reflectance at the spectral variable i of the pixel **x**, n the total number of variables and v_i the variance associated with the spectral variable i.

The "Neighbourhood" lens generates an embedding of high-dimensional data into two dimensions by embedding a k-nearest neighbours graph of the data. A k-nearest neighbours graph is generated by connecting each point to its nearest neighbours. This graph is embedded in two-dimensions using Ayasdi's proprietary graph layout algorithm. Looking at this network, we have to remember that a node in this wireframe represents a group of pixels with similar spectra. Moreover, an edge between two nodes indicates at least one pixel in common. Colouring of nodes is also a way to represent the number of pixels they contain. Considering the histogram provided in Figure 2A, the highest number of pixels per nodes (represented in red in this network) is 647, while the lowest is just one pixel (in dark blue). We discover that a topological network exhibits different geometric features, such as flairs and loops, but also different regions of varying densities. As indicated in the previous section of the paper, it is now necessary to split the network in order to generate different groups of pixels. Groupings have been obtained in an automatic way with agglomerative hierarchical clustering (AHCL). This "auto-grouping" algorithm is based on the nodes' colouring of the network. Basically, this algorithm works to collect connected nodes with similar colour values (i.e. pixel density) together into groups. It achieves this by treating edges that have very different colours on their ends as "weak" edges, while edges that connect nodes of the same colour are "strong". AHCL then uses Louvain modularity to identify clusters of nodes that are connected by these "strong" edges. The Louvain method for community detection is a well-known method to extract communities from large networks.[43] Figure 2B presents the topological network splitting. It is interesting to note that the proposed grouping could correspond to the one we may generate intuitively by hand. Table 1 provides information about groups with the number of nodes, the number of pixels and the percentage they represent with respect to the total number of pixel in

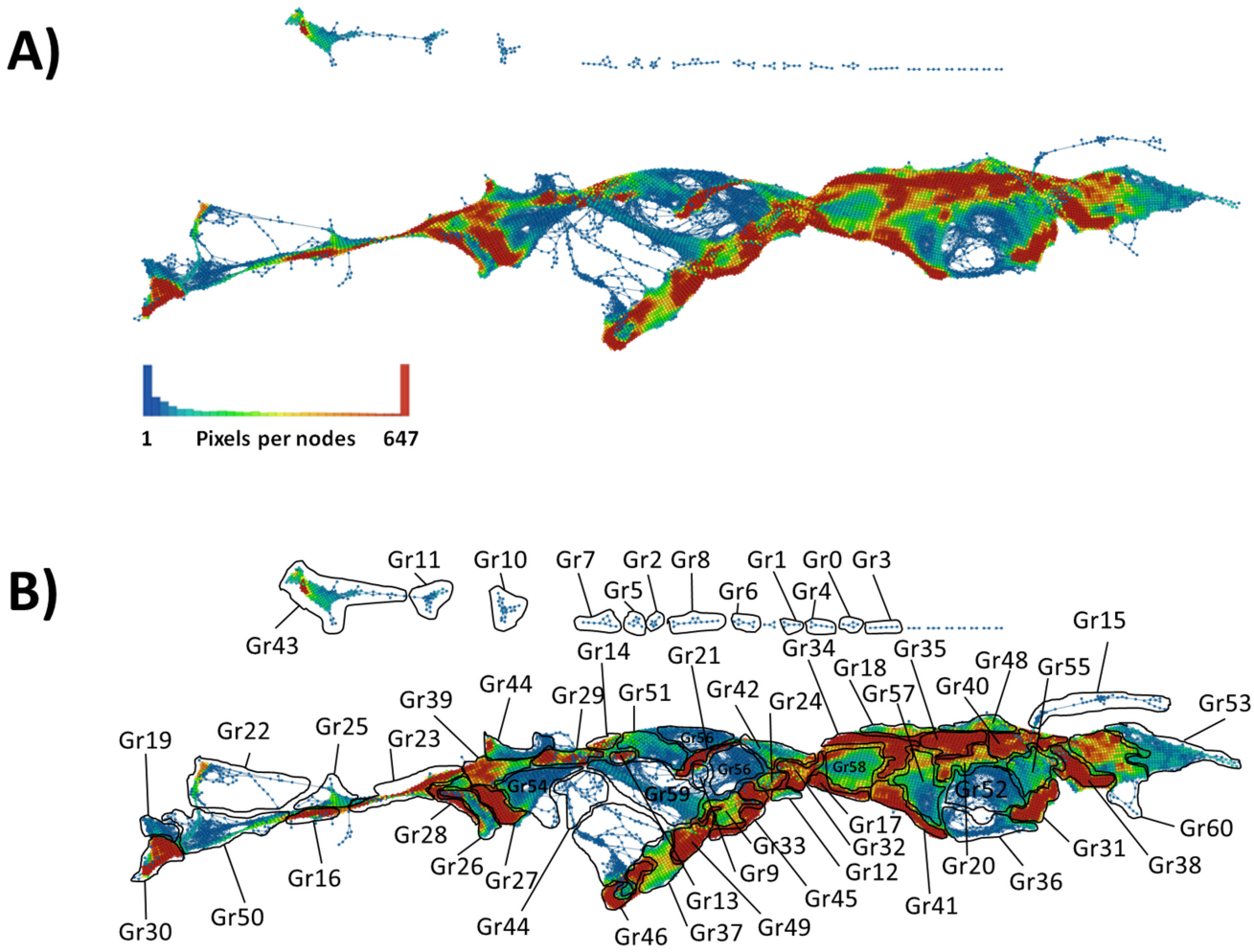

Figure 2. The topological network generated from the HYDICE data set (A) and its splitting in groups of pixels (B).

the data set (i.e. almost 400,000 pixels). Therefore, the first ten bigger groups contain around 5–6% of the total number of pixels each, while the smaller ones contain less than 0.05%. We can thus expect the possibility to observe minor and major contributions together in the final classification map. However, groups with less than 100 pixels have been considered as non-significant. It is now interesting to generate classification maps for the 60 selected groups of pixels. For obvious reasons, it is impossible to find an optimal look-up table in order to encode each group of pixels in one classification map with good contrast. We thus decided to generate a classification map with the first ten most important groups, a second one with the following ten and so on. It is also a good way to potentially observe features corresponding to very small groups with low numbers of pixels. Figures 3B–F show classification maps for all considered groups. Corresponding Matlab fig files can be downloaded from the Supplementary Material section of this paper's online abstract at https://doi.org/10.1255/jsi.2018.a1 for closer observation. Figure 3A presents an image of the considered Washington DC Mall area using the 60th variable (i.e. 758.89 nm) of the data cube. A grid with labelled columns and rows has also been added to this image in order to ease localisation of the particular zones discussed below. By looking at all these figures, it is not hard to see that we can observe more than the six to seven classes (i.e. water, buildings, grass, trees, roads, paths and shadow) usually observed in the literature for the exploration of the same data set.

In the next part, we propose to focus on each category, starting with the contribution of water, in this scene. In Figure 3B, group 30 (5.53% of the total pixels) seems to be linked with deep water such as, for example, in the Potomac River (A1:2, B1), in the Tidal Basin (E1, F1, G1:2, H1:2, I1) and in the Constitution Gardens Pond (E4:5, F4:5).

Group #	Nbr. of Nodes	Nbr. of pixels	% of pixels	
41	173	25724	6,55	**Groups observed in Figure 3B**
45	180	24601	6,26	
34	146	23524	5,99	
49	230	23280	5,92	
40	171	22934	5,84	
35	150	22900	5,83	
32	141	22122	5,63	
30	136	21678	5,52	
38	164	21581	5,49	
60	396	21324	5,43	
46	181	20754	5,28	**Groups observed in Figure 3C**
58	357	18508	4,71	
27	116	17959	4,57	
31	138	17865	4,55	
21	90	13805	3,51	
57	352	13600	3,46	
29	133	12325	3,14	
28	131	12111	3,08	
17	73	10903	2,77	
23	101	10494	2,67	
39	168	10262	2,61	**Groups observed in Figure 3D**
48	198	9432	2,40	
42	174	8452	2,15	
24	103	8219	2,09	
16	65	7947	2,02	
33	143	7827	1,99	
20	84	7327	1,86	
37	160	7181	1,83	
55	324	6949	1,77	
18	81	5575	1,42	
51	261	4643	1,18	**Groups observed in Figure 3E**
54	321	4383	1,12	
43	174	4315	1,10	
52	308	4129	1,05	
53	320	4058	1,03	
14	32	3797	0,97	
26	109	3150	0,80	
25	109	2813	0,72	
44	178	2773	0,71	
13	31	2678	0,68	
50	233	2547	0,65	**Groups observed in Figure 3F**
47	194	2489	0,63	
59	359	2396	0,61	
12	27	2394	0,61	
36	150	1872	0,48	
22	100	1618	0,41	
19	83	1178	0,30	
56	350	999	0,25	
9	18	701	0,18	
11	24	157	0,04	
15	43	91	0,02	**Groups considered as non significant**
10	21	73	0,02	
2	6	33	0,01	
3	6	31	0,01	
5	7	25	0,01	
6	7	22	0,01	
4	6	20	0,01	
7	10	17	0,00	
8	12	12	0,00	
1	5	8	0,00	
0	5	5	0,00	

Table 1. Description of the groups obtained from the splitting of the topological network.

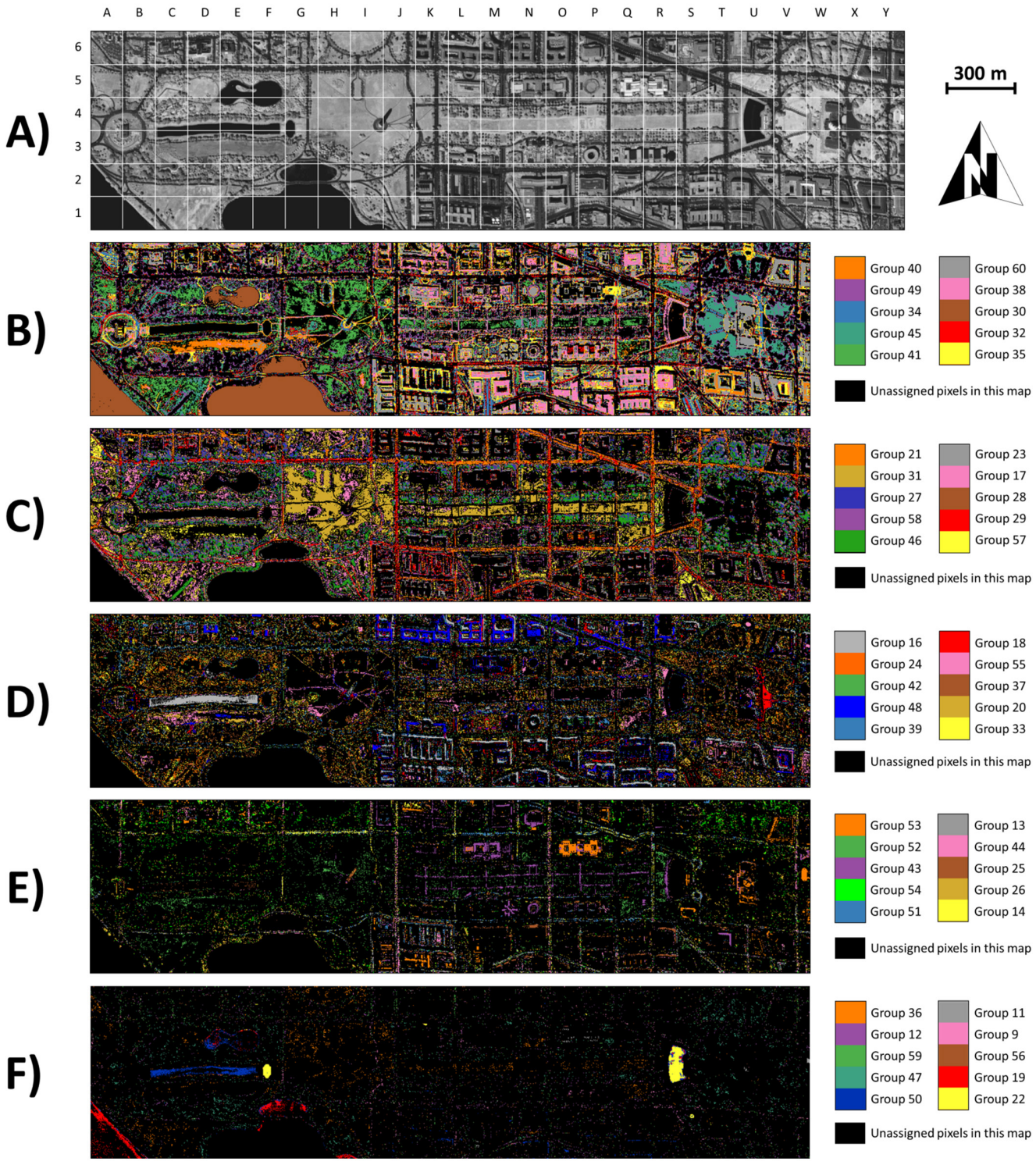

Figure 3. The scene observed at 758.89 nm (A) and corresponding classification maps (B–F).

In Figure 3F, group 50 (0.65% of the total pixels) shows areas with shallower waters, mainly from the Lincoln Memorial Reflecting Pool (lower part of C4, D4, E4, F4). This contribution is also observed to a lesser extent on the left border of the Constitution Gardens Pond (E4:5). It is also interesting to see that Group 22 (Figure 3F, 0.41% of the total pixels) allows us to discriminate even shallower waters in the small basin and fountains. The major contribution is observed for the fountain located on the right-hand side of the Lincoln Memorial Reflecting Pool

(G3:4) and the Capitol Memorial Reflecting Pool (T3:4). Other fountains are also detected, such as the one in front of the Smithsonian National Museum of American History (L5, a 20-pixel surface) or Bartholdi Park (U2, a 26-pixel surface). The author advises readers to download the Matlab fig files in order to observe such details in the images. We can go even further by detecting smaller fountains, such as the one in front of the National Gallery of Art (R6, 3 pixels) or two others in the courtyard of the US House of Representatives (V2, 2 pixels for the first one and 1 pixel for the second). In this first part of exploration, we observe that TDA is able to discriminate water areas with different depths, but also extract groups with very low numbers of pixels. In Figure 3F, pixels of group 19 seems to be linked with pixels of group 30 (Figure 3B) with the highest contributions observed on the bank of the Potomac River (A2, B1), the lower part of the Tidal Basin (G2:H2) and to a lesser extent on the border of the Constitution Gardens Pond (E4:5). It is difficult to say whether this new group of pixels highlights a new water depth or the detection of particular material such as alluvium. It should also be noted that the four groups discussed in this first part are all located at the left end of the topological network (Figure 2B).

Focusing now on buildings, previous papers have often considered them as a unique class. TDA seems to suggest a more detailed analysis considering different roofing materials. The majority of buildings are observed in Figure 3B. Group 60 (5.43% of the total pixels) corresponds to a first roofing material present on the United States Capitol (W3:4) and other buildings (X5:6, Y5:6, X1:2, Y1:2, Q2, K1, L1, 03, C6, L6, P5, R6). A second roofing material is linked with group 38 (5.49% of the total pixels, N2, 01:2, P1, Q1, R1, S1:2, T1:2 U2). Another very specific material is observed on the roof of the US House of Representatives (Group 40, 5.84% the total pixels, V2, W2). Surprisingly, we also retrieve this last contribution not on roofs but along the JFK Hockey Fields (C3, D3, E3, F3) or even on paths around the Washington Monument (H3:4, I3:5, J3:5). Group 35 (5.84% the total pixels) highlights another specific material on the roof of two buildings (L2, M2, T1, U1), in the courtyard (X1:2, Y1:2) or around other buildings (T2, U2, S5) and monuments (I4, A3:4, B3:4). Last, in Figure 3D, group 48 (2.40% of the total pixels) is also linked to another roof type (K6, L6, M6, N6, 06, P6). Once again, all these different groups (60, 38, 40, 35 and 48) are in the same area on the right-hand part of the topological network.

Detection of roads and paths is also very important in remote sensing. Main roads are observed in Figure 3C. Group 29 (3.14% of the total pixels) in red and group 21 (3.51% of the total pixels) in orange represent them. However, all these roads are asphalted. It, therefore, can be concluded that TDA detects two different materials. It is noticeable too that these two groups are quite far from each other in the topological network. At that level, however, it is too soon to draw any conclusions concerning the exact origin of this difference. It could be due to different formulations of bitumen or ageing levels. In Figure 3B, group 32 (5.63% of the total pixels) in red seems to correspond to another type of road. In reality, the observation of the Smithsonian National Air and Space Museum (Q3, R3) indicates quite the opposite. Indeed, highlighted parts of this building correspond to metallic structures. Then it is understandable that the group 32 allows us to detect this particular material in the scene, which is really uncommon. As a consequence, cars in circulation could be detected on main roads (e.g. N2, 02, P2). Moreover, heavy traffic zones or traffic jams are certainly observed (S3, S1, T1, U1, V1, P6 and Q6). Similarly, parked cars are detected along “L’enfant Plaza” (N1:2), at the side of the Capitol Reflecting Pool (T3, U3, T5, U5), around the Lincoln Memorial (A4, B4) or in car parks of some buildings (L3, M3, N5, 05). It is even possible to see smaller details such as streetlights along paths from the Peace monument or the Garfield Circle (V3:4) to the United States Capitol. Group 35 (5.83% of the total pixels) in Figure 3B corresponds to paved zones in courtyard of buildings (X1:2, Y1:2, 06), sometimes on their roof (T1, U1, R2, L2, M2) or their surroundings (T2, U2, R5, S5, D2). This same material can be retrieved on some paths (I3, J4) around the Constitution Gardens Pond (E4:E5, F4:5) and the Lincoln Memorial (A3:4, B3:4). A last contribution in specific zones is represented by group 34 in Figure 3B (5.99% of the total pixels, U1, V1, S1, T1, N1:2, A3, B3:4).

Soil covers and vegetation also remain a central focus in remote sensing. Four soil types are observed in Figures 3B and 3C. The first one (group 45, 6.26% of the total pixels) is particularly present around the United States Capitol (V3:4, W3:5, X3:4), but also to a lesser extent on small parcels of land (W1, X1, P4, H4, C1, D1, C5). The second soil type (group 41, 6.55% of the total pixels) is mainly located around the Washington Monument (H4:6, I3:6, J4:6), along the National Mall (L4, M4, N4,

04, Q4, R4, S4), near the Lincoln Memorial (A4:6) or between the Potomac River and the Tidal Basin (B1, C1, D1). In Figure 3C, two other types of soils (group 31, 4.55% of the total pixels and group 17, 2.77% of the total pixels) are also observed in the same zone. For their part, group 46 (5.28% of the total pixels) and 27 (4.57% of the total pixels) represent trees. When pixels of the two groups are present in the same area, it allows us to detect big trees and smaller ones when pixels of group 46 are only observed. Thus, group 27 should correspond to the shadow of trees. Group 49 (5.92% of the total pixels) in Figure 3B is also linked to the presence of trees. Big trees are mainly observed around the United States Capitol (V5, W5, X3:5, W3, V3) and on both sides of the Lincoln Memorial Reflecting Pool (C4, D4, E4, F4). Zones with small trees are observed, for example, near the Tidal Basin (C1, D1, E2, F2).

Of course, other groups have been generated with TDA, but the corresponding pixels were so spread in the scene that it was difficult to propose a clear identification of materials. This in no way detracts from the fact that TDA is able to extract many different kinds of materials revealing new promising areas in remote sensing.

Conclusion

The main objective of this paper was to introduce the concept of topological data analysis in the framework of remote sensing. Indeed, a large number of papers have already demonstrated that TDA has nice properties for the exploration of large data sets. These first results are encouraging, because it seems possible to extract different kinds of materials which are usually not considered. Moreover, we have to remember that this approach does not impose a data model. In this way, we can imagine that such approach could better manage non-linearities, which is quite unusual for many classical methods in remote sensing. We have seen also that TDA is able to generate classes even if they contain a small number of pixels. This characteristic is very important, because a real exploration of the scene is only obtained when major and minor contributions are detected simultaneously. We are conscious that this paper is a first introduction of TDA in remote sensing and that a deeper study is necessary, but these results seem to promise an excellent behaviour and very good properties.

References

1. J.R. Schott, *Remote Sensing: The Image Chain Approach*, 2nd Edn. Oxford University Press, New York (2007).
2. D. Landgrebe, "Hyperspectral image data analysis", *IEEE Signal Process. Mag.* **19(1),** 17 (2002). doi: https://doi.org/10.1109/79.974718
3. G. Carlsson, "Topology and data", *Bull. Am. Math. Soc.* **46(2),** 255 (2009). doi: https://doi.org/10.1090/S0273-0979-09-01249-X
4. G. Singh, F. Mémoli and G.E. Carlsson, "Topological methods for the analysis of high dimensional data sets and 3D object recognition.", in *SPBG*, p. 91 (2007).
5. P.G. Cámara, "Topological methods for genomics: present and future directions", *Curr. Opin. Syst. Biol.* **1,** 95 (2017). doi: https://doi.org/10.1016/j.coisb.2016.12.007
6. C.W. Bartlett, S.Y. Cheong, L. Hou, J. Paquette, P.Y. Lum, G. Jäger, F. Battke, C. Vehlow, J. Heinrich, K. Nieselt, R. Sakai, J. Aerts and W.C. Ray, "An eQTL biological data visualization challenge and approaches from the visualization community", *BMC Bioinformatics* **13 Suppl 8**, S8 (2012). doi: https://doi.org/10.1186/1471-2105-13-S8-S8
7. J.M. Gilmore, M.E. Sardiu, B.D. Groppe, J.L. Thornton, X. Liu, G. Dayebgadoh, C.A. Banks, B.D. Slaughter, J.R. Unruh, J.L. Workman, L. Florens and M.P. Washburn, "WDR76 co-localizes with heterochromatin related proteins and rapidly responds to DNA damage", *PloS One* **11(6),** e0155492 (2016). doi: https://doi.org/10.1371/journal.pone.0155492
8. J.M. Chan, G. Carlsson and R. Rabadan, "Topology of viral evolution", *Proc. Natl. Acad. Sci.* **110(46),** 18566 (2013). doi: https://doi.org/10.1073/pnas.1313480110
9. G. Sarikonda, J. Pettus, S. Phatak, S. Sachithanantham, J.F. Miller, J.D. Wesley, E. Cadag, J. Chae, L. Ganesan, R. Mallios, S. Edelman, B. Peters and M. von Herrath, "CD8 T-cell reactivity to islet antigens is unique to type 1 while CD4 T-cell reactivity exists in both type 1 and type 2 diabetes", *J. Autoimmun.* **50,** 77 (2014). doi: https://doi.org/10.1016/j.jaut.2013.12.003
10. D. Romano, M. Nicolau, E.-M. Quintin, P.K. Mazaika, A.A. Lightbody, H.C. Hazlett, J. Piven, G. Carlsson and A.L. Reiss, "Topological methods reveal high and

low functioning neuro-phenotypes within fragile X syndrome", *Hum. Brain Mapp.* **35(9),** 4904 (2014). doi: https://doi.org/10.1002/hbm.22521

11. G. Singh, F. Memoli, T. Ishkhanov, G. Sapiro, G. Carlsson and D.L. Ringach, "Topological analysis of population activity in visual cortex", *J. Vis.* **8(8),** 11 (2008). doi: https://doi.org/10.1167/8.8.11

12. J.L. Nielson, J. Paquette, A.W. Liu, C.F. Guandique, C.A. Tovar, T. Inoue, K.-A. Irvine, J.C. Gensel, J. Kloke, T.C. Petrossian, P.Y. Lum, G.E. Carlsson, G.T. Manley, W. Young, M.S. Beattie, J.C. Bresnahan and A.R. Ferguson, "Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury", *Nat. Commun.* **6,** 8581 (2015). doi: https://doi.org/10.1038/ncomms9581

13. J.L. Nielson, S.R. Cooper, J.K. Yue, M.D. Sorani, T. Inoue, E.L. Yuh, P. Mukherjee, T.C. Petrossian, J. Paquette, P.Y. Lum, G.E. Carlsson, M.J. Vassar, H.F. Lingsma, W.A. Gordon, A.B. Valadka, D.O. Okonkwo, G.T. Manley, A.R. Ferguson and TRACK-TBI Investigators, "Uncovering precision phenotype-biomarker associations in traumatic brain injury using topological data analysis", *PLOS ONE* **12(3),** e0169490 (2017). doi: https://doi.org/10.1371/journal.pone.0169490

14. A. Savic, G. Toth and L. Duponchel, "Topological data analysis (TDA) applied to reveal pedogenetic principles of European topsoil system", *Sci. Total Environ.* **586,** 1091 (2017). doi: https://doi.org/10.1016/j.scitotenv.2017.02.095

15. M. Offroy and L. Duponchel, "Topological data analysis: a promising big data exploration tool in biology, analytical chemistry and physical chemistry", *Anal. Chim. Acta* **910,** 1 (2016). doi: https://doi.org/10.1016/j.aca.2015.12.037

16. Y. Lee, S.D. Barthel, P. Dłotko, S.M. Moosavi, K. Hess and B. Smit, "Quantifying similarity of pore-geometry in nanoporous materials", *Nat. Commun.* **8,** 15396 (2017). doi: https://doi.org/10.1038/ncomms15396

17. P.Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson and G. Carlsson, "Extracting insights from the shape of complex data using topology", *Sci. Rep.* **3,** 1236 (2013). doi: https://doi.org/10.1038/srep01236

18. J.E. Ball and L.M. Bruce, "Level set segmentation of remotely sensed hyperspectral images", in *Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International*, **8,** p. 5638 (2005). doi: https://doi.org/10.1109/IGARSS.2005.1526055

19. J.E. Ball and L.M. Bruce, "Accuracy analysis of hyperspectral imagery classification using level sets", in *Proceedings of the 2006 ASPRS Annual Conference* (2006).

20. H.G. Akçay and S. Aksoy, "Automatic detection of geospatial objects using multiple hierarchical segmentations", *IEEE Trans. Geosci. Remote Sens.* **46(7),** 2097 (2008). doi: https://doi.org/10.1109/TGRS.2008.916644

21. J. Li and Y. Qian, "Clustering-based hyperspectral band selection using sparse nonnegative matrix factorization", *J. Zhejiang Univ.-Sci. C* **12(7),** 542 (2011).

22. J. Theiler, G. Cao, L.R. Bachega and C.A. Bouman, "Sparse matrix transform for hyperspectral image processing", *IEEE J. Sel. Top. Signal Process.* **5(3),** 424 (2011). doi: https://doi.org/10.1109/JSTSP.2010.2103924

23. G. Bilgin, S. Erturk and T. Yildirim, "Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines", *IEEE Trans. Geosci. Remote Sens.* **49(8),** 2936 (2011). doi: https://doi.org/10.1109/TGRS.2011.2113186

24. J.C. Tilton, Y. Tarabalka, P.M. Montesano and E. Gofman, "Best merge region-growing segmentation with integrated nonadjacent region object aggregation", *IEEE Trans. Geosci. Remote Sens.* **50(11),** 4454 (2012). doi: https://doi.org/10.1109/TGRS.2012.2190079

25. Q. Lu, X. Huang and L. Zhang, "A novel clustering-based feature representation for the classification of hyperspectral imagery", *Remote Sens.* **6(6),** 5732 (2014). doi: https://doi.org/10.3390/rs6065732

26. X. Huang, X. Liu and L. Zhang, "A multichannel gray level co-occurrence matrix for multi/hyperspectral image texture representation", *Remote Sens.* **6(9),** 8424 (2014). doi: https://doi.org/10.3390/rs6098424

27. J.M. Duarte-Carvajalino, G. Sapiro, M. Velez-Reyes and P.E. Castillo, "Multiscale representation and segmentation of hyperspectral imagery using geometric partial differential equations and algebraic multigrid methods", *IEEE Trans. Geosci. Remote Sens.* **46(8),** 2418 (2008). doi: https://doi.org/10.1109/TGRS.2008.916478

28. S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing", *IEEE Trans.*

Geosci. Remote Sens. **47(1),** 161 (2009). doi: https://doi.org/10.1109/TGRS.2008.2002882

29. Y.-Q. Zhao, L. Zhang and S.G. Kong, "Band-subset-based clustering and fusion for hyperspectral imagery classification", *IEEE Trans. Geosci. Remote Sens.* **49(2),** 747 (2011). doi: https://doi.org/10.1109/TGRS.2010.2059707

30. H. Su, Y. Sheng, P. Du and K. Liu, "Adaptive affinity propagation with spectral angle mapper for semi-supervised hyperspectral band selection", *Appl. Opt.* **51(14),** 2656 (2012). doi: https://doi.org/10.1364/A0.51.002656

31. N. Wang, B. Du and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6(2),** 554 (2013). doi: https://doi.org/10.1109/JSTARS.2013.2242255

32. S.D. Xenaki, K.D. Koutroumbas, A.A. Rontogiannis and O.A. Sykioti, "A layered sparse adaptive possibilistic approach for hyperspectral image clustering", in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, p. 2890 (2014). doi: https://doi.org/10.1109/IGARSS.2014.6947080

33. H. Su and P. Du, "Multiple classifier ensembles with band clustering for hyperspectral image classification", *Eur. J. Remote Sens.* **47(1),** 217 (2014). doi: https://doi.org/10.5721/EuJRS20144714

34. H. Su, Y. Sheng, P. Du, C. Chen and K. Liu, "Hyperspectral image classification based on volumetric texture and dimensionality reduction", *Front. Earth Sci.* **9(2),** 225 (2015). doi: https://doi.org/10.1007/s11707-014-0473-4

35. W. Wang, Y. Qian and Y.Y. Tang, "Hypergraph-regularized sparse NMF for hyperspectral unmixing", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **9(2),** 681 (2016). doi: https://doi.org/10.1109/JSTARS.2015.2508448

36. H. Zhai, H. Zhang, L. Zhang and P. Li, "Reweighted mass center based object-oriented sparse subspace clustering for hyperspectral images", *J. Appl. Remote Sens.* **10(4),** 046014 (2016). doi: https://doi.org/10.1117/1.JRS.10.046014

37. W. Wei, L. Zhang, C. Tian, A. Plaza and Y. Zhang, "Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering", *IEEE Trans. Geosci. Remote Sens.* 1 (2017). doi: https://doi.org/10.1109/TGRS.2017.2735488

38. W. Yang, K. Hou, B. Liu, F. Yu and L. Lin, "Two-stage clustering technique based on the neighboring union histogram for hyperspectral remote sensing images", *IEEE Access* **5,** 5640 (2017). doi: https://doi.org/10.1109/ACCESS.2017.2695616

39. H. Su, Y. Cai and Q. Du, "Firefly-algorithm-inspired framework with band selection and extreme learning machine for hyperspectral image classification", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10(1),** 309 (2017). doi: https://doi.org/10.1109/JSTARS.2016.2591004

40. L. Tong, J. Zhou, X. Li, Y. Qian and Y. Gao, "Region-based structure preserving nonnegative matrix factorization for hyperspectral unmixing", *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10(4),** 1575 (2017). doi: https://doi.org/10.1109/JSTARS.2016.2621003

41. F. Fan, Y. Ma, C. Li, X. Mei, J. Huang and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation", *Inf. Sci.* **397–398,** 48 (2017). doi: https://doi.org/10.1016/j.ins.2017.02.044

42. R. Sibson, "SLINK: an optimally efficient algorithm for the single-link cluster method", *Comput. J.* **16(1),** 30 (1973). doi: https://doi.org/10.1093/comjnl/16.1.30

43. V.D. Blondel, J.-L. Guillaume, R. Lambiotte and E. Lefebvre, "Fast unfolding of communities in large networks", *J. Stat. Mech. Theory Exp.* **2008(10),** P10008 (2008). doi: https://doi.org/10.1088/1742-5468/2008/10/P10008