****
Editor Comments

Associate Editor
Comments to the Author:
The manuscript has been reviewed by three experts, who all agree that the contribution is relevant and interesting. However, they identified several issues that should be addressed in a revision. Therefore, I recommend that the manuscript undergoes a major revision before being reviewed again.

I would encourage the authors to take the reviewer comments as an opportunity for improving their manuscript. In particular, all reviewers mention that the clarity of presentation should be improved. It might be advisable to address this by rewriting at least the critical sections, not just making minor changes. This would also make it clearer that the manuscript has a scientific contribution on its own, and is not just a extended version of the ECCV paper.

****

Reviewer Comments

Reviewer: 1

Recommendation: Author Should Prepare A Major Revision For A Second Review

Comments:
Summary: This paper suggest to take a transductive view on Zero-shot learning. It proposes to align attribute and word embeddings using CCA on the test data, and to use a graph based label propagation for final classification.

Paper strengths:
  + Applying transductive zero shot classification is interesting and novel
  + Extensive experiments on three datasets, showing promising results

Paper weaknesses
  - Unclear contributions wrt to preliminary version [15], and why a different method is used.
  - Experiments lack insight into the proposed paradigm (eg, influence of # projection dimensions, influence of # samples used for CCA).
  - Experiments compare different things, ie DAP and ALE do not assume the knowledge of the test items, and therefore unfair comparison.

First of all, congrats to the authors for their accepted ECCV work, well done. This paper is an

extension of the ECCV paper [15], and is relevant for the community.
I like the transductive view on zero-shot learning, but unfortunately after reading the paper I still don't grasp it fully, the paper lacks some important insights.
I think the paper has a lot of potential, if it addresses the comments below. Since a few comments ask for big changes, I rate a major revision.

The transductive view on zero-shot learning is interesting, although rather different from the scenario used in the traditional zero-shot methods such as DAP [28/29] and ALE[1]. Therefore, I'm unsure if a direct comparison is "fair", therefore claiming to have the "best results" so far should be taken with a grain of salt. The chosen setup is really different and that should be highlighted in the experimental section.

The paper is written reasonable ok, however it is still quite hard to follow due to the large amount of different symbols, with very similar meanings.
The section about the hypergraphs is difficult to follow in general. Neither is it clear why hypergraphs are used and not graphs (as far as I understood and looking at Fig 2, only edges between two nodes are used).
From the last paragraph in section 4.2 I understand that the prototypes are nodes in the graphs, but how are they defined (they don't have a visual feature vector).
Finally, I miss the description how the result of the random walk is used as classification. This section really lacks a intuitive description of the idea and the solution!

I think, the paper will improve if it explains more clearly the contributions over [15] in the intro and not hidden in section 2.
And, if it explains more clearly the differences with the label propagation method in [15], and perform a comparison.
Moreover, since the transductive view is rather new, a thorough comparison with [37] and other transductive methods is useful, eg why is CCA selected and not another method.
The same holds for transfer learning, the difference between domain shift (general transfer learning) and projection domain shift (as proposed in this paper) is unclear for me.
And a statement like: "renders any efforts to align the two domains directly unfruitful" (in the related work section) asks for an (experimental) validation.

In general, I miss experiments to increase the understanding of the proposed method(s), and I'm thinking about experiments and questions like:
- influence of projection dimensionality in CCA
- how many examples are required to construct CCA? Knowing all test items beforehand is a rather restrictive assumption, but knowing a few could make a lot of sense.
- can the CCA be performed using the validation images of the train images?
- which part (CCA or Label propagation) is responsible for the improvement? Ie what is the performance of DAP/kNN/SVMs on the CCA space? What is the performance of the label propagation methods on the original space(s)?

I think these additional experiments are crucial for understanding and to value the proposed methods.

Minor comments
- Figure 4, typo: Gij at the left and Gij at the right.
- Section 4.2: why is the notation p(k -> l) is used, instead of p(l|k)?
- Section 4.2: the symbol for complexity seems off.
- Section 4.2: how is the prior probability for G^i and G^{ij} defined (Eq 14/15), and that should be a single equation.
- Equation 10: typo, double + sign.
- Table 1: lots of white space, this could be presented more elegant.
- Section 6.2.2: description about Figure 3 (line 40-45), CUB -> USAA.
- Figure 3: it will be easier to compare when the y-scales are in similar ranges (esp fig 3.b), make a distinction between the transductive and non-transductive experiments.
- Section 6.2.2: confusing description, when using Decaf and Overfeat are there 6 or 9 available views (I think 9, but line 54 states 6).
- Figure 6: it would be good to include N=0 in the graphs and to compare to ALE[1].
- Table 2: T-5/B-5 is unexplained

Additional Questions:
1. Which category describes this manuscript?: Research/Technology

2. How relevant is this manuscript to the readers of this periodical? If you answer Not very relevant or Irrelevant please explain your rating under Public Comments below.: Very Relevant

1. Please evaluate the significance of the manuscript's research contribution.: Good

2. Please explain how this manuscript advances this field of research and/or contributes something new to the literature. : This paper proposes a transductive view on zero-shot classification, which is new and seems a promising direction.

3. Is the manuscript technically sound? In the Public Comments section, please provide detailed explanations to support your assessment: Partially

4. How thorough is the experimental validation (where appropriate)? Please discuss any shortcomings in the Public Comments section.: Insufficient; clearly inferior to state of the art, or necessary tests are absent

1. Are the title, abstract, and keywords appropriate? If not, please comment in the Public Comments section.: Yes

2. Does the manuscript contain sufficient and appropriate references? Please comment and include additional suggested references in the Public Comments section.: References are sufficient and appropriate

3. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on? If not, please explain your answer in the Public Comments section.: Yes

4. How would you rate the organization of the manuscript? Is it focused? Please elaborate with suggestions for reorganization in the Public Comments section.: Could be improved

5. Please rate the readability of the manuscript. Explain your rating under Public Comments below. : Difficult to read and understand

6. How is the length of the manuscript? If changes are suggested, please make explicit recommendations in the Public Comments section.: Too short – would benefit from additional material, details, or description.

7. Should the supplemental material be included? (Click on the Supplementary Files icon to view files): Does not apply, no supplementary files included

8. If yes to 7, should it be accepted:

Please rate the manuscript overall. Explain your choice.: Fair

Reviewer: 2

Recommendation: Author Should Prepare A Minor Revision

Comments:
The manuscript is well written, the theory is sound with extensive experiments and a thorough discussion of the results. I have some minor comments that may further increase the general value of the paper

- I would suggest to explicitly indicate the dimensionality of \Phi^i (i.e. \mathbb{R}^{n_T\times\m_i}) and W^i (in lines 10-12 of page 5). In particular a paragraph may be dedicated to explain why the embedding space dimensionality should be the sum of the separate original space dimensionalities. What about projecting all the views to a shared subspace and reducing the dimensionality instead of enlarging it? Subspace methods have proven to be effective for different domain adaptation problems (e.g. [12]) and can be good solutions also for heterogeneous zero-shot learning.

- Since \psi_c^\{\mathcal{X}} is synthetically defined, is it possible to understand a posteriori

how much
it is reliable as a prototype?

- I find a bit confusing the use of "k" in section 4.1. At the beginning it refers to every node in view i (lines 32-33 page 6), but then in the second column of page 6 it is used in the sentence "We select the first k highest values from ...". Please rephrase the sentences to explain better what "k" refers to.

- In line 35 of page 6 "k = 1, ...n_T+c_T": if I understand correctly both the unlabelled samples and the prototypes are considered as nodes. Please add an explicative sentence here.

some typos
- caption of figure 2: G^{ij} at the right --> G^{ji} at the right
- equation (10): "++" in the denominator
- caption of table 1: D is not indicated/explained
- line 27 page 11: "from from"

Additional Questions:
1. Which category describes this manuscript?: Research/Technology

2. How relevant is this manuscript to the readers of this periodical? If you answer Not very relevant or Irrelevant please explain your rating under Public Comments below.: Relevant

1. Please evaluate the significance of the manuscript's research contribution.: Good

2. Please explain how this manuscript advances this field of research and/or contributes something new to the literature. : This paper faces the problem of projection domain shift which arises when using an intermediate semantic representation for zero shot learning and presents interesting extensions with respect to the original conference publication in [15]. In particular, besides defining a projection for each view of the data to a multi-view embedding space, here the authors propose a heterogeneous hypergraph. A label propagation method on this hypergraph exploits the manifold structure of the data distributions over different views. The learned embedding space also enables new ways of sharing semantic and visual information.

3. Is the manuscript technically sound? In the Public Comments section, please provide detailed explanations to support your assessment: Appears to be - but didn't check completely

4. How thorough is the experimental validation (where appropriate)? Please discuss any shortcomings in the Public Comments section.: Compelling experiments; clearly state of the art

1. Are the title, abstract, and keywords appropriate? If not, please comment in the Public Comments section.: Yes

2. Does the manuscript contain sufficient and appropriate references? Please comment and include additional suggested references in the Public Comments section.: References are sufficient and appropriate

3. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on? If not, please explain your answer in the Public Comments section.: Yes

4. How would you rate the organization of the manuscript? Is it focused? Please elaborate with suggestions for reorganization in the Public Comments section.: Satisfactory

5. Please rate the readability of the manuscript. Explain your rating under Public Comments below. : Easy to read

6. How is the length of the manuscript? If changes are suggested, please make explicit recommendations in the Public Comments section.: About right

7. Should the supplemental material be included? (Click on the Supplementary Files icon to view files): Does not apply, no supplementary files included

8. If yes to 7, should it be accepted: As is

Please rate the manuscript overall. Explain your choice.: Good

Reviewer: 3

Recommendation: Author Should Prepare A Minor Revision

Comments:
* Keywords: Deduction & Theorem Proving & Knowledge Processing is not applicable

* Introduction: The introduction is mostly good, although the language does not make the

problem setting clear to readers new to the area of research. For example, it isn't clear without much more context what a "prototype" or a "semantic representation" is in this setting. It is also unclear what it means when it refers to "the transductive embedding space". Providing some grounding for a new reader would increase interest in the work.

It is nice work, though there are three main issues I would like to see addressed:

1) Readability. The grammar and English usage is good, but I found it took several reads, significant effort, and the reading of several reference papers to understand and gain an intuition for what is being done, and I am still uncertain about key aspects. One pattern in the manuscript is that the technical explanations start abstract and heavy with notation or equations, and the intuition or examples come much further in the manuscript, if at all. Perhaps it reads fine if you already understand the work, but as a first-time reader, I might have given up on it if I weren't a reviewer, and for maximum impact you want your paper to be accessible to TPAMI readers. This is a difficult thing to fix, I've made many suggestions below that I hope help.

2) The paper makes strong claims about being the only paper to identify and address domain shift, but the manuscript doesn't provide an intuition for how the way they applied CCA rectifies domain shift, and the experiments don't give experimental evidence that it does. The paper does not give an intuition for how CCA with the domain-biased word vectors and attributes with target (zero-shot) images results in embeddings that are not domain shifted, and for readers without experience in this area, the claim seems dubious. As for the experiments, the text for Fig 3 says the experiments show that domain shift is rectified, but those experiments do not isolate CCA as beneficial; for the experimental settings where CCA is used, they also use TMV-HLP (the hypergraph algorithm), which conflates the two parts of their approach. It also isn't clear how the control experiment (e.g. 'V' data point) was set up (e.g. is f^A(X_T) used?). Because the experiment always combines CCA and TMV-HLP, it doesn't clearly show that domain shift existed or is being addressed by CCA. Could you instead compare nearest neighbors with an without CCA? This would demonstrate that NN before and after transduction has addressed domain shift in the way the paper claims it should.

3) Related to (2) above I have major misgivings about how well the model could still classify the non-zero-shot (auxiliary classes). My best guess is that CCA skews the embedding space in favor of the zero-shot classes at the expense of the target classes. For example, what would Zebra look like in Fig 1 if you showed it after CCA? In most real world settings where one has an instance to categorize, one would not be told whether it is from a known or unknown class. If this is a drawback of the approach, it makes me very uncomfortable that the paper doesn't address this or even mention it. There is a sentence in the conclusion that appears to hint at this but doesn't directly address it ("The current framework needs to be extended to first distinguish these two types of data before performing zero-shot recognition"). If this is a drawback, be upfront and honest about it, similar to Socher et al. NIPS 2013 on Zero-shot.

=== Detailed comments ===

Most of my comment here will be about readability and ways to improve the text of the manuscript, though I also raise issues with some of the claims made in the paper.

Text locations are designated with page number, columns side, and line number, e.g. P1 R46 is page 1, right column, line 46.

From the start, make it clear that the transductive setting means that you are assuming that you have all your zero-shot instances at training time. There is a difference between this setting and another zero-shot setting which assumes that zero-shot instances are available one at a time and test time. You have "transductive" in the title, but this is a big enough difference that it is best to be clear.

P1 L31: typo, should be "or solely based [on] high-level description"

P1 L35: typo, should be "humans' ability to [recognize] without seeing examples"

P1 L45 - R56: In the problem statement, the abstract concepts of "semantic representation" and "prototype" are used. It will help readers if you ground these with an example upfront instead of following the paragraph with the example. With the example at the end, the reader may have to go back and re-read the paragraph to understand it. You could make this easier for the reader.

Fig 1: The figure provides a nice visualization, but the caption here is unnecessarily abstract, and the figure comes a page before the explanation, so it's helpful to provide a better explanation in the caption. Instead of saying prototypes and image feature projections, say they are attribute vectors and image features mapped using a regressor into attribute space.

P2 L29-L43: This text gives some description of Fig 1 except it leaves out key details as to how it was generated. After reading the rest of the paper, I understand how the prototypes and images were projected into the same space, but the text here doesn't give a good intuition, they're only described as "image feature projections". It also isn't clear that the domain shift shown isn't a failing of the particular regressor used, which doesn't learn the dimensions jointly---can you definitively claim based on the visualization that this is a common problem? You state here that you are the first to identify and address this problem, but that is an overreach if this figure is the evidence of the problem. Also, as I mentioned earlier, I also want to see the embedding of Zebra after CCA to see the effect on non-zero-shot classes.

P2 L49-52: This can be written to more clearly give a sense of why a prototype centered among class members still isn't good enough. In particular, the problem is that the clusters for the different classes overlap. This is what the text is trying to get at, but the central issue of overlap is buried down in an adjective at the end of the phrase ("struggle to assign the correct class labels to these highly overlapped data points").

P2 L54-56: It is incorrect that the issue of overlapping classes and limitation of a single prototype has never been identified. This was one of the focuses of Prototype Theory that was studied by psychophysicists before Comptuer Vision researchers, e.g. Eleanor Rosch in the 1970s. It will rub many readers the wrong way for you to claim you are the first to identify this problem.

P2 R8: This is the first time the term "views" has been introduced, and instead of being defined it is left up to the reader to understand what you mean from context. It makes the rest of the paragraph hard to read.

P2 R20: this line refers to "multiple semantic spaces" instead of "views". It is too much to expect the reader to tease apart the subtle difference between these two terms from context. Please provide a definition for this as well. (If they are referring to the same thing, then choose one and use it throughout.) Things get more confusing when the text refers to "multi-view semantic space alignment" at P2 R24.

P2 R48: The solution to the prototype sparsity problem is to "explore" the manifold structure. That's a vague term, it would be better to use a word that provides some better intuition as you introduce this part of your approach.

P2 R52-58: The description of heterogenous hypergraphs at this point in the text is heavy on terms but doesn't convey any meaning to a reader new to the concepts, so this section of the paragraph doesn't provide useful information to many of your readers. If you want to stay at a high level here, find a way to convey the idea of what's being done without relying on terms that will only become clear later.

P4 R3: Be more explicit that i is indexing into I and that m_i is the dimensionality of one semantic view (e.g. attribute or word vector). The way those are embedded in the sentence, I found myself debating between two possible interpretations (also because of an equation I read later, see P5 L19-20).

P4 R22-23: I could not find a description of how attribute words are turned into vectors. I only found here that "A is typically manually defined using a standard ontology". I was unclear through the paper whether each attribute is embedded individually or whether they were combined in some way and embedded as a single attribute vector per class. Please make that explicit at this point in the manuscript.

P4 R40-41: the SV regressors used to map instances into semantic space are introduced here, and they play a key role as the starting point, the basis for learning a multi-view space, and a means for experimental comparison, but how they were trained is not mentioned at all. They are also the regressors that are used to demonstrate the key idea in Fig. 1. The text doesn't even say what dimensionality the regressors project to.

P4 R51-56: This section describes the multi-view embedding, but it starts by defining variables and follows with an optimization equation. The section does not give an intuition for what CCA is doing in this setting, which is very useful for making sense of the math. Please address in an introductory explanation how it could be that associating the domain-shifted semantic views with the zero-shot instances improves the domain shift for zero shot classes. Is this at the expense of the auxiliary classes?

Eq (1): Indicate what variable is the optimization minimizing over.

P5 L11-12: Either somewhere around here or in the experimental section it would be good to say what the dimension of the embedding space is, e.g. the size of W_i that you use. I could not find this number anywhere in the paper.

P5 L19-20: I found it confusing that it says "The dimensionality of the embedding space is the sum of that of \Phi^i, i.e. \sum_{i=1}^n_V m_i". The dimensionality of the space is determined by the size of the Ws. Do you mean that the points lie in a subspace of that dimension? Or do you size W_i to map to that dimensionality?

P5 L54: "After alleviating the projection domain shift problem... ". At this point in the paper I am entirely unconvinced as I don't have an intuition for the algorithm and I have seen no experimental results. I suggest toning down the claim here.

P5 R17-21: At this point on my first read, I began to lose track of what vectors go through which functions and combined in which ways. You could address this by adding a box and arrow diagram to the paper illustrating how the data is being processed.

P5 R32-37: The motivation here for hypergraphs is going to be lost on many readers on their first read. Can you find a way to better illustrate the failing of a graph and the usefulness of a hypergraph? It would also help to define what a hypergraph is.

P5 footnote 4: "the prototypes are updated by one-step self-training as in [16]". This seems like potentially important information. As a reader, I need to read [16] to understand what self-training even is and why it's useful. Can you at least say something at a high level about what it is and why you do it? Is it a common step in these approaches?

P6 L17-18: having the same number of similar and dissimilar pairs is a strange assumption given that you have several classes.

P6 L16: "within the set" - I didn't know here what "set" you were referring to, probably because hyperedge was not defined - that would be an opportunity to introduce it as a set.

P6 L36-37: "includes the nodes in view j that are the nearest to node ..." - how is "nearest" defined? A distance cut-off? A number k of nearest nodes? Everything else appears to build on this, but it's left vague.

P6 R8-22, R47-50: These sections describe the normalizations and manipulations done to make the hypergraph work. While I'm glad the manuscript includes the details, it reads as though these are hacks that were thrown in to make it work. How much are these needed? Is there there a more prinicpled intuition behind them?

P6 R59-60: "With the pairwise similarity, one can now create the hypergraphs", which comes after a full page of describing hyperedge formation and hyperedge similarity, which sounded like it was hypergraph creation. If all that is left is a pruning, please describe it that way. Or if the process up to this point was described declaratively, then it may be easier for the reader to understand to also have an imperative description. Have you considered putting in some pseudocode?

P7 Fig 2: (1) The image here could be helpful in introducing all the math in the previous section instead of having it follow. (2) typo in the superscripts at "and G^[ji] at the right". (3) I expected here to see the query node clustered with the rest of the hyperedge, was that a misreading or an inaccuracy in the picture?

P7 L34-49: I read over this section and got nothing out of it. It's fine to have a section that's aimed at domain experts if that's the intention. Perhaps you preface with a word or two to indicate if that's the case.

P7 R24-26: "A typical solution to this is to fuse hypergraphs..." - this again seems like an important detail that's been left unexplained.

P7 R31: "Now we have two types... and 2-graphs G^P = {G^i}" - I don't recall seeing an explanation of what the 2-graph is.

P7 R33-35: "To propagate information from prototypes/labelled nodes ... a classic strategy is random walk". Please add something here that gives an intuition about how doing a random walk results in labels being assigned, before going into the math.

P8 R37-38: Perhaps because I never fully got how the attributes are represented, I was suprised here that you can reverse the mapping to give a set of attributes. Wouldn't that be a lossy or irreversible mapping? Or is this only giving a single attribute?

P9 L39-42: Did you try experiments on ImageNet using just word vectors (no attributes)? There are several other works that show results on the 800/200 split, e.g. PST, DeVise, and use features to DECAF. Is the issue that the running time for this approach is too high in the number of images?

P9 R8: Somewhere in here you need to give the parameters that you used. What dimensionality word vectors? How are the attributes turned into vectors? What dimensionality is the multi-view embedding space? Do these choices make a difference in the results? How did you choose them? Did you put together a validation set?

P9 R15: In the apples-to-apples comparison using the features that come with AwA, please give your number in the text. The only number written here is the other technique's result. As for the result here, which is 49% vs. 48.3%, did you check whether the 0.7% difference is significant?

P9 R30-31: "We will have more in-depth analysis on this result later" - I assume you mean later in the manuscript? It would be nice to give a Figure or a section number to guide the reader.

P10 Fig 3(a) and L45-49: (1) It isn't clear how the experiments with V and A were performed. (2) This result doesn't support your claim that CCA addressed domain shift because it could have been TMV-HLP that improved the performance, not CCA. To show the benefit of CCA, you need to use the embedding space without TMV-HLP.

P11 - In the secion on "Heretogeneous hypergraph vs. other graphs", I don't see mention of the ECCV results and the improvement from the graph label propagation algorithm. It at least seems worth a mention that they differ in algorithm there was an improvement in performance in this version.

P10 L58: Instead of "However, our TMV-HLP on all views ... improves" it is more fitting to say "Similarly, our ..."

P10 L60-R34: In bold you say ".. and the more views the better", though this is shown not to be true in Fig 3(b) on the last bar. Here you've included another 3 views, and the performance degraded. Clearly you can't make such a strong statement.

P10 R39: typo - V_D was repeated in list where you meant to have V_D, A_D.

P10 53: grammar - "the number of views to learn [an/the] embeddeding space doubles..."

P11 Fig 4: I'm not convinced the Overfeat t-SNE plot is a fair apples-to-apples comparison in 2 dimensions. How many dimensions in the Overfeat features? How many in the multi-view embedding space? They are more separable in 2 dimensions,

P11 L33: "There is no validation set" - there is always a validation set, you just remove some of your training data while you tune parameters.

P11 Fig 5: The legend says "Beofre T-embed", how is that done? My understanding was that the multi-view embedding must be done in order to run TMV-HLP.

P12 L16-17: "Our TMV-HLP algorithm is computationally efficient": It's a matter of application whether squared in the number of images is computationally efficient. If one is running on ImageNet, this will not be efficient enough. The scale differential increases from 30 minutes on AwA to about 18 days on ImageNet. You need to soften your claim here.

P12 L32-33: "contrast it to the conventional N-shot learning without the prototypes" - this is confusing, the prototypes can be word vectors, e.g., and are how labels are assigned to the images. How do you do zero-shot without any prototypes? I found this section to be quite confusing.

P12 L38: "we modify the SVM- used in [29] into SVM+" - please say a little more about how.

P12 Table 2: make it easier on the reader by putting an explanation of T-5 and B-5 in the caption. It also helps if you order the list of animals for the abbreviations in the caption in the same order they are shown in the table.

P13 L22: "Zero-attribute learning": this isn't a good name for what is being described here, which is assigning a class from a list of attributes. The same indicates that attributes are not available.

P14 Table 3: (a) I found this table and the text describing it very confusing. The query is a list of attributes, but the table shows a class as the query. This is somewhat explained, but the indirection should be removed from the table or describe the columns differently (e.g. ground truth instead of query). (b) How many attributes were added in the 3rd row? Can they be listed in the caption?

Additional Questions:
1. Which category describes this manuscript?: Research/Technology

2. How relevant is this manuscript to the readers of this periodical? If you answer Not very relevant or Irrelevant please explain your rating under Public Comments below.: Very Relevant

1. Please evaluate the significance of the manuscript's research contribution.: Good

2. Please explain how this manuscript advances this field of research and/or contributes something new to the literature. : The manuscript extends work published in ECCV that uses a combination of techniques to perform transductive zero-shot image classification in domains where semantic information is available for classes (class attributes, word vectors). They show impressive performance on the task relative to other methods using transduction, e.g. on the

AwA (animals with attributes) data set. They introduce TMV-HLP, an algorithm for generating a hypergraph and propagating label information across it.

3. Is the manuscript technically sound? In the Public Comments section, please provide detailed explanations to support your assessment: Appears to be - but didn't check completely

4. How thorough is the experimental validation (where appropriate)? Please discuss any shortcomings in the Public Comments section.: Lacking in some respects; some cases of interest not tested

1. Are the title, abstract, and keywords appropriate? If not, please comment in the Public Comments section.: No

2. Does the manuscript contain sufficient and appropriate references? Please comment and include additional suggested references in the Public Comments section.: References are sufficient and appropriate

3. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on? If not, please explain your answer in the Public Comments section.: Could be improved

4. How would you rate the organization of the manuscript? Is it focused? Please elaborate with suggestions for reorganization in the Public Comments section.: Satisfactory

5. Please rate the readability of the manuscript. Explain your rating under Public Comments below. : Difficult to read and understand

6. How is the length of the manuscript? If changes are suggested, please make explicit recommendations in the Public Comments section.: About right

7. Should the supplemental material be included? (Click on the Supplementary Files icon to view files): Does not apply, no supplementary files included

8. If yes to 7, should it be accepted: After revisions. Please include explanation under Public Comments below.

Please rate the manuscript overall. Explain your choice.: Good