Dear Editor,
Please find attached the author response to the reviewers' comments of the manuscript "Hydrological characterization of cave drip waters in a porous limestone: Golgotha Cave, Western Australia" by Mahmud et al. for your consideration. We believe we have addressed all questions raised by the two reviewers of the manuscript.

As requested, we have included a detailed response to reviewer's questions (below). A revised version will be provided once a decision has been made to accept the manuscript.

If you have any additional questions, we would be more than happy to address them.

With kind regards,

For the authors,
Kashif Mahmud

---

Comments to the Author
Interactive comment on "Hydrological characterization of cave drip waters in a porous limestone: Golgotha Cave, Western Australia" by Mahmud et al.

Reviewer(s) Comments:
Anonymous Referee #1
In the submitted manuscript Mahmud et al. provide a detailed study on the dynamics of cave drips at a large karst cave in South West Australia. The authors use a large set of automatically recorded drip rate records to classify and cluster the different drips by their statistical properties and knowledge from previous research. They show that established classification schemes do not apply to their data set but their new clustering method provided a clear distinction of 4 clusters of drip types within the cave. The most prominent one, cluster 1, consisted of drips that were mostly controlled by matrix flow, which is in accordance with previous classification using LiDAR imaging. The other clusters were expressed by a stronger hydrological variability in terms of mean discharge and flow variability.

We appreciate the reviewer's comments.

The manuscript is generally well-written, the methods are clearly explained and the conclusions are well supported by the results and discussion. Some improvement is necessary in terms of structure and detail: Some parts of the methods appear in the results section and should be moved to the methods description. Also the elaborations about the drip characteristics are a bit too detailed and may be shortened to improve the readability of the manuscript.

We will reorganize the method/result sections and remove the unnecessary portion that describes the previous paper.

Finally, another subsection addressing the impact of the results of this study would be very valuable for the karst and cave hydrology communities. I am confident that this can be done in the frame of minor revisions.

The impacts of the findings of this study have been briefly identified in the last paragraph of the conclusion, however we will elaborate this discussion further, including presenting a summary of the implications of the current study, as well as bringing together an overall summary tying in the relevant findings from the previous two papers by Mahmud et al. We will also recommend potential future research directions. This will appear as the following subsection:

**Implications of the findings and future research:**
We will emphasize the overall impacts of this study from different perspectives, starting with temporal analysis, multisite analysis, karst modelling, and then paleoclimate records.

Starting with the time-series analysis, this research presents a methodology that can be applied globally for drip logger data. The results show that some data-integration is necessary to avoid artefacts from slow drip sites. For sites where there is significant matrix flow, our study has demonstrated that the Smart and Friederich (S&F) classification is not appropriate. Therefore, this study proposes an alternative hydrological classification scheme that incorporates cave sites influenced by matrix flow. The times series approach adopted in this study also opens the way for improved analysis and classification of hydrology time series in general e.g. tests for non-linearity, autocorrelation, cluster analysis, etc. and all of these will certainly benefit our understanding of the hydrology of karst systems.

In this study, we also extend the analysis of drip time series to multiple sites, whereby we take advantage of the ensemble of loggers to extract common properties by clustering, which would not be possible with single site analysis. The results show that by considering multiple simultaneous time series, one can make better inferences about water flow and unsaturated zone properties. The main impact is to recommend the use of spatial networks of loggers over individual loggers – it should be noted that currently, most researchers deploy only a few loggers to understand the flow to individual sites. This study also proposes a possible methodology for the analysis of such datasets.

Regarding application of our findings, we believe that our methodology based on drip logger datasets can provide direct evidence of deep drainage, and therefore the timing of diffuse recharge, which could be used for basic model calibration. Spatial drip data (possibly combined with Lidar) is beneficial to infer flow types (e.g. the proportion of fracture vs matrix, etc.) which could be used for model configuration to produce realistic karst recharge (Hartmann et al. 2012), and hence large-scale groundwater estimation (Hartmann et al. 2015). Another potential application is the integration of flow types in groundwater models through inverse modelling. Such data could also be used to constrain water isotope model configurations used for forward modelling speleothem $\delta^{18}O$ (Bradley et al. 2010, Treble et al. 2013). Overall, the findings of this work will definitely provide a better understanding of processes that control vadose zone flow and transport processes, which would ultimately help develop approaches to incorporate these processes into simulation models (Hartmann and Baker 2017).

The analysis, presented here and combined with the findings of our previous two papers (Mahmud et al. 2016, Mahmud et al. 2015), provides valuable information for paleoclimatologists and geochemists wishing to sample stalagmites. While these studies have characterised Golgotha Cave, they could be applied to any other cave system globally. We can summarize our previous work as follows: 1) we have devised a classification for flow-type based on stalactite morphology (Mahmud et al. 2015); 2) quantified the recharge response of each flow type to infiltration (Mahmud et al. 2016); 3) combined findings in points 1-2 to estimate the total volume of cave discharge; 4) compared cave discharge with infiltration to estimate the total recharge volume and identify highly focused areas of recharge (Mahmud et al. 2016). The current study has further developed the spatial and temporal statistical relationships between the flow sites, permitting both quantification and visualisation of the hydrology between the ground surface and the cave ceiling. More generally, these studies illustrate the heterogeneity between flow sites and what causes this, as well as putting forth methods that can be applied to any cave system to better understand diffuse recharge and paleoclimate records from speleothems.

We further propose some ideas for future research that have evolved from this study. For example:

a) Combining drip logger network with a surface weather station and soil moisture network to constrain the water balance with site specific measurements using modelled input time series derived from nearby meteorological stations. Additionally, employing sap flow meters to constrain tree water use.
b) Combining the logger network, which constrains diffuse recharge, to a bore network that measures groundwater level, to understand the relative importance of diffuse and river recharge.
c) Combining cave drip logger data with surface geophysics data to track water movement.

Please see the attached and commented pdf for more detailed specific and technical comments.

1) Line 27: The typo

Will be corrected.

2) Line 50: The typo

A comma will be added.

3) Line 54: This may depend on the more or less developed connection between the surface and the cave, doesn't it?

Yes indeed, will be added.

4) Lines 57-58: remove unnecessary brackets

Will be removed.

5) Line 131: Provide short information about method.

Additional explanation will be provided in terms of modelled evapotranspiration (ET) data which is core in determining the water budget and hence the total amount of infiltration. ET data is collected from the Australian Water Availability Project (AWAP) and referenced in the text (Raupach et al., 2009). According to Raupach et al. (2009), the ET is modelled using the mathematical equations of Priestley-Taylor.

6) Line 139: The second part of this section is a bit too detailed, please shorten.

We will focus on these to minimize the description of previous works.

7) Line 160: better use points instead of lines to avoid misinterpretation

We think point plots would make it difficult to grasp the trend of these dense time series, and hence we prefer to stick with the line plots.

8) Line 216: which processes does the offset account for? different flow lengths? Please provide short explanation.

The offset *O* (in hours) is needed to align two time-series such that they present maximum correlation. This offset accounts for the lag time based on the maximum correlation between two-time series in order to match those time series. This will be explained.

9) Line 223: Please provide some short description of the MDS method for readers that are not familiar with the method.

Multi-dimensional scaling (MDS) starts by defining a distance between a set of objects. In our case, each drip logger is an object and a specific distance between drip loggers is considered to characterize the similarity between any two loggers. This definition will be added.

10) Line 248: please move apart the 10^4 and the figure title, please provide x- and y-axis labels

The figure Y-axis ticks will be modified, however we feel adding the axis levels to each of the subplots will make the diagram congested. So, we described the axis levels in the figure caption.

11) Line 262: Same as above: figure needs axis labels

We feel adding the axis levels to each of the subplots will make the diagram congested. So, we described the axis levels in the figure caption.

12) Lines 268-269: Please mention in methods that this is part of the analysis.

Will be mentioned in method section.

13) Line 283: please mention in caption that ln is the log10 (isn't it?)

Yes it is, will be cited.

14) Line 292, 295: here, and some lines before some accidental breaks should be removed

Will be removed.

15) Line 306: K-means requires a pre-definition of the number of clusters. How was this done here?

We have defined the cluster number in section 4 (Line 227) based on the number of flow categories identified by Mahmud et al. (2016).

16) Line 316: or their flow paths pass the same hydrological domain, the karst matrix.

That could be another possibility and therefore will be added.

17) Line 332: Another subsection discussing the implications of this research concerning the understanding of cave hydrology, karst recharge, and paleoclimate reconstruction is missing here. Also, how the newly gained knowledge could be incorporated into hydrological models to simulate the unsaturated zone (see refs below) would be very valuable.

Bradley, C., Baker, A., Jex, C.N. & Leng, M.J. 2010. Hydrological uncertainties in the modelling of cave drip-water δ18O and the implications for stalagmite palaeoclimate reconstructions. Quaternary Science Reviews, 29, 2201–2214, doi: 10.1016/j.quascirev.2010.05.017.

Treble, P.C., Bradley, C., et al. 2013. An isotopic and modelling study of flow paths and storage in Quaternary calcarenite, SW Australia: Implications for speleothem paleoclimate records. Quaternary Science Reviews, 64, 90–103, doi: 10.1016/j.quascirev.2012.12.015.

Hartmann, A., Lange, J., Weiler, M., Arbel, Y. & Greenbaum, N. 2012. A new approach to model the spatial and temporal variability of recharge to karst aquifers. Hydrology and Earth System Sciences, 16, 2219–2231, doi: 10.5194/hess-16-2219-2012.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y. & Wagener, T. 2015. A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. Geoscientific Model Development, 8, 1729–1746, doi: 10.5194/gmd-8-1729-2015.

> Earlier in this response letter we have mentioned adding a subsection with all these suggested references. This new section discusses the implications of this research concerning the understanding of cave hydrology, karst recharge, modelling, and paleoclimate reconstruction.

Reviewer(s) Comments:

Anonymous Referee #2

This manuscript is a follow up on drip monitoring data that were published in previous works including a 2016 paper in HESS. Whereas classification of flow regimes in the previous paper(s) was based also on morphological characteristics of the stalactites, here a similar clustering is based solely on a cluster analysis of the drip data. Beside the cluster analysis, there are new histograms and analysis of autocorrelation, which may add some qualitative understandings of the karst flow regime in these stalactites (seasonality, annual precipitation variability). I am not sure the "delta" from the previous works on this data that is presented in the current manuscript is worth a new HESS paper. I am sure that in the present way it is written it is not. Therefore, I recommend on a major revision in which: 1) the description of previous methods and results will be decreased to minimum; 2) Elaboration on the new statistical methods and results that are used here, 3) the presentation and discussion concerning the histograms and autocorrelation analysis will be upgraded significantly; and 4) the "delta" from our understanding of the system we had before this analysis will be declared more explicitly.

> We will remove the unnecessary portions that describe the previous paper and elaborate the result section significantly with the description of new statistical analysis. We will also add a section about the wider impacts of the research in the field of karst and cave hydrology that distinguishes the added findings compared to our previous works in this domain (discussed above). Some parts related to the site description and methods have been kept identical in the first version, but these will be stripped to the minimum during the revision and substantially revised, and this will make space for increased discussions and presentation of the results.

Major Comments

1) Even though it is declared in the manuscript and figure captions, it's inappropriate that more than half of the paper including 3 figures and 2 tables are repetition of methods and results of a previous work. It doesn't look good, especially the almost identical figures, avoid.

We will remove Figure 2 completely and cite our previous work (Mahmud et al. 2016) for the climate data and drip time series. Figure 3 will be redesigned by focusing only the cave floor images, complementing with the ceiling image and photographs of underlying stalagmites shown in Fig. 3 of previous paper (Mahmud et al. 2016). We will significantly modify the tables to represent the new findings with minimal overlap of previous outcomes.

2) Lines 83-212, old stories, to be reduced to a 1/3 of current.

We will condense the content more and emphasize on our previous works (Mahmud et al. 2016, Mahmud et al. 2015) to follow for more detail.

3) Lines 213-228 these are the new methods: Elaborate in explaining them, equations, figures that illustrate, etc,. MDS, K-means, these are not general statistics (this section (4) has to be as long as sections 2+3 at least).

We employed a data analysis technique called Multi-dimensional scaling (MDS), that allows a data dimensionality reduction i.e., mapping complex multidimensional data on a low-dimensional manifold. MDS is a technique concerned with embedding a set of points in a low-dimensional space so that the distances between the points resemble as closely as possible a given set of dissimilarities between objects that they represent (Birchfield and Subramanya 2005). MDS requires a distance matrix to be computed, in which a single scalar number characterizes the similarity between any two time-series. It takes an input matrix giving dissimilarities between pairs of items and outputs a coordinate matrix whose configuration minimizes a loss function. MDS is also known as Principal Coordinates Analysis (PCoA) and works differently from principle-component analysis (PCA), which operates on a covariance matrix, MDS operates on a distance or dissimilarity matrix (Pisani et al. 2016). Even if PCA and MDS methods can return the same results in specific contexts, MDS can be considered as a more general method that maintains its validity in a rigorous sense also for non-euclidean distances, i.e., the distance matrix (*d*) chosen in this study. MDS is used to translate these distances into a configuration of points defined in an n-dimensional Euclidean space (Cox and Cox 1994). A MDS results in a set of points arranged so that their corresponding Euclidean distances indicate the dissimilarities of the time series.

The basic steps of performing the MDS algorithm are:

1. Construct the distance matrix d: One key component in clustering is the function used to measure the temporal similarity (or distance) between any two time-series being compared. To define an appropriate measure of similarity between time series, we determine two factors: firstly, the offset (*O*) to match two time-series based on their maximum correlation, and secondly the complement of the correlation coefficient (1-*R*) between the time series (Jex et al. 2012). Initially, we compute the cross-correlation function and *O* is defined as the lag time based on the maximum correlation between two time-series. Next, we define R as the correlation coefficient with the time series being moved by the offset amount *O* to have maximum correlation coefficient. Finally, the distance matrix (*d*) is computed for each pair of loggers using the following equation (Jex et al. 2012):

i. $d = O\,(1 - R)$

2. Compute the inner product matrix $\mathrm{B} = -\frac{1}{2}JDJ$, where $\mathrm{J} = I - \frac{1}{n}\mathbf{1}\mathbf{1}^{T}$ is the double-centering matrix and **1** is a vector of all ones.
3. Decompose B as $\mathrm{B} = V\Lambda V^{T}$, where $\Lambda = \mathrm{diag}\,(\lambda_1, \dots, \lambda_n)$, the diagonal matrix of eigenvalues of B, and $\mathrm{V} = [\mathrm{v}_1, \dots, v_n]$, the matrix of corresponding unit eigenvectors. Sort the eigenvalues in non-increasing order: $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$.
4. Extract the first p eigenvalues $\Lambda_p = \mathrm{diag}\,(\lambda_1, \dots, \lambda_p)$ and corresponding eigenvectors $\mathrm{V}_p = [v_1, \dots, \mathrm{v}_p]$.
5. The corresponding Euclidean distances of the set of points, indicating the dissimilarities of the time series are now located in the $n\times p$ matrix $\mathrm{X} = \left[x_1, \dots, \mathrm{x}_p\right]^{T} = \mathrm{V}_p {\Lambda_p}^{1/2}$.

The k-Means clustering algorithm is then used to divide these points into k clusters, which corresponds to a categorization of the drip data time series. k-means clustering, or Lloyd's algorithm (Lloyd 1982), is a method of vector quantization that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

The algorithm proceeds as follows:
1. Choose *k* initial cluster centers (*centroid*): Here, we use *k*=4 clusters as this was the number of flow categories identified by Mahmud et al. (2016).
2. Compute point-to-cluster-centroid distances of all observations to each centroid. There are two steps to follow: first assign each observation to the cluster with the closest centroid. Then individually assign observations to a different centroid if the reassignment decreases the sum of the within-cluster, sum-of-squares point-to-cluster-centroid distances.
3. Compute the average of the observations in each cluster to obtain *k* new centroid locations.
4. Repeat steps 2 and 3 until cluster assignments do not change, or the maximum number of iterations is reached.

All these will be added in the manuscript.

4) What does the MDS analysis add to the correlation matrices analyzed in Mahmud et al. 2016. ?

This paper goes beyond the analysis done by Mahmud et al. (2016), where the correlation was considered without offset, and the quantitative drip data were not used for clustering, but only for validation. We could summarize both the papers like below in terms of investigating the dependence between multiple drip time series:

Mahmud et al. (2016): correlation between time series
Mahmud et al. (2017): computation of a distance between time series + offset correction + MDS-based clustering

Overall this study establishes a novel way to find consistent characterization of cave hydrology, which can be obtained by performing together both methodologies of Mahmud et al. (2015) and Jex et al. (2012). It relies on a metric that defines drip logger time series as similar if they are well correlated and only have a small offset with one another, and therefore these time series should cluster together. The MDS analysis supports this hypothesis and moreover, displays the spatial patterns of the flow paths between the surface and the cave chambers. This technique shows potential to classify, quantify and visualise the observed relationships between infiltration through the fractured limestone rocks and surface climate inputs.

Specific Comments

5) L 22 –Abstract should not include references.

Will be removed.

6) L 27 Capital C

Will be corrected.

7) L 37 for development of karst in relation to fractures, beds, bedding plains see also Kurtzman et al., 2009, Geosphere, v. 5; no. 2; p. 126–139;

The reference will be added.

8) Line 91 "high matrix porosity" – give numbers 0.1? 0.2? 0.3?

The porosity is reported in Smith et al. (2012) as 0.3 – 0.5. Will be added.

9) Lines 215-221: define cross-correlation function; is O and R calculated to all n(n-1)/2 pairs of drip data, elaborate, explain the method. End of lines 216 and 218 – redundancy. What are the dimensions of the distance matrix – d?

Cross-correlation is a measure of similarity of two time-series as a function of the displacement of one relative to the other. The function is an estimate of the covariance between two time-series, $y_{1t}$ and $y_{2t}$, at lags $k = 0, \pm1, \pm2, ....$

Yes, both *O* and *R* are calculated to all n(n-1)/2 pairs of drip data. We will reorganize this section with detailed explanation on both MDS and K-means algorithms, as we stated before in comment 3.

The distance matrix d is square, symmetric, and has dimension equal to the number of drip loggers.

10) Lines 223 – 228, what is MDS? Is it of the family of the classic principle-component analysis (PCA)? In the current application how many dimensions? Equations, figures, tell us the method? What is the K-means algorithm – elaborate on top of citing?

We have discussed above the MDS and k-means algorithm earlier in this response letter. It will be explained in details in the revised manuscript.

11) Figure 4. Put all vertical axis the same - 10000 or 1X10-4, do not mix; Color choice not good – try blue for wet seasons ad red for dry, or rainy 2013 in contrast with other play with it and choose the better, print and see if it is good on printing as well.

Y-axis ticks will be modified as well as different colour choices will be explored to find the best one.

12) L 245. These histograms are not skewed. Maye be add a sub figure to figure 4 with the most characteristic normal, skewed, be-modal histograms, including a continuous line of the pdf to illustrate.

The histogram plots were not clearly explained and will be improved in the revised manuscript. We agree with the reviewer’s suggestion to highlight the extreme skewed or bimodal histograms, and will add those as sub-plots.

13) Figure 5 and relevant text. – The most contrasting ACF are usually at time lag of 150-200 (1 season in the 2-seasons Mediterranean climate), I would plot these ACF until lags of 365 days to enhance the seasonal understanding, that may be gained.

We tested the ACFs until lags of 365 days, however in general the yearly signal is quite weak. In some drips, we got some negative correlation, but it is very insignificant and no process can explain that negative yearly correlation.

14) Figure 6 and 7, choose more contrasting colors and increase symbol size.

Contrasting colors will be used with larger symbol size.

15) L 292 and 296 typo mistake of new line.

Will be corrected.

16) L313-314 “inconsistency” - unclear

We observe that the clustering generally agrees with the morphology-based flow classification of our previous work (Mahmud et al. 2016) with few exceptions. For example, site 2vi has really high discharge with high variability, showing inconsistent drip rate.

17) L339 the beginning of line is unclear

The histogram distributions of various drip data time series can illustrate the differences between the flow classifications. Will be clarified.

Additional references:

Birchfield, S.T. and Subramanya, A. (2005) Microphone Array Position Calibration by Basis-Point Classical Multidimensional Scaling. IEEE Transactions on Speech and Audio Processing 13(5), 1025-1034.

Bradley, C., Baker, A., Jex, C.N. and Leng, M.J. (2010) Hydrological uncertainties in the modelling of cave drip-water δ18O and the implications for stalagmite palaeoclimate reconstructions. Quaternary Science Reviews 29(17-18), 2201-2214.

Cox, T. and Cox, M. (1994) Multidimensional scaling, Chapman and Hall, London.

Hartmann, A. and Baker, A. (2017) Modelling karst vadose zone hydrology and its relevance for paleoclimate reconstruction. Earth-Science Reviews 172, 178-192.

Hartmann, A., Gleeson, T., Rosolem, R., Pianosi, F., Wada, Y. and Wagener, T. (2015) A large-scale simulation model to assess karstic groundwater recharge over Europe and the Mediterranean. Geosci. Model Dev. 8(6), 1729-1746.

Hartmann, A., Lange, J., Weiler, M., Arbel, Y. and Greenbaum, N. (2012) A new approach to model the spatial and temporal variability of recharge to karst aquifers. Hydrol. Earth Syst. Sci. 16(7), 2219-2231.

Jex, C.N., Mariethoz, G., Baker, A., Graham, P., Andersen, M., Acworth, I., Edwards, N. and Azcurra, C. (2012) Spatially dense drip hydrological monitoring and infiltration behaviour at the Wellington Caves, South East Australia. International Journal of Speleology 41(2), 283–296.

Lloyd, S. (1982) Least squares quantization in PCM. IEEE Transactions on Information Theory IT-28(2 pt 1), 129-137.
Mahmud, K., Mariethoz, G., Baker, A. and Treble, P.C. (2017) Hydrological characterization of cave drip waters in a porous limestone: Golgotha Cave, Western Australia. Hydrol. Earth Syst. Sci. Discuss. 2017, 1-19.
Mahmud, K., Mariethoz, G., Baker, A., Treble, P.C., Markowska, M. and McGuire, L. (2016) Estimation of deep infiltration in unsaturated limestone environments using cave LiDAR and drip count data. Hydrol. Earth Syst. Sci. 20, 359-373.
Mahmud, K., Mariethoz, G., Pauline, C.T. and Baker, A. (2015) Terrestrial Lidar Survey and Morphological Analysis to Identify Infiltration Properties in the Tamala Limestone, Western Australia. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 8(10), 4871 - 4881.
Pisani, P., Caporuscio, F., Carlino, L. and Rastelli, G. (2016) Molecular Dynamics Simulations and Classical Multidimensional Scaling Unveil New Metastable States in the Conformational Landscape of CDK2. PLoS ONE 11(4), 1-22.
Smith, A.J., Massuel, S. and Pollock, D.W. (2012) Geohydrology of the Tamala Limestone Formation in the Perth Region: Origin and Role of Secondary Porosity, p. 63.
Treble, P.C., Bradley, C., Wood, A., Baker, A., Jex, C.N., Fairchild, I.J., Gagan, M.K., Cowley, J. and Azcurra, C. (2013) An isotopic and modelling study of flow paths and storage in Quaternary calcarenite, SW Australia: implications for speleothem paleoclimate records. Quaternary Science Reviews 64(0), 90-103.