



SDI FINAL EVALUATION FORM 1.1

PART 1:

Journal Name:	British Journal of Mathematics & Computer Science
Manuscript Number:	Ms_BJMCS_26476
Title of the Manuscript:	A Metric for Data Set Comparison Via QR-Factorization
Type of Article:	Short Research Article

PART 2:

FINAL EVALUATOR'S comments on revised paper (if any)	Authors' response to final evaluator's comments
The previous version of comments seems too implicit. I am wondering if the authors know orthonormal basis is not unique in a vector space. The main concept of this paper is based on this sentence "The work is conducted under the perspective that two matrices A and B will be considered as effectively the same if their ranges are the same." on page 103. If two matrices have the same ranges, they might quite different. For example, the identity matrix and any full matrix in $M_n(\mathbb{R})$. The notation about the matrix size in this revised version makes me very confusing. In page 103, if the author concern about the column basis, then we often have $m \leq n$ in practice (m is for dimension and n is for data size). In the case of $m \geq n$, the matrix is usually not full rank. Hence, the author should modify the first paragraph of page 103 very carefully. In page 104, equation (3.5), the full rank is a very important criterion. I recommend the author emphasize this criterion again. One problem in the symmetry property is that the matrix size become square. Why? In practice, the matrix usually not be square. And in page 105, equation (3.9), the matrix A becomes none square. Why? Please don't use the notation A and B in equation (3.8) to illustrate the face $(AB)^* = B^*A^*$. This makes the reader confusing. In page 108, the notation $sp(A_1)$ is not consistent to $S(A)$.	We are unclear on what the evaluator means by "too implicit" This is true, however both the $n \times n$ identity matrix and any full rank $n \times n$ matrix form a basis for the same space. We agree that there may be many orthonormal bases for a vector space. By construction, however, the orthonormal matrices of a QR factorization are unique (with aforementioned constraints) and the argument is that this particular basis may be used to compare the data sets. We have removed the sentence and merged the surrounding sentences. We perform the calculations with data sets that are overdetermined, though we are not solving $Ax=b$, as is often the application of QR factorizations. We assume that we have more data points than dimensions. We agree. It has been added. We do not require A to be square, it has been made consistent across sections. As the reviewer pointed out earlier, there are typically more data points than dimensions in A. The notation has been updated to reflect generic matrices U and V to avoid confusion. The notation has been updated to be consistent throughout the article.