

Addressal of Reviewer Feedback

Joe Zemmels, Susan VanderPlas, Heike Hofmann

Reviewer Comments #1

The authors have not addressed the concerns of the initial review and have in some ways doubled down on their positions. The authors have impractical expectations for algorithm research and an incorrect view of the CMC methods. The NIST CMC methods are not a polished finalized approach for 3D surface comparison. They are works in progress. The fact that details of their algorithm change from paper to paper reflect ongoing improvements. It is not realistic for the authors to create a package (cmcR) that will be a perfect implementation of CMC for use in crime labs with actual casework. The method is not finalized yet.

By no means do we expect the CMC algorithm to be finalized or that a “perfect” implementation does or will ever exist. We appreciate that the NIST authors have continued to iterate on and improve previous versions of the CMC methods. Our criticism of the changes made from paper-to-paper is that there is no discussion of why the changes were made, which connects to the obstacle of reproducing the authors’ work. Our expectations are simply that an algorithm, specifically in a high-stakes application like forensic science, be reproducible, and we respectfully disagree that this expectation is “impractical.” Stated plainly, our overall stance is as follows: if the code or data exist, then you should share it. Having to re-implement an algorithm or re-collect data that already exists is a waste of time and resources.

I am confused about the authors expectations and I believe the overall tone is still inappropriate. The authors are criticizing published progress of an ongoing research project.

Our criticism is intended to extend only so far as it comes to the computational reproducibility of the CMC algorithms in the published works. Our expectations are the same as with any rigorous scientific method: that the results be reproducible. As stated in the paper, by “reproducibility” we do not mean the qualitative reproducibility that we have been able to achieve in the paper. We mean reproducibility in the sense defined by [NASEM] provided in the paper.

The NIST authors should be encouraged to continue to publish continual improvements to their methods rather than waiting until there is a single final approach. It is more important to get updates as the algorithm is revised, improved, and tested than it is to have a paper about an open-source implementation of this work. The accurate comparison of surface topographies is a very difficult research challenge. Comparison is affected by the firearm type used, ammunition, caliber, etc. . . . The discipline is still far from a simple clean algorithm that works in all situations.

We completely agree that it is important to continually improve upon previous versions of the algorithm and that comparing surface topographies is a difficult task. We respectfully disagree with the assertion that publishing updates to an algorithm is “more important” than publishing an open-source implementation of the algorithm. We believe that it is just as important to get a method in the hands of many for the sake of experimenting with and validating the method than it is to develop improvements upon the method. As Reviewer 1 states [below], “[i]t is now up to additional research groups to take the initial results and extend them.” We believe that this is precisely what we have accomplished in this paper.

The feeling I get is that the authors believe that NIST has such a ‘final algorithm’ but that they are hiding details preventing others from implementing it. Don’t get me wrong, I commend the authors for creating an open-source implementation of the CMC methods; however, (and I could be wrong) I do not believe the creation of an open-source tool merits a peer-reviewed journal paper. It would be a paper if the authors created the open-source tool and used it to answer a research question (as they start to do see the discussion

of the five dimension that have a demonstrable impact below).

We do not believe that a “final algorithm” exists, nor will ever exist in the sense that development and improvement of the algorithm will cease. We do believe that NIST has an implementation of the algorithm that they have used to obtain publishable results and, as such, should also publish their code for the sake of reproducibility. As Reviewer 1 states [below], the work published thus far by NIST is a “conceptual framework” (we use the term “conceptual description” in the paper) for the CMC methods rather than an actual, reproducible algorithm. We argue that an actual published implementation, such as the one available in cmcR, is much more useful, in the literal sense that it can actually be used, to the forensics community. NIST could have provided their implementation code and data along with their conceptual framework and a caveat that it is a work in progress, yet they chose not to.

The authors responded to the initial review by saying that their goal was to “show that our implementation could produce results that are qualitatively similar to published results”. And they say that they achieved this result “we have been able to qualitatively reproduce the results of the original implementations.”. It is therefore confusing that they also state “we assert that we were not able to obtain “all the information [we] needed for the CMC implementation from the published CMC papers.”. I really don’t understand how they both were and were not able to implement the methods.

See the definition of reproducibility provided in the paper.

They go on to say “Our implementation, while providing qualitatively similar results to those in published papers on the Fadul and Weller data sets, is based on a large amount of guesswork and lengthy experimentation due to the ambiguity in the original CMC papers.” To me this says that the NIST algorithms have been published in sufficient detail to be reproduced.

See the definition of reproducibility provided in the paper.

It is unrealistic to expect that every detail of the algorithm be explicitly specified. For example, the authors seem shocked and upset that they know of no “open-source implementation of the multivariate band-pass Gaussian regression filter used in surface metrology (ISO 16610-71,2014)”. They specifically cite the ISO standard that describes the mathematical method. The math described in the ISO standard is explicitly what is used. As one further example, the authors state “It is a fact that rotating the contents of a 2D array by any angle other than a multiple of 90 degrees requires interpolation of, and therefore a change in, the values of the array. This requires the user to choose the algorithm used to perform the interpolation” but it is unrealistic to expect algorithm papers to describe how they do their rotation. There are several methods for rotation that would have no effect on the CMC score because the CMC method is somewhat “coarse”. The authors would have to use a very non-standard rotation method to affect a correct CMC implementation.

We fully agree that a human-readable description of an algorithm should not contain every detail of an implementation. However, without the precise details it is difficult, if not impossible, to computationally reproduce the results of an algorithm. This is why we argue in the paper that authors need to supplement their human-readable description with code and data used to arrive to published results if they want their method to be reproducible.

The authors state “While it is clear that the creators of the CMC pipeline have a working implementation, the wider forensic science community only has access to conceptual descriptions of the pipeline and some summary statistics describing its performance.” and “we argue that it is not acceptable in a forensics/legal setting to only provide a conceptual description of an algorithm for which there clearly exists an actual implementation (i.e., a “step by step approach”).” I disagree with this assumption. The NIST authors do not have a working implementation that is sufficient for use by others in the forensic community.

Their code is research level and runs far too slowly for use in practice. The CMC methods are not “camera ready algorithms”. They are ideas meant to spur additional research. They are intended to inspire researchers (like the authors) to build off the basic framework, to make changes, and to test on large datasets. NIST developed the conceptual framework for these approaches. It is now up to additional research groups to take the initial results and extend them.

Wouldn't it be mutually beneficial for other researchers to have access to the "research level" version of the algorithm so that they can, for example, determine if alternatives will improve the processing speed?

The authors point to a few new figures (13, 14) to demonstrate the correctness of their implementation by reporting a variance ratio. I am not convinced that this is a good way to compare results; why not use a more established method? Boiling the CMC results from an entire dataset down to a single point is insufficient to demonstrate the correctness of their algorithm.

I agree with the authors that no paper is perfect and that there are questions one could ask of the NIST researchers after having read their papers. But I disagree with the overall tone and assumptions made in this manuscript. As the other reviewer suggested, the authors could directly contact NIST (as can anyone). The authors replied that they have since reached out and have obtained additional information. This is how research works.

In conclusion, the manuscript seems to try to cover too many bases (describing reproducible research, criticizing the NIST papers, describing the CMC methods, describing their implementation, assessing the dimensions that have demonstrable impact, and demonstrating that their implementation is qualitatively similar to the NIST papers). The best part of the paper is the exploration of conditions that have "demonstrable impact on effectiveness". This is a worthwhile endeavor and I would encourage the authors to drop the description of the NIST method, the attack of the NIST papers (including details like criticism of missing Gaussian filters and missing rotation descriptions) and rather focus on the exploration of these effects (the five dimensions listed) on CMC. It is more common for a research paper to describe the research and then simply mention that their implementation is available open-source.

I do not believe the paper has been sufficiently modified to address the issues preventing publication.

Reviewer Comments #2

1. Introduction, 2nd to last sentence: Could you please acknowledge estimation procedures that use sampling here, for example MCMC. It would be remiss of me to let this sentence sit as it is and have my forensic colleagues repeatedly hammered in court because MCMC estimates are rarely, if ever, identical, but do fall within a certain range of variation (if the sampler has converged).

Setting a random seed is a sufficient condition for ensuring computational reproducibility of MCMC or any other random sampling procedure.

2. "open-source pattern-matching algorithms in forensics (glass (Park and Carriquiry, 2019),": I have two issues with this sentence. Firstly the goal of forensic glass analysis does not involve pattern matching. It may have a "match step" - but that is not essential as it is neither required by the court nor necessary for the computation of a likelihood ratio. Secondly, Carriquiry and Park are not expert in forensic glass analysis, nor is their work definitive. There are plenty of other papers. Evett's first paper (1977) describes in good detail how the t-test may be used in this situation. Lindley's subsequent paper in 1978 gives a Bayesian solution - again easy to implement. There are also papers by Curran and by Lucy both of which come with free open source software. I also think this sentence is untrue "the vast majority of proposed methods do not adhere to the definition of reproducibility in National Academy of Sciences, Engineering, and Medicine (2019)." and antagonistic and does not need to be in the paper.
3. 'firearm's unique "fingerprint" left on a cartridge case (Thompson, 2017).' - I realise you're citing a paper, but uniqueness is an unsupportable, and unnecessary requirement in forensic science. It would be good to note this. David Kaye has some good thoughts on this topic Law, Probability and Risk (2009) 8, 85-94.