Response to reviewers

Eric Hare, Heike Hofmann, Alicia Carriquiry

October 24, 2016

Reviewer 1

1. The authors have not provided a comprehensive literature review, to the point that the paper appears to be taking credit for the work that has been done by Dr. Nick Petraco, his collaborators, and others. I provided the references for Dr. Petraco's work in my last review.

   Most of the references that were suggested by reviewers earlier were incorporated into the revised version of the manuscript. Regarding the "overlap" with Petraco's work, please see the attached document entitled x3prplus for an in-depth discussion of our use of exactly one function in Petraco's package.

2. As I read the paper it is unclear how the matching criteria is different than the matching criteria currently in use by the tool mark's community, especially those proposed by Monkres et al (2013) in Comparison and Statistical Analysis of Land Impressions from Consecutively Rifled Barrels and the PhD dissertation work published in Technometrics by Amy B. Hoeksema with her PhD advisor Max Morris in 2013. Dr. Hoeksema's dissertation also has a wonderful explanation of Chumbley's U-statistic, which appears to be similar to the author's proposed statistic, which is a PhD dissertation in the department of the authors of the paper currently under review. My opinion is that a discussion and comparison of these classical methods with their new comparison statistic is necessary to demonstrate how the author's work advances the science of land impression comparisons.

   First, it seems more appropriate to refer to Hoeksema's U statistic since she was the one to propose its use in her dissertation. As the reviewer notes, this work was carried out in our own department, and we were well aware of it. Thus our surprise with the review, since to say that our work is a repeat of Amy Hoeksema's work is so inaccurate as to make us wonder whether the reviewer is familiar with either approaches. On October 11th, 2016, we sat down with Max Morris to discuss the similarities and differences between the Hoeksema U-statistic, and the methods used and features derived in our paper. Max was at least as surprised as we were to read the reviewer's comment, since the differences between Amy's work and ours are stark. The Hoeksema U-statistic characterizes toolmark signatures by producing two sets of correlation values. One set is produced by doing "coordinated shifts" after finding the ideal matching window, and computing each correlation value. The other set is computed by doing "independent shifts". When the two toolmarks match, we'd expect that the correlation values from the coordinated shifts would be significantly higher than those computed from the independent shifts, and this is assessed using a Mann-Whitney U statistic. While very interesting and worth further discussion, Max Morris agreed that this is far different from the methods we have proposed. In our work, we use full bullet signatures from pairs of bullets and extract several features based on these. The features include a cross-correlation, as well as the number of matching striae, the number of consecutive matching striae, the difference in the height of the signatures at each peak and at each valley, and several more. At no time is a Mann-Whitney U statistic used, windows extracted, or any coordinated/independent shifts used. These are completely different approaches where the only "similarity" is that we also compute a correlation, albeit between different variables!

3. The main methods used in the paper under review are extremely dependent on Dr. Petraco's work (implemented in an R library x3pr and available at https://github.com/npetraco). The authors built a new library titled x3prplus (available at https://github.com/heike/x3prplus) on top of x3pr. While

working on my review, I went through a large amount of the two R libraries on github. The new library mainly appears to be wrappers for Dr. Petraco's work. I find it very disconcerting and highly inappropriate that the authors are relying on Dr. Petraco's work to such a high degree and that they are not referencing his work in their paper.

The reviewer must have looked at the wrong code because no one who is reasonably familiar with R would claim that x3prplus is a "wrapper" for x3pr! Please see the attached document for a rebuttal to these claims. In x3prplus, all we did is call ONE of Petraco's functions, read.x3p, to read 3D files. Everything else is completely different, as can be seen in the attached document. Perhaps our choice of name for the package has contributed to the confusion.

Reviewer 2

4. This paper is uninteresting because it is not novel, even when considering the applied nature of the proposed work. It is very common for statisticians recently involved in forensic science to assume that nothing has been done prior to their own involvement, and this seems to be the assumption made by the authors. Unfortunately, much work on the subject of automated bullet comparisons has been done and published (but not necessarily in the statistical literature) and the authors of this paper are simply re-inventing the wheel without providing anything new to the debate.

   We thank the reviewer for his/her lecture, but it is difficult to take his/her comments seriously when they are offered with absolutely no evidence. Exactly whose work are we re-inventing? And where has work similar to ours been published? If the reviewer is going to make statements such as the ones above, please, try to back them up with some facts.

5. The paper is pedestrian and incomplete because extremely similar work has been proposed by Jain, Chumbley, Prokoski, Petraco, Zheng and so on (the list is available to this reviewed, who is happy to share it with the editor if needed). The authors even ignored a milestone paper published by their own department head... The authors of the current paper have not made any attempt to review the relevant literature: if anything, they merely took the references that I provided in my first review and added them to their bibliographic review, while dismissing their work. For example, the authors' treatment of Riva et al's work is not appropriate: it is not because they consider cartridge cases that their methodology could not be applied to landmarks on bullets! I believe that if the authors had bothered reviewing the literature they would not have ended up re-inventing the wheel and they may have actually done something useful for the community.

   Really? We have referred to many of those papers in the manuscript and have shown how our approach is quite different. Indeed, even PCAST (the President's Council of Advisors in Science and Technology) in its report, mentioned that our work was the only promising work and they actually looked at all the literature themselves. So we respectfully disagree with Reviewer 2. The Riva and Champod paper published in 2014 is indeed referred to in the manuscript and of all the published work, it is the one that most resembles what we have done. The reviewer poo-poos our saying that one difference is that Riva and Champod worked on casings and we worked on bullets. The reviewer would be well advised to read our comment more carefully; the reason it is important is that Riva and Champod make use of the circular structure of their data to pre-process their images and develop their score. A big component of their algorithm is the alignment of images, which they do by rotating the images on the xY plane. This is not something that cannot be extended to the case of bullets, were the XY plane is not flat and where rotation makes no sense. The approach they use to extract features from the aligned images is completely different from ours and for the reviewer to say that we are re-inventing the wheel is a bit of a scandal. The fact that Riva and Champod have come up with a nice automated algorithm to match breech face and firing pin images does not mean that we cannot come up with a DIFFERENT automated algorithm to match bullet lands. This said, we have added additional commentary about Riva and Champod in our manuscript.

6. The paper is scientifically questionable because (1) nothing in this paper is novel (as discussed above), (2) the authors are extremely confused as to what type of inference they are proposing. They emphasize that legal and scientific scholars advocate the use of a Bayesian framework for the quantification of the weight of forensic evidence, yet the entire paper is about calculating error rates. I assume that the authors understand that in order to make an error, one has to make a decision, and whether in a Bayesian or frequentist framework, a decision is related to posterior probabilities and utility functions, not to Bayes factors. Bottom line, if the authors are endorsing the use of a Bayesian framework, they spent a great deal of effort addressing the wrong problem. If the authors are not endorsing a Bayesian framework, then they should say so, and scientifically support their work on based on its own merits.

Again, we have to disagree with the reviewer and again we take issue at his/her bandying about such incendiary comments with nothing to back them up. We are not confused, in particular not on the issue of inference. In the intro we mentioned the type of inference that is often drawn in legal settings, mostly to satisfy an earlier reviewer (an actual constructive reviewer) who requested that we add a bit of general background in the introduction. We then proceed to explain that this is not what we pursue in our paper; all we want to present is an automated algorithm to decide whether two bullets may have been fired from the same gun. No more and no less. Yes, we are focusing on error rates, and for the reviewer's peace of mind, yes, we realize that there is a decision implicit in our trying to compute an error rate. The decision is simply whether two bullets may have been fired from the same gun. We have gone back to the manuscript and added some additional language to explain the purposes of the work we present.

7. The paper is ethically problematic because it seems that most of the code made available by the author is similar to, if not require the use of, legacy code provided by another author working in the same area, and who is absolutely not referenced in the current paper.

We referenced the OpenFMC organization and devoted a paragraph to its discussion. This is the organization that is the recommended citation for the package which we use a function of, `x3pr`. We have added a reference to Dr. Petraco directly in this discussion. See the below citation, given by `x3pr` as of October 17th, 2016:

```
citation("x3pr")
```

```
##
## To cite package 'x3pr' in publications use:
##
##   OpenFMC (2014). x3pr: Read/Write functionality for X3P surface
##   metrology format. R package version 1.0.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {x3pr: Read/Write functionality for X3P surface metrology format},
##     author = {{OpenFMC}},
##     year = {2014},
##     note = {R package version 1.0},
##   }
##
## ATTENTION: This citation information has been auto-generated from
## the package DESCRIPTION file and may need manual editing, see
## 'help("citation")'.
```

8. Overall, this paper does not advance the field of forensic science one bit: (1) the statistical analysis proposed by the authors is not novel at all, thus we are not learning anything from it; (2) the final results reported by the authors are useless since (i) they are related to the wrong type of inference (error rates vs. Bayes factors, and (ii) they are based on such a small and distorted dataset that they can't even be used in court to support the admissibility (or rejection) of firearm evidence.

We are very clear about the limitations presented by a small dataset, particularly in the conclusion. Unfortunately, this are the data that we had to work with. With that said, the idea of statistical modeling is that once new data are obtained, it will be rather trivial to compute the set of features used in our model, and retrain the model with a broader and more varied dataset. Whether performance on holdout samples would remain as strong remains to be seen. But because this code is open-source and in the public domain, the algorithm can be assessed on new data by anyone willing to invest the time to run it. While the dataset is small (only 10 barrels) we have been told by multiple firearm

examiners that it is a challenging dataset on which to develop methods because the 10 barrels were consecutively manufactured. Thus, one would expect that thre striae made by those barrels will be more similar across them than across two randomly selected barrels from a large population. We are collaborating with Alan Zhang in NIST to assemble a larger database that will be publicly available within a year or two.

Reviewer 4

9. This paper describes a statistical experiment to to determine whether two bullets were fired from the same gun barrel. The authors formulate a simple model, and perfom a version of least squares.

   We feel this is a rather inaccurate summary of the work done in our paper. We do use a random forest model, but this is not the focus. Rather, the data analysis steps culminating in feature extraction are the focus. These features, computed on pairs of bullet lands, are then fed into a random forest model, but we do formulate this model. Furthermore, we do not perform a version of least squares - The random forest uses decision trees. (Is the reviewer thinking of regression trees? If so, that is not what we are implementing.) One purpose of our work is to highlight the individual data analysis steps that compromise a full analysis of bullet striations, beginning from an open-source data file, and ending with a predicted probability of a match.

10. No new statistical methodology is introduced, so the impact of this work is very narrow. There is also no attempt to relate the work to any previous statistical studies or tests, as there is only one reference.

    Did the reviewer read the correct paper? The revised copy of our paper contains over thirty references, not just one. Many are from previous statistical studies on bullet matching (e.g., Biasotti 1959, Chu 2010, Riva 2014) while many more are references on statisitcal tests themselves (e.g., Therneau 2015, Liaw 2002, Cleveland 1979)

11. I believe that this work should be submitted to a more specialized journal, but Annals of Applied Statistics is definitely the wrong venue.

    We believe that the focus on a timely application and the novel combination of statistical and machine learning methods to address the problem warrant publication in the Annals of Applied Statistics.

Reviewer 5

12. The authors have addressed the majority of my comments. Exceptions are: Comment 1): The authors corrected the second sentence in the abstract as suggested, yet did not correct the first sentence of the Introduction, which was a copy-paste of the second sentence in the abstract, and this first sentence in the Introduction is now still incorrect. This sentence, currently reading "Firearm examination is a forensic tool used to determine whether two bullets were fired from the same gun barrel." should be replaced by "Firearm examination is a forensic tool used to help the court determine whether two bullets were fired from the same gun barrel." in order to be correct and consistent with the abstract.

    Thank you. We apologize for the oversight, and have made this correction.

13. Comment 2): The weakness of the paper is that it lacks an explanation for why a binary "match" vs "non-match" approach was chosen. It does not explain what the value of the observations is with regard to the two competing propositions, i.e., the question of interest to the forensic scientist. If the authors want this to be a good paper, I encourage them to provide such an explanation.

    We have attempted to extend this discussion. Although we have assessed our algorithm in terms of a binary "match" vs. "non-match" scale, we didn't attempt to reduce the full scope of the problem down to this decision. As a matter of fact, the output from our model are predicted probabilities of a match, or scores from the random forest. These scores can be seen in Figure 19. For purposes of assessment, we considered a predicted probability of at least 50% as indicative of a match. However, in a real world application of this algorithm, the scores themselves are most certainly relevant. We apologize that we didn't make this distinction as clear in the previous revision, and we hope this alleviates some of the reviewer's concerns.

14. Comment 3): The term "ballistics" was corrected as suggested. Yet there remain two occurrences of "ballistic" as an adjective in lines 477 ("ballistic images") and 490 ("ballistic imaging"). I interpret these as referring to images or imaging of the trajectory of bullets. If this is not what was meant, I strongly encourage the authors to replace "ballistic images" and "ballistic imaging" with phrases that represent what they actually meant.

    We have corrected the use of ballistics in these instances.

15. Comment 7): An attempt was made to visually distinguish between the two bullets, yet note that the shading that was added is barely visible when the figure is printed on a color laser printer.

    Thank you for pointing this out, as we did not attempt to check the printed copy of the paper. We have corrected this to improve the printed appearance.

Reviewer 6

16. The revised paper "Automatic Matching of Bullet Lands" describes an automated procedure for matching bullets that have been fired from the same gun. It also compares the accuracy of the procedure on a test set of bullet images from the Hamby, Brundage and Thorpe study. The key conclusion is that the automated algorithm outperforms human matching, and achieves high accuracy. The methods used in the paper are statistically simple, but also robust. The problem is important and the paper fills a need.

    We thank you for the kind words.

17. Some aspects of the algorithm are judgmental, and I expect that at each step in the algorithm one might be able to find an alternative that is incrementally smarter. But I think that is permitting the best to be the enemy of the good.

    We agree wholeheartedly with the reviewer. Several best judgment decisions were made during the course of the algorithm. We attempted to be explicit when certain parameters were chosen (e.g., the smoothing factor parameter) and include a discussion of how the results change when these parameters are varied. As such, thus algorithm could be modified to use best parameters as chosen from cross-validation, for instance.

18. The paper is clearly written and sufficiently concise. The description and motivation for the work is good. And the code is in the public domain, which is a great contribution.

    Thank you once again for the comments.

x3prplus

Background

We have compiled this document in response to a serious accusation raised by a reviewer of our paper Automatic Matching of Bullet Lands. The accusation is reproduced below:

*The main methods used in the paper under review are **extremely dependent on Dr. Petraco's work** (implemented in an R library x3pr and available at https://github.com/npetraco). The authors built a new library titled x3prplus (available at https://github.com/heike/x3prplus) on top of x3pr. While working on my review, I went through a large amount of the two R libraries on github. **The new library mainly appears to be wrappers for Dr. Petraco's work**. I find it very disconcerting and highly inappropriate that the authors are relying on Dr. Petraco's work to such a high degree and that they are not referencing his work in their paper.*

In this document, we will show quite easily that this accusation is demonstrably false. To begin, we can easily access all exported functions of the x3pr package with a call to `getNamespace`. If we do so for x3pr, we obtain the following list:

```
ls(getNamespace("x3pr"))
```

```
##  [1] "launch_3Dview"          "lms2x3p.file"
##  [3] "make.x3p.header"        "plot3D.x3p.file"
##  [5] "print.x3p.file.info"    "read.digital.surf.file"
##  [7] "read.digital.surf.header"  "read.digital.surf.profiles"
##  [9] "read.x3p"               "read.zeiss.lms.file"
## [11] "read.zeiss.lms.header"  "read.zeiss.lms.profiles"
## [13] "sur2x3p.file"           "write.x3p"
```

Performing the same task for x3prplus:

```
ls(getNamespace("x3prplus"))
```

```
##  [1] "boot_fit_loess"      "bulletAlign"         "bulletAlign_nist"
##  [4] "bulletCheckCrossCut" "bulletGetMaxCMS"     "bulletGetMaxCMS_nist"
##  [7] "bulletSmooth"        "CMS"                 "fit_loess"
## [10] "fortify_x3p"         "get_bullet"          "get_crosscut"
## [13] "get_grooves"         "get_peaks"           "get_peaks_nist"
## [16] "getCircle"           "getTwist"            "maxCMS"
## [19] "plot_3d_land"        "predCircle"          "predSmooth"
## [22] "processBullets"      "read.x3pplus"        "sample.x3p"
## [25] "smoothloess"         "striation_identify"  "unfortify_x3p"
```

We will use these functions to illustrate the vast separation of functionality between the two packages.

Function Similarities

Did we attempt to take credit for work that was done by Dr. Petraco? Below is a list of all the functions available in the x3pr package (accessed 10/10/2016). Further down, you can see the list of all the functions available in x3prplus. As is clear from the function names and descriptions, **there is virtually no overlapping functionality**. In general, Dr. Petraco's functions are about reading and parsing file formats. Our functions assume the reading step has already taken place, and compromise a much broader set of statistics and data science functionality including fitting models, deriving features, detecting grooves in bullet lands, and more.

x3pr Function	Comparable x3prplus Function	Description
launch_3Dview	None	Shiny app to view a 3D surface
lms2x3p.file	None	Convert a Zeiss LMS file to an X3P format
make.x3p.header	None	Put together the XML format x3p header
plot3D.x3p.file	None[1]	3D plot of a surface in X3P format
print.x3p.file.info	None	Format (most) x3p header file info
read.digital.surf.file	None	Read in a Digital Surf file
read.digital.surf.header	None	Read the header off a Digital Surf file
read.digital.surf.profiles	None	Read in the surface data of a Digital Surf file
read.x3p	read.x3pplus	Read an x3p file
read.zeiss.lms.file	None	Read a surface file in Zeiss LMS format
read.zeiss.lms.header	None	Read the header off a Zeiss LMS file
read.zeiss.lms.profiles	None	Read out surface data in Zeiss LMS format
sur2x3p.file	None	Convert a read in .sur file to x3p
write.x3p	None[2]	Write a x3p format file

1: `x3prplus` includes an internal function, `plot_3d_land`, which uses the plotly package to produce a surface plot of a bullet land. This function is not accessible to users and uses no functionality from the `x3pr` function `plot3D.x3p.file`.

2: The `unfortify_x3p` function converts a data frame representation of a bullet land into the surface matrix version native to `x3pr`. However, this does not depend on the `write.x3p` code, nor does it have any of the disk writing functionality.

Function Wrappers

It is rather trivial to assess the claim that `x3prplus` "mainly appears to be wrappers" for `x3pr`. In fact, we can very simply determine the unique number of `x3pr` functions in use by `x3prplus` to determine whether there is any evidence that `x3prplus` is merely a wrapper.

x3prplus Function	x3pr Function Used	Description
boot_fit_loess	None	Fit a LOESS model with bootstrap samples
bulletAlign	None	Align two surface cross cuts according to maximal correlation
bulletAlign_dtw	None	Align two surface cross cuts according to maximal correlation
bulletAlign_nist	None	Align two surface cross cuts according to maximal correlation
bulletCheckCrossCut	None	Identifying a reliable cross section
bulletGetMaxCMS	None	Identify the number of maximum CMS between two bullet lands
bulletGetMaxCMS_nist	None	Identify the number of maximum CMS between two bullet lands
bulletSmooth	None	Smooth the surface of a bullet
CMS	None	Table of the number of consecutive matches
fit_loess	None	Fit a loess curve to a bullet data frame
fortify_x3p	None	Convert a list of x3p file into a data frame
get_bullet	None	Deprecated function use get_crosscut
get_crosscut	None	Read a crosscut from a 3d surface file
get_grooves	None	Find the grooves of a bullet land
get_peaks	None	Identify the location of peaks and heights at a crosscut
get_peaks_nist	None	Identify the location of peaks and heights at a crosscut
getCircle	None	Estimate center and radius
getTwist	None	Estimate the twist in a bullet land
maxCMS	None	Number of maximum consecutively matching striae
predCircle	None	Estimate predictions and residuals for a circle fit of x and y
predSmooth	None	Estimate predictions and residuals for a smooth of x and y

x3prplus Function	x3pr Function Used	Description
processBullets	None	Process x3p file
read.x3pplus	read.x3p	Read X3P File with option to transpose matrix
sample.x3p	None	Sample every X element of a data frame
smoothloess	None	Predict smooth from a fit
striation_identify	None	Match striation marks across two cross sections
unfortify_x3p	None	Convert a data frame into an x3p file

In total, **there is exactly one single function call made to an `x3pr` function**. This is a function we wrote, `read.x3pplus`, which extends `read.x3p` with a simple option to transpose the resulting matrix. Not a single other `x3pr` function is ever called in `x3prplus`. Comparatively, we call dozens of unique functions belonging to the `ggplot2` package. It is clear that we aren't attempting to be a wrapper for `x3pr` anymore than we are a wrapper for `ggplot2`. In fact, the set of functionality of both packages is wildly different, as can be seen from a simple read of the function descriptions in the two tables above.

In other words, the claim that `x3prplus` is "extremely dependent on Dr. Petraco's work" is demonstrably false. As a matter of fact, we have an internal version of the code which replaces the single use of Dr. Petraco's code, in our `read.x3pplus` function, with a natively written XML parser. Nonetheless, we would prefer to build on Dr. Petraco's work, which is most certainly in the spirit of open-source software and academic research. This is also likely why Dr. Petraco has made his code publicly available on GitHub.

Referencing x3pr

The question remains of whether despite only using a single function of `x3pr`, and our package bearing virtually no similarlity in terms of functionality, did we fail to provide proper attribution of the package. Dr. Petraco contacted Eric regarding this issue, in an email which was CCed to his major professor (Dr. Heike Hofmann), the director of CSAFE (Dr. Alicia Carriquiry), the Department Chair of Statistics at Iowa State University (Dr. Max Morris) and the CSAFE contract officer at NIST (Susan Ballou). Dr. Petraco claimed "I am puzzled however as to why no citation appears to this work anywhere in your text as `x3pr` contributed critically to it." The paper referenced was a pre-publication version published to Arxiv, and was still in the revision stage. Eric acknowledged the omission of the direct reference to `x3pr` in the pre-publication paper, and added it that afternoon, within hours of receiving the email:
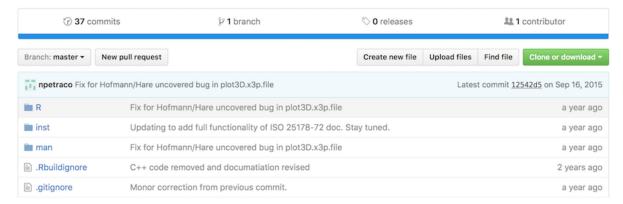
https://github.com/erichare/imaging-paper/commit/c63cfeee7134222a9b760752fca80084aadad866

Further, numerous other references to Dr. Petraco and `x3pr` have existed within the context of our package `x3prplus`, including but not limited to:

1. The README for the package: https://github.com/heike/x3prplus
2. Presentations given on x3prplus and our algorithm
   - https://github.com/heike/x3prplus/blob/presentations/jsm-2016/slides.Rmd
   - https://github.com/heike/x3prplus/blob/presentations/graphics-group-2015/slides.Rmd
   - https://github.com/heike/x3prplus/blob/presentations/forensics-group-late2016/slides.Rmd
3. Within the package description itself.
4. Directly in the name of the package `x3prplus`

As a matter of fact, we submitted a pull request in order to fix an issue in `x3pr` which was detected over a year ago. Dr. Petraco merged it in to fix the bug, as can be seen by this screenshot taken October 11th. This is also the most recent change that Dr. Petraco has made to his package, and predates serious work on `x3prplus`.

Basic read/write capability for the x3p surface metrology format in R

ⓟ **37** commits	⑂ **1** branch	◇ **0** releases	⚏ **1** contributor

Branch: master ▾	New pull request		Create new file	Upload files	Find file	Clone or download ▾

⫟ **npetraco** Fix for Hofmann/Hare uncovered bug in plot3D.x3p.file		Latest commit 12542d5 on Sep 16, 2015
📁 R	Fix for Hofmann/Hare uncovered bug in plot3D.x3p.file	a year ago
📁 inst	Updating to add full functionality of ISO 25178-72 doc. Stay tuned.	a year ago
📁 man	Fix for Hofmann/Hare uncovered bug in plot3D.x3p.file	a year ago
📄 .Rbuildignore	C++ code removed and documatiation revised	2 years ago
📄 .gitignore	Monor correction from previous commit.	a year ago

Although using `x3pr` merely to read in the data, we felt that it was important to reference the work that had been put into it. Given this, the reviewer finding our package "very disconcerting" and "highly inappropriate" is quite frankly a very strange assertion, particularly given the lack of any evidence provided in the review.