A note on the use of simple suspect list metrics in the evaluation of profiling methods and Investigative Procedures*

Richard Z. Gore, PhD
Professor Emeritus
UMASS Lowell
Richard_Gore@uml.edu

Comments on the NIJ project to create Geographic Profiling Software Evaluation Criteria
Submitted to NIJ – August 24, 2005

*Do not cite without permission: Manuscript containing material in this note under review.

A note on the use of simple suspect list metrics in the evaluation of profiling methods and Investigative Procedures*

Richard Z. Gore, PhD
Professor Emeritus
UMASS Lowell
Richard_Gore@uml.edu

Introduction

The NIJ charged a committee to come up with recommendations (Rich & Shively, 2004) for the establishment of performance measures to evaluate Geographical profiling software. The committee failed to settle on a single measure and offered several performance metrics. This result was probably necessitated by the committee's recognition of the limitations inherent in all the methods suggested, particularly when wider practical issues about operational alternatives and force setting are present. Rossmo's (2005) response to the committee's evaluation suggestions included a host of spatial and statistical problems he identified with these suggested metrics. His conclusion that an area proportion method was the best evaluation approach seems sound as a solution to the basic question of which software package produced the most efficient estimate, from an investigative perspective, of the offenders basing point. However, Rossmo (2005) and Rich & Shively (2004) both indicated that other unaddressed issues exist around the question of geographical profiling evaluation. Rossmo (2005) noted that these metrics fail to adequately capture the total benefit that can be derived from the application of geographic profiling. He notes:

> Rossmo (2005, p10). "Another way of looking at this is to imagine a database containing 1,000 suspects in a sexual homicide case with crime scene DNA evidence. How many of these suspects need to be tested before the offender is identified? This is the appropriate test of a geographic profile's accuracy."
>
> Rossmo (2005, p. 13) "The geoprofile was used, amongst other criteria (e.g., uncorroborated alibi, similarity to a composite sketch, etc.) to help investigators prioritize an initial list of 312 suspects for DNA testing. The individual ultimately convicted of the crime series was ranked sixth out of 312 based on the geoprofile alone (he actually scored higher because of the other criteria)."

The Rich & Shively (2004, p. 6) report contains the statement:

"It was also noted that the presence of sophisticated automated computer systems

* Do not cite without permission: Manuscript containing material in this note under review.

> in law enforcement agencies offers the analyst a host of “data reduction” tools that may prove just as valuable (or more valuable) to crime analysts and investigators compared to geographic profiling software.”

The above statement from Rich & Shively (2004) is central to the broader and more important question regarding evaluating the worth of any procedure associated with the investigative process. Can evaluation criteria be established by which force decision makers could answer such questions as is a combination of data reduction techniques as or more investigatively efficient than using a specific geoprofiling package? Other examples of such questions regarding the investigative process might be, does the application of a specific geographical profiling method add benefit to a specific offender profiling ranking model or how much value is actually derived from allowing investigators to always work their hunches, informer information, tips and other investigative leads first as opposed to integrating these names as they appear into a new ranked list based on all available intelligence. At first glance these comparative evaluations might seem to be comparing apples and oranges or that it would take a major accounting effort to evaluate, an effort that would be both expensive and burdensome on investigators. These conclusions may not be entirely true. In the physical sciences numerous examples abound of how a simple change in analytical framework can turn a very complex analysis into a much simpler one. In the above examples we tend to see the problem as comparing quantitative results in a geographic framework with offender profile data from a qualitative framework, thus the apples and oranges conclusion. In real world administrative environments our education and experience has generally condition us to measure all such benefits in terms of cost savings. Since all of the elements underlying such questions generally have an associated cost, transforming the question into cost space offers a means to answer them. Although effective, evaluation by cost accounting is a bottom-up procedure that is expensive in money and time as well as generally expanding the paper work burden at all levels which can impact morale and productivity. The policing related questions posed above are generally associated with a common property other than cost and that is suspects. Can these questions find a satisfactory solution by a transformation of the evaluation process into suspect space?

Suspect Space

Many of the difficulties associated with the evaluations discussed above, in the committee report, and by the posted expert commentaries can be diminished by using Suspect Space metrics. Some of what follows may seem obvious but my intention is to illustrate a simple conceptual framework or way of thinking about the criminal investigative process in general and later explain how this is related to the problem of software evaluation. Further, I don’t differentiate between single and serial crimes as both are amenable to some form of spatial analysis and should both be considered under the heading Geographic Profiling applications (see addendum).

What is Suspect Space? Suspect Space is simply the list of all possible suspects for any crime (one crime or a crime series) existing in either a random or ranked state. Its range is from one to the whole population. There is no “no suspects or zero suspects”. A crime dictates an offender. The non-existing set element “Zero suspects” simply implies that everyone who might be included in a hypothetical suspect pool is still a suspect. It must be emphasized that at every stage of an investigation an actual or potential suspect list exists even if it is only implicitly defined**!!** Taking the set or collection "individuals" the goal of the investigative process is to use rational assumptions based on experience (statistical data) and any specific personal data (description, MO, relationship to victim, etc) to reduce the members of this universal set to a manageably sized subset that becomes a workable suspect list. It should be noted that presently this reduction step is often not sufficiently effective in producing a workable list with the result that some crimes are just recorded but never actively investigated. There is clearly one basic geographic restraint on the universal set "individuals". The universal set "individuals" is everyone on earth. For any crime there should initially be a hypothetical subset, called suspects, of this universal set. Without assumptions or actual intelligence information, the subset "suspects" would equal the universal set "individuals". The first step in any investigation is to reduce the hypothetical subset of suspects by some geographic restraint that defines the problem area. These boundaries can have a rational justification based on experience or may simply include the area containing all the events associated with the case that have a geographic expression augmented by some arbitrary buffer. It is now theoretically possible to generate an actual list of suspects for this study area. Without any case specific information, the initial suspect list will be drawn using some form of filter from past offenders. At the completion of an investigation the summary report should include documentation by the investigators and/or crime analysts of the filtering profiles and methods employed that is sufficiently detailed to allow a post-mortem recreation of all lists generated during the investigation. Additionally, files of the actual lists and the date generated and a list of all suspects produced from canvassing, informants, tips and the dates these names surfaced should also be available in the summary. During the actual investigation suspect lists might be generated for only part of the problem area incrementally guided by a geographic profile. It is essential to know the methods employed to generate those lists if the analysis proposed here is to produce meaningful results. This investigative book-keeping would not add much work to the analyst as these lists could be copied into a Summary folder as they are generated. A metadata file detailing the filter or technique used would be an additional task required during the investigation. This procedure might also save time occasionally by avoiding the "now what exactly did we do here" question when an investigation is looking for new direction.

In suspect space, the application of any data analysis method can alter the list by not moving, or moving suspects up or down, by shortening the list, or both. The value of any analytical method would be measured by how effective it was in moving a suspect, who upon detection is identified as the offender, toward the top, shorting the list or both.

The only significant measure of the value of any analytic technique to a police agency is the resources saved by using that technique versus not using that technique. If the costs of an investigation are conceptually divided into two categories, a suspect space metric can be a good surrogate in comparisons for the actual monetary saving for one of these. Category one of the investigative tasks would be defined as the process of creating and then reducing the list of suspects down to one suspect. The second category would include any additional investigative efforts that might be required to arrest that final suspect. Moving the suspect who later proves to be the offender up by ranking a list or shortening the list as opposed to working a non-ranked list clearly saves investigative time. Any list changing metrics would probably correlate well with the actual monetary cost saving in expense Category one. Although, the committee recommended metrics potentially can answer the limited question of which among a series of software packages gives the most effective theoretical result, they tell little about how effective the package will be in a particular force environment. In the end, police investigate and arrest people not parcels so the real measure of any technique is ultimately how effectively it helped to locate the offender not his basing point. Rossmo (2005) pointed out some spatial difficulties with metrics that only measure how effective a technique is in locating the basing point. He notes that there is not necessarily any direct correlation between the amount of area investigated to reach the actual offender and the number of offenders investigated. The implications of this is that the actual distance between an estimate of a basing point and the actual basing point and the ratio of the area searched to total problem area offer only a relative means of comparing geographic profiling methods not potential investigative impacts. Only when the potential suspect spatial distribution is uniform can these geographic measures yield any clue as to how effective the tool was in saving police resources.

Suspect List Metrics

How would the Suspect Space metric (SSM) actually be calculated? As mentioned above, any technique or its software actualization can alter suspect space in three ways. Therefore, data analysis techniques would have to be classed into one of these three categories. The software discussed in the report produces results that can only order or rank a suspect list after the initial defining of the problem area**.**

If a series of rapes committed by a male is assumed, initially it might be concluded that a suspect list containing all known male rapists in the problem area offered the highest probability of containing the actual offender. If this list was initially random and the offender was on it, then following the ordering of the list, using the geographic profiling method, a SSM could be defined. One suspect space list ranking metric (SSR) could be defined as the offenders list position as a former suspect from the top divided by the list length. This is simply a measure of what portion of the suspect list had to be investigated to reach the offender. After several applications of the method in different investigations, the SSR should be lower than .5, if the geoprofiling method is producing any real value. If the offender proves not to be on this list, then a new list, possibly consisting of all known male sex offenders, minus those already cleared (only necessary if the sex offender filter did not remove the rapists) could be worked up and ordered by the values on the software generated probability surface. At the top of this new ordered list, the list of previously investigated rapists would be placed but of course the investigation would proceed with the first suspect below the already investigated rapists. Should this next list fail to contain the offender, than a list of all male violent offenders might be tried. If this also failed, then the analyst might try a list of all males offenders under some age, dictated by victim evidence (at this stage, if not sooner, alternative list shortening strategies would be very helpful). With each new ordered list all suspects investigated on earlier lists would be placed on top. No matter what the criteria, the numerator of the SSR is known and an estimate of the denominator or total theoretical size of the problem area suspect list can be calculated. The metric described in the above hypothetical case would be a measure of the overall effectiveness of the investigation which included the ability of the investigation team to produce short lists containing the offender as well as the ranking ability of a particular geographical profiling package. A better measure of the software's value alone in this force setting would be to take the final list described above, order this whole list with the geoprofiling package and then calculate the SSR. Depending on what is being compared the determination of the denominator may require careful consideration but the data for this determination would come from the same sources as were used to establish the working lists for parts of the problem area. Those data sources are likely to be the force intelligence databases, those of cooperating agencies or in cases calling for more massive canvassing, the census data. If partly through an investigation the offender's bother-in-law turns him in, the SSR numerator could be determined by using what ever criteria was being employed to generate the suspects for each search cell until the cell containing the offenders now known basing point is reached. The denominator would be an extension of that procedure over the whole problem area.

In the future when effective list shortening techniques are discovered, their effectiveness might be evaluated as the ratio of list size after application of the technique to the original list size. Both of the suspect list metrics discussed here are somewhat comparable and may allow some valid comparisons between list ranking and list shortening techniques. Gore et al. (2005) offers an example of using an SSM to compare three methods of list ranking. It should be noted that not all geographic profiling methods have the same restraints imposed by the current crop of software packages and these (See Gore and Pattavina, 2004; Gore, Tofiluk, & Griffiths 2005) may offer an alternative approach in cases where there is only one to three incidents.

In addition to being a metric that allows for a measure of the operational effectiveness between geographic profiling software, it allows for the comparison of results between geographic and non-geographic data analysis methods as well. For example, weights based on offender statistics for a particular crime type and area could be applied to various suspect descriptive characteristics in order to generate a score that could be used to order the suspect list. When the offender is discovered the SSR for the offender profile ranked list would be calculated (SSR1). That list is also the random list used by the geographic profiling technique to generate its ordered list from which its SSR (SSM2) is calculated. These metrics offer a test of the proposition mentioned in the committee report that various combinations of data reduction methods might be superior to geographic profiling (I assume that the caveat "as currently practiced" is implied as we all hope for future progress). In essence, this would be a comparison of offender profiling with geographic profiling. If the multi-case SSR2 average had a consistently better (lower) value than the SSR1 average, we could state that the geographic profiling method being employed was more effective than the results derived from that specific weighted offender profile. The above evaluation would place no control on the methods employed in the actual investigation and there is no suggestion here of setting up experimental trials which can present serious ethical problems when a dangerous serial offender is at lodge. If the investigation's final report is sufficiently documented these "what if we had done" evaluations can be made post mortem. There are wrinkles (that generally have solutions) in cross method and between case comparisons that consistently have different sized lists. These variants can center on how the values of the denominators are to be determined so that the results are a measure of department savings and not simply a measure of how efficient each method was in working its respective list. It may also require a modified metric other than those discussed in this note. Comparisons between compounded methods could also be made with a SSM. This ability to allow comparisons between most means of solving crimes that involve a reduction and/or ranking of the suspect subset can't be too strongly emphasized. If the final report documentation is adequate, the closed cases can be used to test how effective newly proposed ranking tools of any type would have been in effecting the investigative efficiency of that closed case.

The cost associated with all data analysis methods often greatly transcends those of the cost of the software license. There is the time required to conduct the analysis and the training costs required to allow proficient use of the software by the operator. Personnel turnovers generate recurring training expenses. The Suspect Space metric creates information that can be used to effectively evaluate among and between many different investigative approaches. This potential may allow for the identification of the best mix of analytical methods that balance software related costs with software effectiveness.

Conclusion

Metrics based on Suspect Space would seem to offer the best means currently available to evaluate the relative effectiveness between various suspect list altering procedures in identifying an offender. Suspect list metrics are also probably the best currently available means to evaluate a procedures actual benefit to a police department in terms of investigative cost saving if a relative value of cost savings between alternatives is sufficient for decision makers. The use of a Suspect space metric such as the rank on a list or reduction in list size may seem a little strange at first but remember all the proposed measures in both geographic space and suspect space are only indirect measures of this benefit. That method, which best tracts the actual benefit, is relatively easy to implement, and is more universally applicable as a means of judging effectiveness across a wide range of existing and future methods of data analysis should be encouraged.

Addendum

I believe a more general definition of geographic profiling would also be beneficial to the formulation of an evaluation process. I propose the following ad hoc definition: "Geographic profiling is the process of identifying an offender or offenders of a crime or series of crimes by the application of spatial analysis to any geographic information capable of yielding results that can be used as the basis for ranking and/or reducing the size of the suspect pool. Current practice actualized in existing software assumes that an offender has a single basing point. It focuses on the generation of probability surfaces defined by cells that cover the area of investigation. Each cell contains a value of the probability that it contains the basing point. These values are then used as the basis for suspect ranking."

References

Gore, R.Z. and Pattavina, A. (2004) Applications for Examining the Journey to Crime Using Incident-Based Offender Probability Surfaces. *Police Quarterly Vol. 7 No. 4,* 457-474.

Gore, R.Z., Tofiluk, N.J. & Griffiths, K.V. (2005) Single Incident Geographic Profiling. *In Wang, F. (ed.), Geographic Information Systems and Crime Analysis. Idea Group Publishing* Chap 8. 118-136

Rich, T., & Shively, M. (2004). A methodology for evaluating geographic profiling software. Cambridge, MA: Abt Associates.

Rossmo, D.K. (2005) An Evaluation of NIJ's Evaluation Methodology for Geographic Profiling Software. Available on the MAPS website.