DOJ Forensic Discovery Policy

- **Laboratory case file:**

  - May include written communications between experts or between experts and prosecutors.

  - Must review these to decide which, if any, are protected and which should be disclosed under *Brady/Giglio*, Jencks, or R. 16.

  - Discuss best practices regarding how to preserve such electronic communications, per DOJ March 2011 e-communications memo

Practical Suggestions

- Pay Special attention to:
  - Cases with classified information
  - When forensic reports may reveal the identities of cooperating witnesses, undercover officers, or disclose pending covert investigations

- If circumstances warrant (e.g., tests in another case may be relevant), review materials outside case file.

Practical Suggestions

- Communicate!

- Take the time to learn the science

NATIONAL ATTORNEYS GENERAL
TRAINING & RESEARCH INSTITUTE
THE TRAINING & RESEARCH ARM OF THE NATIONAL ASSOCIATION OF ATTORNEYS GENERAL

Practical Suggestions

- **Fair and Thorough Discovery Practices Bolster both your Reputation and that of your office**

Proudly Featuring

Boar's Head

PREMIUM DELI MEATS & CHEESES

"Compromise Elsewhere"

NATIONAL ATTORNEYS GENERAL
TRAINING & RESEARCH INSTITUTE
THE TRAINING & RESEARCH ARM OF THE NATIONAL ASSOCIATION OF ATTORNEYS GENERAL

- **Opinions Supporting Defense Theory**
- *United States v. Seaman*, Not Reported in M.J., 2009 WL 4048019 N.M.Ct.Crim.App. 2009)
- *Murrell v. Giroux*, 2015 WL 5009022 (M.D. Pa. 2015)
- *Brim v. United States*, 2015 WL 1646411 (C.D. Calif. 2015)
- *United States v. Wood*, 57 F.3d 733, 737 (9th Cir.1995)
- *Benn v. Lambert*, 283 F.3d 1040, 1060-62 (9th Cir. 2002)
- *United States v. Severns*, 559 F.3d 274 (5th Cir. 2009)
- *United States v. Crasten*, 2013 WL 462056 (W.D. Va. 2013)
- *Batchilly v. Nance,* Not Reported in F.Supp.2d, 2010 WL 1253921 (S.D.N.Y. 2010)
- *Villasana v. Wilhoit*, 368 F.3d 976, 979 (8th Cir.2004)
- **Disclosure of Inconclusive Results**
- *United States v. Howard*, 516 Fed.Appx. 409 (6th Cir. 2013)
- *Green v. Swarthout*, Not Reported in F.Supp.2d, 2014 WL 3748618 (N.D. Cal. 2014)
- *Miaram v. People,* 2016 WL 373963 (E.D.N.Y.2016)
- **Impeachment of Expert—Credentials**
- *Hall v. Beard*, 2017 WL 1234212 (C.D. Cal. 2017)
- *O'Brien v. United States*, 962 A.2d 282 (D.C. 2008)
- **Evidence collection/protocol**
- *Overton v. Jones*, 155 F.Supp.3d 1253 (S.D.Fla. 2016)
- *United States v. Gonzalez*, 938 F.Supp. 1199 (D.Del. 1995), *aff'd* 127 F. 3d 1097 (3d Cir. 1997)
- **Failure to disclose serious misconduct in other cases**
- *United States v. Brown*, 2015 WL 1268159 (D. Mass. 2015)
- In re Oscar Ray Bolin, Jr., 811 F.3d 403 (11th Cir. 2017)
- **Impeachment of Expert—Under Investigation**
- *United States v. Olsen*, 704 F.3d 1172 (9th Cir. 2013); *Lindh v. Murphy*, 124 F.3d 899, 901 (7th Cir. 1997

RE: Boston Travel

From: "Ambrosino, Michael (USADC)" <(b) (6)   >
To: "Hunt, Ted (ODAG)" <(b) (6)
Cc: "Antell, Kira M. (OLP)" <(b) (6)
Date: Fri, 22 Sep 2017 11 50 43 0400

Ted: I totally understand your concern. (b) (5)

mta

From: Hunt, Ted (ODAG) [mailto (b) (6)
Sent: Friday, September 22, 2017]
To: Ambrosino, Michael (USADC) <(b) (6)
Cc: Antell, Kira M. (OLP) (JMD) <(b) (6)
Subject: RE: Boston Travel

Hi Mike,

(b) (5)

Thanks,

Ted

From: Ambrosino, Michael (USADC) [mailto: (b) (6)   ]
Sent Thursday, September 21, 2017 4 39 PM
To: Hunt, Ted (ODAG) <(b) (6)
Subject: FW Boston Travel

Ted:

Just wanted to put this on your radar. (b) (5)

Is there anything we can do here to get Bruce to this conference? mta

From: Budowle, Bruce [mailto (b) (6)   unthsc.edu]
Sen Thursday September 21
To: (b)(6) Shelly Cox US Courts
Cc:] <(b) (6)   >; Daniel Capra <(b) (6) fordham.edu>; Antell, Kira M. (OLP) (JMD) <(b) (6)
Subject: RE: B

I was not aware that there would not be travel support and thus I will have to decline attending. I have my funds committed for the rest of the year and cannot apply the travel for other events. Sorry. Thanks for the invitation.

From: (b)(6) Shelly Cox US Courts (b) (6)   ]
Sent: Thursday, Septe   7 7:40 AM
To: Budowle, Bruce <(b) (6)   @unthsc.edu>
Subject: Boston Travel

Good morning. I am sorry that we have not been able to connect by phone. Given your busy schedule, I am sending some information to assist with your travel plans. I think you are aware that the AO does not pay travel expenses for symposium speakers. The advisory committee members are staying at the Hilton Back Bay on Dalton Street The Hilton is currently sold out for the evening

of Oct 26. There will be shuttle service from the Hilton to all group events. There are many hotel options at different price points near the Hilton and/or Boston College. I selected a few to highlight available rates and options.

These hotels are within walking distance to the Hilton so it would be convenient to connect with the shuttle service to/from events.

Sheraton
39 Dalton St / 328 ft from Hilton
(617) 236-2000
$550 standard online rate

The Colonnade
120 Huntington Ave / 0.3 mi from Hilton
(617) 424-7000
$475 standard online rate

Courtyard Marriott
88 Exeter St / 0.5 mi from Hilton
(617) 437-9300
$371 standard online rate

If you prefer something closer to Boston College to facilitate cab or Uber usage, here are recommendations from the school.

Sheraton Needham
100 Cabot St / 3.5 mi from BC
(781) 444-1110
$237 standard online rate

Marriott Newton
2345 Commonwealth Ave / 4.5 mi from BC
(617) 969-1000
$179 standard online rate

Crowne Plaza Newton
320 Washington St / 5.6 mi from BC
(617) 969-3010
$217 standard online rate

I hope this helps narrow the field so you can secure a room and make flight arrangements. I am at your disposal should you have further logistical questions.

Kind regards,
Shelly
(b) (6) (direct)

Good afternoon, Dr. Budowle. Professor Capra at Fordham asked that I reach out to you regarding logistics for the conference at Boston College in October. I am happy to call if you provide a number. Alternatively, my direct dial is (b) (6)   . I look forward to speaking with you. Thank you.

Best,
Shelly

Shelly Cox
Administrative Specialist

UNITED STATES COURTS
Administrative Office
of the United States Courts

Rules Committee Support Office
Administrative Office of the United States Courts
Thurgood Marshall Federal Judiciary Building
One Columbus Circle NE | Room 7-300
Washington, DC 20544
(b) (6) _____ (direct) | (b) (6) _____ (main)
(b) (6) _____

RE: NDCA Conference - Criminal Breakout Session

From: "Cadet, Chinhayi (USACAN)" <(b) (6)   >
To: "Hunt, Ted (ODAG) (JMD)" <(b) (6)
Date: Fri, 09 Mar 2018 12:09:44 -0500

Hi Ted,

Yes, I can give you a call this afternoon. The draft format document is a work in progress, and you and I can make any changes that you would like to that document before the call. I have a meeting that I'm sure will be over by 2:00 (5:00 Eastern time) but might be over by 1:30 (4:30 Eastern time). How late will you be in today and what works best for your chedule today?

Best,
Chinhayi

From Hunt, Ted (ODAG) [mailto (b) (6)
Sent: Friday, March 09, 2018 7:13 AM
To Cadet, Chinhayi (USACAN) (b) (6)
Subject: RE: NDCA Conference - Criminal Breakout Session

Chinhayi,

I'd like to speak with you about the format for the PCAST session before we have our phone conference as a group. Is there ome time that you are available oon?

Ted

From: Cadet, Chinhayi (USACAN) [mailto   (b) (6)]
Sent Thursday, March 8. 2018 6 56 PM
To: Hunt. Ted (ODAG) <   (b) (6) >; Chris Fabricant <(b) (6)   @innocenceproject.org>; Haywood Gilliam
Cc: Ellen Leonida (b) (6)   @fd.org) <(b) (6)   @fd.org>
Subject NDCA Conference Criminal Breakout Session

Hi All,

Thanks so much for agreeing to participate in our Oxford Style debate on the PCAST Report at the NDCA Conference As you know, Ted and Chris will debate, and Judge Gilliam will be moderating. Ellen and I are looking forward to working with you and building a great presentation for Napa The topic of the debate is as follows

Criminal Breakout - An Oxford-Style Debate on the PCAST Report
The President's Council of Advisors on Science and Technology issued a report on Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature Comparison Methods ("PCAST report"). The report identified "gaps" in several forensic science disciplines (including DNA, bite marks, hair comparison), and recommended actions that should be taken to "strengthen the scientific underpinning of the forensic disciplines" and to "promote the more rigorous use of these disciplines in the courtroom." Some forensic scientists, however, think that the PCAST report goes too far. We will have an Oxford style debate on the merits of the PCAST report.

We would like to schedule a call for next week to discuss the debate, logistics, and any ideas you have about the direction of the debate. For discussion, we have attached a draft format for the debate. Please reply to this email with your availability for a conference call, and please let us know if any of the following time slots work for you:
Tuesday, March 13 – 12:00 pm – 5:00 pm (3:00pm - 8:00pm Eastern Time)
Thursday, March 15 –2:00pm- 5:00pm (5:00pm – 8:00 pm Eastern Time)
Friday, March 16 – 10:00am – 1:00pm (1:00pm – 4:00pm Eastern Time)

Let us know if you have any questions. Thanks!

Best,
Chinhayi

RE: NDCA Conference - Criminal Breakout Session

From:	"Cadet, Chinhayi (USACAN)" <█(b) (6)█
To:	Ellen Leonida <█(b) (6)█ @fd.org>, Haywood Gilliam <█(b) (6)█ >
Cc:	Chris Fabricant <█(b) (6)█ @innocenceproject.org>, "Hunt, Ted (ODAG)" <█(b) (6)█
Date	Mon, 12 Mar 2018 20 29 38 0400

Hi All,

Thank you to everyone for responding with your availability for our conference call. Please mark your calendars for Tuesday, March 20 at 11:00 am (2:00 pm Eastern Time) for our call. I will circulate a call-in number tomorrow.

Best,
Chinhayi

From: Ellen Leonida [█(b) (6)█ @fd.org]
Sent: Monday, March 1
To: Haywood Gilliam <█(b) (6)█
Cc: Chris Fabrica___ █(b) (6)█ cenceproject.org>; Cadet, Chinhayi (USACAN) <█(b) (6)█ >; Hunt, Ted (ODAG) (JMD) <█(b) (6)█
Subject: Re: NDCA Conference - Criminal Breakout Session

I also think we can work this out on our call. The "voting" is just a vehicle to make the debate interesting. It's completely anonymous and nobody retains the results. Last year we had people put pre-printed cards ("for" or "against") in envelope   Hopefully thi year we can et it up o people can anonymou ly te t in their vote

3/20 before noon works for me.

Ellen Leonida
Assistant Federal Public Defender
450 Golden Gate Avenue, Room 19 6884
San Francisco. CA 94102
Phone █(b) (6)█

CONFIDENTIALITY NOTICE: This transmission is intended only for the use of the individual or entity to which it is addre ed and may contain information that i privileged and confidential If the reader of thi me age i not the recipient, you are hereby notified that any disclosure, distribution or copying of this information is strictly prohibited. If you have received this transmission in error, please notify us immediately by telephone and return the original document to us at the above address via U.S. Postal Service.

▼ Haywood Gilliam---03/11/2018 08:00:03 PM---From: "Cadet, Chinhayi (USACAN)" <█(b) (6)█

From: Haywood Gilliam/CAND/09/USCOURTS@USCOURTS
To "Hunt, Ted (ODAG)" █(b) (6)█
Cc: "Cadet, Chinhayi (USACAN)" <█(b) (6)█ >, "Chris Fabricant" <█(b) (6)█ @innocenceproject.org>, Ellen Leonida/CANF/09/FDO@fdo
Date: 03/11/2018 08:00 PM
Subject: Re: NDCA Conference - Criminal Breakout Session

Thanks, Ted. I can explain how the polling works when we talk. The aggregate results (what percentage voted yes and what percentage voted no) would be shown to the audience during the session. But no one would know how any individual voted, and the data doesn't get retained. I understand the potential concern, and look forward to discussing on our call.

Sent from my iPhone

On Mar 11, 2018, at 9:51 AM, Hunt, Ted (ODAG) <█████ (b) (6)█ wrote:

All,

I guess I don't understand the purpose of having a voting poll if the results are not to be made public, either to conference attendees or to those outside the meeting   If that is the case, then what is the purpose of collecting this data?

As previously stated, welcome a full debate on this topic, including wide audience participation. However, in light of the timing of the conference, which closely coincides with the Federal Rules Advisory Committee Meeting, I still don't feel that collecting and counting votes that reflect impressions about this topic (for ether public or private use) is appropriate

I look forward to our upcoming call to further discuss the session.

Thanks,

Ted

From Cadet, Chinhayi (USACAN) [mailto (b) (6)   ]
Sent: Saturday, March 10. 2018 6:04 PM
To: Chris Fabricant < (b) (6)   @innocenceproject.org>; (b)(6) Haywood Gilliam US Courts
Cc: Ellen Leonida (b) (6)   @fd.org) < (b) (6)   @fd.org>; Hunt, Ted (ODAG) <(b) (6)   >
Subject RE NDCA Conference   Criminal Breakout Session

I totally agree that an electronic polling option would be great. This seems like a nice compromise position. Ted?

From: Chris Fabricant (b) (6)   @innocenceproject.org]
Sent Saturday, March 10, 2018 1 18 PM
To: (b)(6) Haywood Gilliam US Courts
Cc: Cadet, Chinhayi (USACAN) (b) (6)   Ellen Leonida (b) (6)   @fd.org) < (b) (6)   @fd.org>; Hunt, Ted (ODAG) (JMD) < (b) (6)   >
Subject Re NDCA Conference   Criminal Breakout Session

Greetings all,
I am available 3/20 between 10 and 2 as well. I am traveling 3/21, unfortunately. I like the electronic polling proposal. I want to give some thought to the new, narrower topic, but it's also a little unclear to me what precisely we're to focus on. Is it the efficacy of conducting 'black box' studies, consistent with the PCAST criteria for the appropriate design of such studies?
Thanks,
Chris

On Sat, Mar 10, 2018 at 1:21 PM, <(b)(6) Haywood Gilliam US Courts (b) (6)> wrote:

> Thanks, Chinhayi. I am available on 3/20 between 10 and 2, and on 3/21 between 3 and 4.
>
> One thought: what about the idea of seeing if we can get an electronic polling system, so that we can keep the idea of voting without requiring anyone to take a public position? I've seen those systems used for similar panels at a number of conferences with good results. I think having a vote is a nice way to increa e audience engagement, o perhaps thi would be a way to keep that dimen ion without putting anyone in an awkward position. I have no idea what the cost or logistical hassle would be to get one of those, though, so just some food for thought when we talk. I'll look forward to our call on the 20th or 21st.
>
> Be t,

<image001.jpg>

Haywood S. Gilliam Jr.
United States District Judge
Northern District of California
1301 Clay Street
Oakland, CA 94612

From:	"Cadet, Chinhayi (USACAN)" < (b) (6) >
To:	(b)(6) Haywood Gilliam US Courts " < (b) (6) >
Cc:	Chris Fabricant < (b) (6) @innocenceproject.org>, Ellen Leonida (b) (6) @fd.org)" < (b) (6) @fd.org>, "Hunt, Ted (ODAG) (JMD)" (b) (6)
Date	03/09/2018 04 23 PM
Subject	RE NDCA Conference Criminal Breakout Session

Hi Judge Gilliam, Chris, and Ellen,

<u>**Scheduling**</u>
The following times work for Ted and me   Do any of these time slots work with your schedules?
Tuesday, March 20   6 00 am 8 45 am (9 00 am 11 45 Eastern time) and 10 00am 4 00pm (1 00pm 7 00pm Eastern time)
Wednesday March 21 at 3 00pm 4 00pm (6 00pm 7 00pm Eastern time)

<u>**Format**</u>
Ted pointed out that just a week after our NDCA presentation, the Federal Rules Advisory Committee will meet in Washington, D.C., to discuss whether Federal Rule of Evidence 702 should be amended (or a new rule created) in light of the PCAST Report   Consequently, it may be best to have a standard debate instead of an Oxford style debate   In other words, given the sensitive nature of the topic and timing issues, it may to best to avoid having public votes. Additionally, as the PCAST report makes a number of recommendations, we should discuss how to narrow the focus of which specific issues we will focus on during the debate. Accordingly, a revised draft format is attached, which clarifies the focus of the debate topic and allows for audience participation with questions (and absent voting). Thank you.

Best,
Chinhayi

From: (b)(6) Haywood Gilliam US Courts   (b) (6)   ]
Sent: Friday, March 09, 2018 10:46 AM
To: Cadet, Chinhayi (USACAN) <(b) (6)
Cc: Chris Fabricant <(b) (6)   @innocenceproject.org>; Ellen Leonida (b) (6)   @fd.org) <(b) (6)   @fd.org>;
Hunt, Ted (ODAG) (JMD) (b) (6)
Subject: Re: NDCA Conference - Criminal Breakout Session

Thanks, Chinhayi. I'm traveling for a conference next week, so won't be able to do a call at any of the proposed times (flying on Tuesday and Friday and booked on Thursday). Can we look at some options the following week?

Best,

    image001 jpg

Haywood S. Gilliam Jr.
United States District Judge
Northern <u>Di trict of California</u>
1301 Clay Street
Oakland, CA 94612

From:	"Cadet, Chinhayi (USACAN)" <(b) (6)
To:	"Hunt, Ted (ODAG) (JMD)" <(b) (6) Chris Fabricant <(b) (6) @innocenceproject.org>, Haywood Gilliam
	(b) (6)
Cc	"Ellen Leonida (b) (6) @fd org)" (b) (6) @fd org
Date	03/08/2018 03 56 PM
Subject	NDCA Conference Criminal Breakout Session

Hi All,

Thanks so much for agreeing to participate in our Oxford-Style debate on the PCAST Report at the NDCA Conference. As you know, Ted and Chris will debate, and Judge Gilliam will be moderating. Ellen and I are looking forward to working with you and building a great presentation for Napa. The topic of the debate is as follows:

<u>Criminal Breakout - An Oxford-Style Debate on the PCAST Report</u>
The President's Council of Advisors on Science and Technology issued a report on Forensic Science in Criminal Courts:

Ensuring Scientific Validity of Feature Comparison Methods ("PCAST report"). The report identified "gaps" in several forensic science disciplines (including DNA, bite marks, hair comparison), and recommended actions that should be taken to "strengthen the scientific underpinning of the forensic disciplines" and to "promote the more rigorous use of these disciplines in the courtroom." Some forensic scientists, however, think that the PCAST report goes too far. We will have an Oxford style debate on the merits of the PCAST report.

We would like to schedule a call for next week to discuss the debate, logistics, and any ideas you have about the direction of the debate. For discussion, we have attached a draft format for the debate. Please reply to this email with your availability for a conference call, and please let us know if any of the following time slots work for you:

Tuesday, March 13 – 12:00 pm – 5:00 pm (3:00pm - 8:00pm Eastern Time)
Thursday, March 15 –2:00pm- 5:00pm (5:00pm – 8:00 pm Eastern Time)
Friday, March 16 – 10:00am – 1:00pm (1:00pm – 4:00pm Eastern Time)
Let us know if you have any questions. Thanks!

Best,
Chinhayi
[attachment "Criminal break format.docx" deleted by Haywood Gilliam/CAND/09/USCOURTS] [attachment "Criminal break format revised.docx" deleted by Haywood Gilliam/CAND/09/USCOURTS]

--
M. Chris Fabricant
Director, Strategic Litigation
40 Worth Street, Suite 701
               York 10013
(b) (6)
         >

<u>Criminal break-out session format (3:00 – 4:00 pm):</u>

3:00 – 3:10: moderator introduces the proposition and the debaters. The proposition is that crime labs and criminal courts should adopt the recommendations of the PCAST report, given:

      1) The experimental design (black box study) and criteria that PCAST sets forth by which forensic disciplines should be validated

      2) Method by which error rates are determined

3:10 – 3:30: 10 minutes each for the advocates to set forth their positions

3:30 – 3:34: 2 minutes each for the advocates to respond to the other's opening

3:34 – 3:54: 20 minutes of audience questions (supplemented by questions from the moderator)

3:54 – 4:01: 3 minute closing arguments

RE: NDCA Conference - Criminal Breakout Session

From: "Hunt, Ted (ODAG)" <(b) (6) ▮▮▮▮▮

To: "Cadet, Chinhayi (USACAN)" <(b) (6) ▮▮▮▮▮

Date: Fri, 09 Mar 2018 13:48:34 -0500

I will be working late tonight, so let me know when you will be available later and we will set a time.

Thanks,

Ted

From Cadet, Chinhayi (USACAN) [mailto (b) (6) ▮▮▮▮▮
Sent: Friday, March 9, 2018 12:10 PM
To Hunt, Ted (ODAG) (b) (6) ▮▮▮▮▮
Subject: RE: NDCA Conference - Criminal Breakout Session

Hi Ted,

Yes, I can give you a call this afternoon. The draft format document is a work in progress, and you and I can make any change   that you would like to that document before the call   I have a meeting that I'm   ure will be over by 2 00 (5 00 Eastern time) but might be over by 1:30 (4:30 Eastern time). How late will you be in today and what works best for your schedule today?

Be t,
Chinhayi

From: Hunt, Ted (ODAG) [mailto▮▮▮▮▮ (b) (6)
Sent: Friday, March 09, 2018 7:13 AM
To: Cadet, Chinhayi (USACAN) <▮▮▮▮▮ (b) (6) >
Subject: RE: NDCA Conference - Cr mina Breako   ession

Chinhayi,

I'd like to   peak with you about the format for the PCAST   e   ion before we have our phone conference a   a group   I there some time that you are available soon?

Ted

From: Cadet, Chinhayi (USACAN) [mailto▮▮▮▮▮ (b) (6)
Sent: Thursday, March 8, 2018 6:56 PM
To: Hunt, Ted (ODAG) <▮▮▮▮▮ (b) (6) Haywo ▮▮▮▮▮ >; Chris Fabricant <(b) (6) ▮▮▮▮▮ @innocenceproject.org>; Haywood Gilliam <▮▮▮▮▮
Cc: Ellen Leon da (b) (6) ▮▮▮▮▮ @fd.org) <(b) (6) ▮▮▮▮▮ @fd.org>
Subject: NDCA Conference - Criminal Breakout Session

Re: Ballistics evidence rulings

Yes please. Many thanks!

Sent from my iPhone

On Jun 25, 2018, at 6:50 PM, Hunt, Ted (ODAG) <(b) (6)   wrote:

All –

Happy to help. D.C. recently went through a Daubert challenge to ballistics (based in part on a PCAST-centric challenge) and filed an extensive brief with supporting exhibits.

If you're intere ted, I can pa   tho e material on a a tart

Ted

(b) (5)

(b) (5)

Re: Ballistics evidence rulings

From: "Hur, Robert (USAMD)" <(b) (6)   >
To: "Hunt, Ted (ODAG)" <(b) (6)
Cc: "Medinger, Jason (USAMD)" (b) (6)   "Schenning, Stephen (USAMD)"
<(b) (6)   >, "Harding, Robert (USAMD)" (b) (6)   , "O'Malley, Kristi
(USAMD)" <(b) (6)   >, "Kaplan, Lindsay (USAMD)" <(b) (6)   >, "Gardner,
Daniel (USAMD)" (b) (6)
Date: Mon, 25 Jun 2018 21:16:16 -0400

Yes please! Thanks Ted.

Sent from my iPhone

On Jun 25, 2018, at 6:50 PM, Hunt, Ted (ODAG) <(b) (6)   wrote:

All –

Happy to help D C recently went through a Daubert challenge to balli tic (ba ed in part on a PCAST centric challenge)
and filed an extensive brief with supporting exhibits.

If you're interested, I can pass those materials on as a start.

Ted

RE: Ballistics evidence rulings

From: "Hunt, Ted (ODAG)" <(b) (6)
To: "Hur, Robert (USAMD)" <(b) (6) >
Cc: "Medinger, Jason (USAMD)" <(b) (6) >, "Schenning, Stephen (USAMD)" <(b) (6) >, "Harding, Robert (USAMD)" <(b) (6) >, "O'Malley, Kristi (USAMD)" <(b) (6) >, "Kaplan, Lindsay (USAMD)" <(b) (6) >, "Gardner, Daniel (USAMD)" (b) (6)
Date: Tue, 26 Jun 2018 09:54:27 -0400
Attachments: Firearms Pleading Appendix Compressed pt3of4,pg401-1100.pdf (19.16 MB)

This is 4 of 5.

From: Hur, Robert (USAMD) <(b) (6) >
Sent: Monday, June 25
To: Hunt, Ted (ODAG) (b) (6) >
C (b) (6) >; S MD)
<(b) (6)) <(b) (6) >; O'Malley, Kristi (USAMD)
<(b) (6) >; Kaplan, Lindsay (USAMD) (b) (6) ; Gardner, Daniel (USAMD)
<(b) (6) >
Subject: Re: Ballistics evidence rulings

Yes please! Thanks Ted.

Once the neighborhood of optimal θ has been identified, the process is repeated within this neighborhood using full resolution versions of Z_A and Z_B. The peak correlation value, to be denoted as $ACCF_{max}$, is defined as:

$$ACCF_{max} = \max(\mathrm{corr}(\theta_i)), \forall\, \theta_i = I_{\theta 2}, \tag{9}$$

where the set $I_{\theta 2}$ corresponds to a set of angles within the neighborhood identified in the first step, and their resolution is 1 degree.

We now discuss an example of the registration process used to find the best matching relative position between a pair of images. Figure 8-6 shows the topographies obtained by applying the pre-processing and filtering algorithms to a pair of breech face impressions found on two different cartridge cases fired by the same firearm. The striated marks seen on these breech face impressions seem to be unique to the gun that fired these cartridge cases.

The results of comparing these breech face signatures can be seen in Fig. 8-7. The top image in Fig. 8-6 and the left image in Fig. 8-7 show the breech face for casing #3 fired by Sig Sauer #30 in the original De Kinder collection. The bottom image in Fig. 8-6 and the right image in Fig. 8-7 show the breech face for casing #7 fired by the same gun. The middle image in Fig. 8-7 is a "split screen" image derived from the other two. The right side of this image is the right half of the optimally aligned (rotated and translated) breech face signature of cartridge #7. The two halves match very well, which indicates a high degree of similarity between the two images.

Figure 8-8 shows the corresponding plot of $\mathrm{corr}(\theta_i)$. As seen in this image, the $\mathrm{corr}(\theta_i)$ plot peaks at a relative orientation close to zero degrees, achieving a correlation value $ACCF_{max}$ of 0.65.
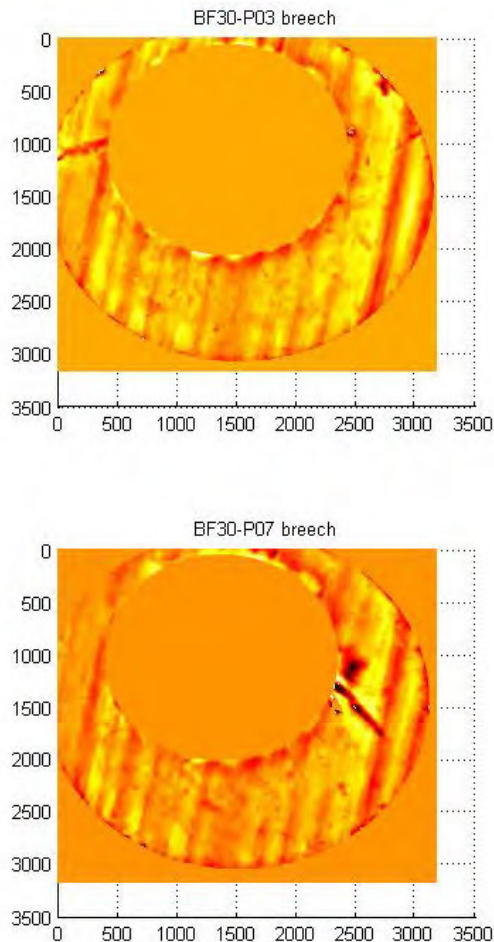
Figure 8-6. Topography images of breech face impressions of two Remington casings fired by the same Sig Sauer #30 gun, one of the original guns studied by De Kinder et al.[7]. The scales represents the pixel positions in the images. The actual diameter is approximately 4 mm.
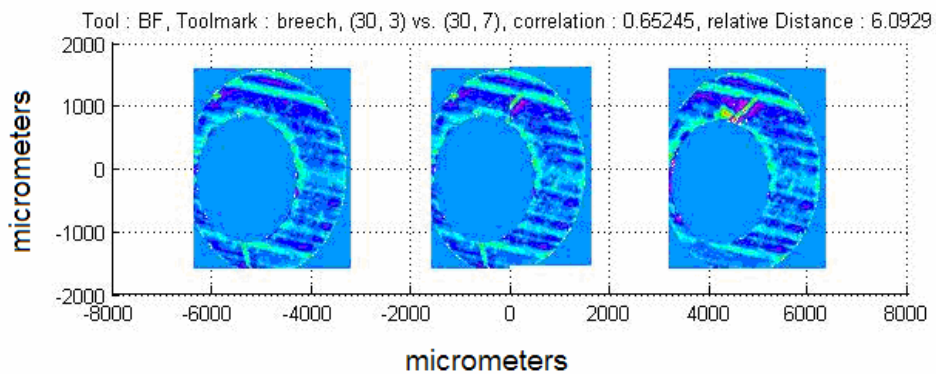

Figure 8-7. Results of alignment and similarity computation for the pair of matching breech face impressions shown in Fig. 8-6. Left−casing 3, right−casing 7, middle−a split screen image comparing the two images. Note: these images have been rotated by about 90° from those of Fig. 8-6.
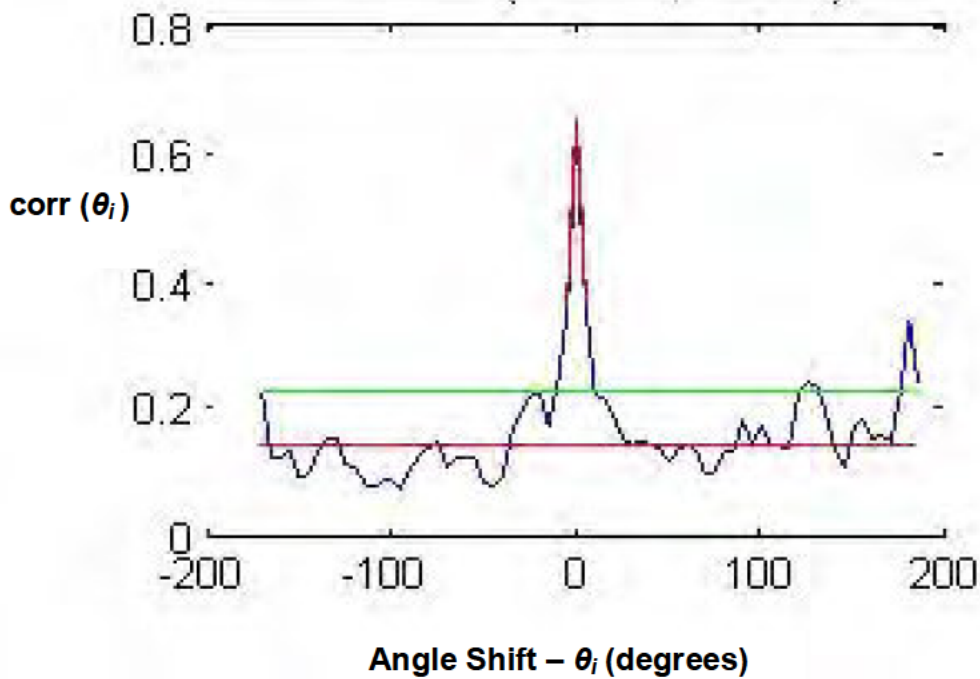
57

Figure 8-8. corr (θ_i) vs. alignment angle shift θ_i for a pair of topography images, showing a clear peak near 0°.

8.5 Consideration of Additional Metrics

In order to improve the ability of the code to produce accurate matches of casings fired by the same gun, we consider other metrics besides $ACCF_{max}$ for assessing similarity. These are described briefly below.

8.5.1 Relative Distance

The correlation value of 0.65, shown in Fig. 8-8, seems to be rather significant because it is "much higher" than the correlation values for all other rotation angles. In order to quantify the degree of significance of this peak, we introduce an additional metric, developed at IAI, which we refer to as the "relative distance." We define the relative distance as follows:

$$\text{relDist}(ACCF_{max}) = \frac{ACCF_{max} - \text{median}(\text{corr}(\theta_i))}{\text{std dev}(\text{corr}(\theta_i))}, \forall\ \theta_i = I_\theta . \tag{10}$$

Having computed the correlation between a pair of cartridge case breech face signatures for all relative orientations of interest, we define the relative distance as the difference (relative to one standard deviation) between the maximum correlation value and the median correlation value for all relative orientation angles under considerations. Shown in Fig. 8-9 is a histogram of the correlation values plotted in Fig. 8-8. From these data, the relative distance is computed to be 6.09, a very high value indicating that the peak correlation value is significantly greater than the median of correlation values obtained for all other rotational angles under consideration.
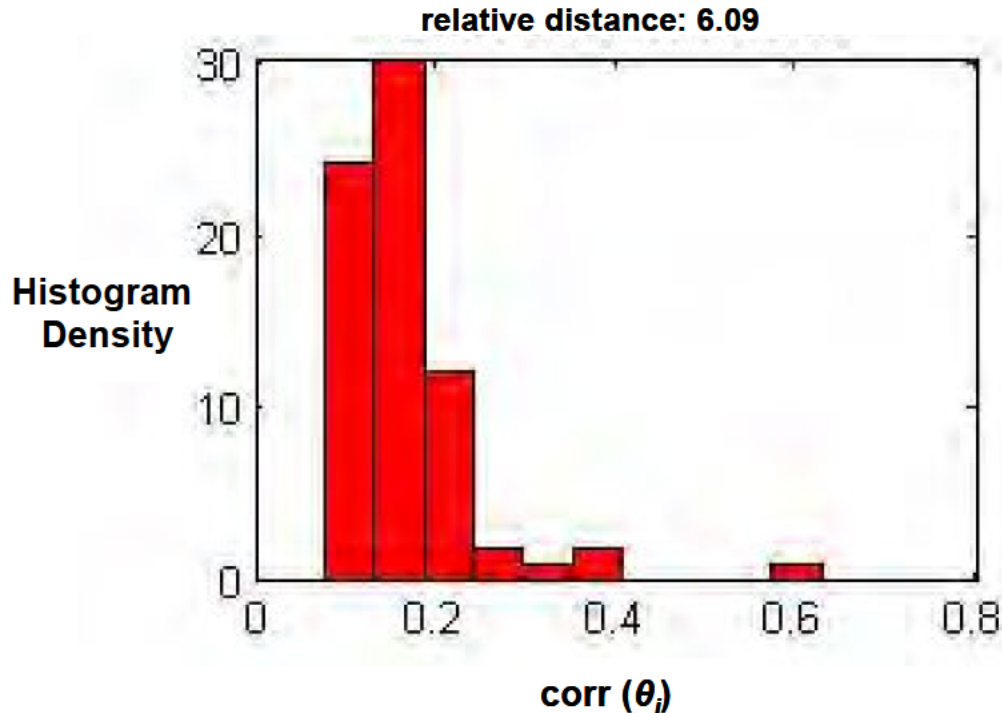
58

20220314-16423

Figure 8-9. Histogram of cross-correlation values for the data shown in Fig. 8-8.

The relative distance can be interpreted in a number of ways. One interpretation is that the relative distance provides a metric of how "out of the ordinary," or significant, a particular relative orientation is with respect to a representative sample of random orientations. Another interpretation is that the median of the correlation values obtained for all rotational angles provides us a "baseline" of the correlation values, which would be achieved by pairs of non-matching breech face signatures. In other words, if we assume that the median of the correlation value for a representative sample of relative orientations is approximately the same for pairs of matching and non-matching cartridge cases (which will be true if the optimal peak for matching pairs is sufficiently narrow), then the relative distance provides a metric of the probability of obtaining the particular maximum correlation value given that the pair under consideration is a non-matching pair. Viewed from this perspective, a relative distance of 6.09 is a very convincing indication of a matching pair.

Figure 8-10 shows a graphical representation of results obtained from a comparison of matching and non-matching signatures of breech face impressions. The peak correlation and relative distance results from the comparison of each pair of such signatures is indicated by either a blue square (for non-matching pairs) or a red rhomboid (for matching pairs). The horizontal axis of Fig. 8-10 corresponds to $ACCF_{max}$ while the vertical axis corresponds to relDist. As expected, the majority of matching breech impression pairs achieve both high maximum correlation and high relative distance values, while the non-matching pairs only achieve relatively low values. Nevertheless, there are a few matching pairs that fail to achieve high values and a few non-matching pairs that achieve relatively high values. These pairs might become false negative identifications and false positive identifications respectively.
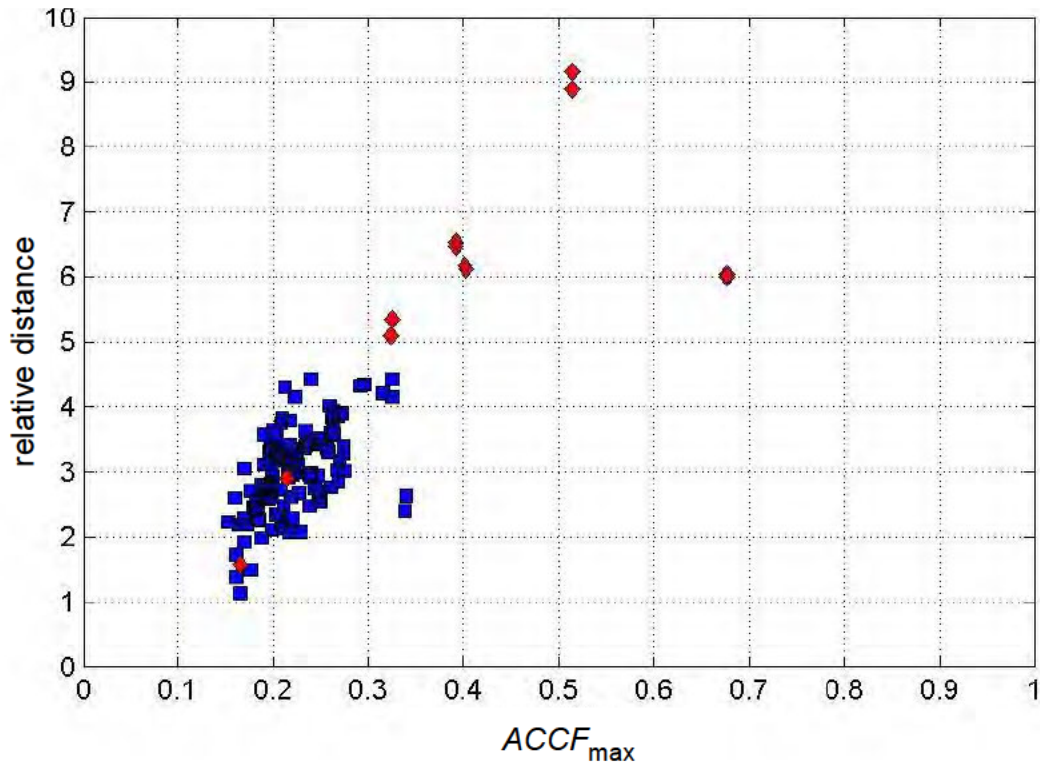
Figure 8-10. Graphical representation of correlation ($ACCF_{max}$) and relative distance results for the comparison of matching (red) and non-matching (blue) pairs of breech face signatures of Remington casings fired from guns 007, 009, 213, 375, 430, and 535 in the De Kinder collection.

The graphical representation shown in Fig. 8-10 lends itself to the definition of non-overlapping regions of the 2D space corresponding to matching, non-matching, and possibly "undetermined" pairs of signatures. One can construct an overall "score" metric based on a linear combination of the maximum correlation and the relative distance, which effectively projects the two dimensional results into a single dimension. The linear transformations used for this study (based on empirical observations) were:

$$s = Gd * \text{relDist} + Gc * (ACCF_{max} - \Delta\text{corr}),$$ (11)

with the empirical constants $Gd = 0.77$, $Gc = 6.4$ and $\Delta\text{corr} = 0.1$ for firing pin impressions, and $Gd = 0.89$, $Gc = 4.5$ and $\Delta\text{corr} = 0.1$ for breech face impressions. The constant Δcorr is the x-intercept in Fig. 8-10. We have used the relative distance and the combined parameter s shown in Eqn. 11 above for one of the correlation calculations shown later in Sec. 8.

8.5.2 Parameters Related to rms Roughness

Although the $ACCF_{max}$ can be used for signature comparison, we observed during previous 2D bullet signature measurements [17] that $ACCF_{max}$ does not characterize the uniqueness of a topography image. Based on the definition of the cross-correlation function, if two compared signatures have the same shape but different vertical scales, their $ACCF_{max}$ is still 100 %. We have, therefore, developed a parameter called the signature difference, D_s, which is highly

60

correlated with $ACCF_{max}$ but which directly quantifies bullet signature differences [26]. For 3D topography comparisons, the 3D version of the parameter D_s is calculated in the following way:

- At the registration position where the $ACCF_{max}$ value between topography B and A occurs, after shifting along the x- and y-directions and rotating around the z-axis, construct a new topography $Z_{B-A}(x)$, which is equal to the difference of the compared topography Z_B and the reference topography Z_A.
- Calculate the areal rms roughness for the new topography $Z_{B-A}(X)$, $Sq(B - A)$.
- Calculate the topography difference D_s between topography B and A defined as the ratio

$$D_s(3D) = Sq^2(B - A) / Sq^2(A). \qquad (12)$$

A similar parameter to D_s is the difference in mean square roughness between the two surfaces being compared. This metric is given by $Sq^2(B) - Sq^2(A)$. Both metrics are directly related to scale differences in roughness between two surfaces that might have otherwise similar shapes. We have calculated these parameters in some of the studies undertaken here but have not yet performed a systematic appraisal of their usefulness as metrics for correlating two surfaces, which would effectively supplement information provided by the cross-correlation maximum.

8.6 Uncertainty Arising from Topography Measurements

As suggested by the NA Committee, uncertainty in the $ACCF_{max}$ results was estimated by repeating topography measurements on two of the casings over four days. The correlations between topography images of the same surfaces and the variations in those correlations enable us to estimate uncertainty in the correlation results due to variation in the topography measurements. The two casings were the Remington casings #3 and #7, both fired by the Sig Sauer 535 gun in the De Kinder collection.

The results for $ACCF_{max}$ correlation values for pairs of these casings are shown in Fig. 8-11 for breech face impressions and Fig. 8-12 for firing pin impressions. The $ACCF_{max}$ values are shown plotted along the x-axis, and the relative distance values are plotted along the y-axis. The results of Fig. 8-11 clearly separate themselves into three groupings: correlations among the breech-face-impression images of Remington casing #3, correlations among the breech-face-impression images of Remington casing #7, and correlations between the images of #3 and #7. The results also include the images from these two casings taken when they were first measured as two of the casings in the set of 70 casings. The grouping of diamonds in Fig. 8-11, for example, includes the pair-wise correlations between the four topography images of the Remington #3 breech face impression measured over the four days and the original image measured about half a year earlier. Altogether there were 20 correlation values calculated over all of the pair-wise comparisons for the five images. Likewise the grouping of squares comprises the 20 pair-wise correlations for images of the Remington #7 casing fired by the Sig Sauer 535. The triangles comprise the intercomparisons of the #3 casing images discussed above with the #7 casing images. There are 50 of these. For all groupings, each pair-wise comparison is included twice, with the reference and compared casings switching places.
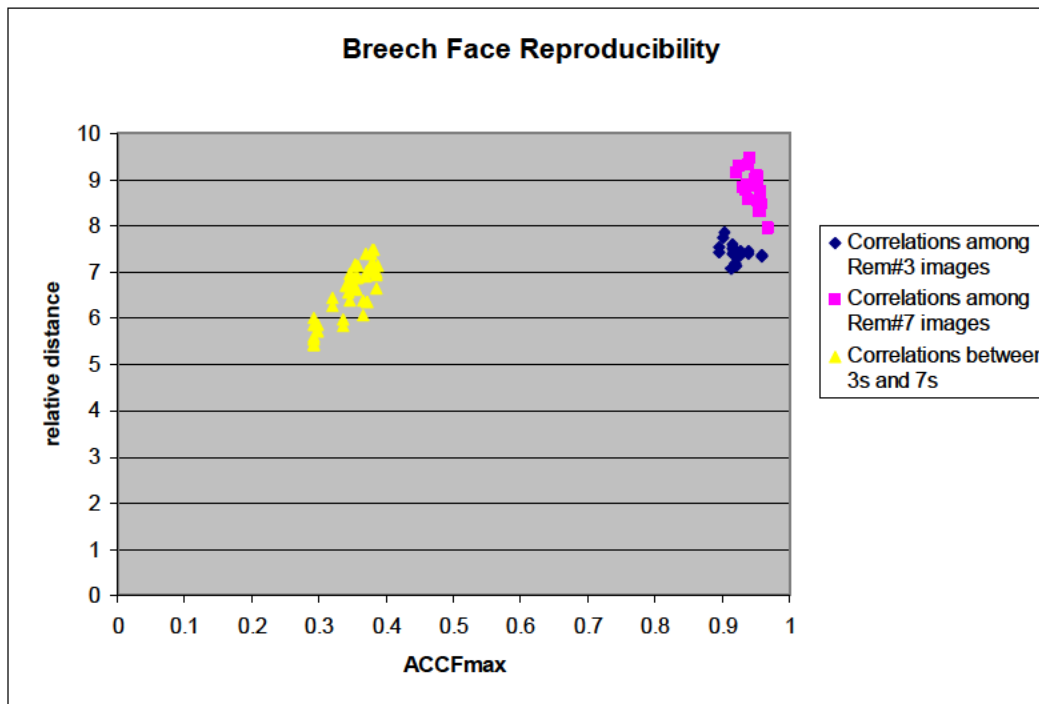
61

Breech Face Reproducibility

Figure 8-11. Reproducibility of topography measurements of Remington #3 and #7 breech face impressions as given by the $ACCF_{max}$ and the relative distance parameters.
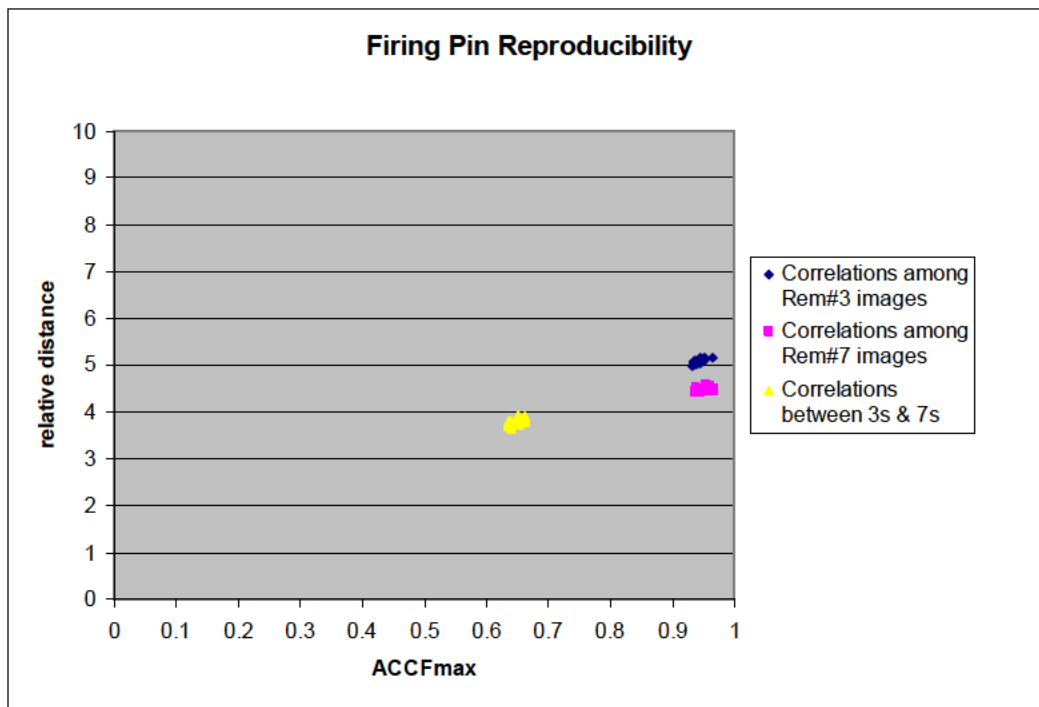
Firing Pin Reproducibility

Figure 8-12. Reproducibility of topography measurements of Remington #3 and #7 firing pin impressions as given by the $ACCF_{max}$ and the relative distance parameters.

Reproducibility of the firing pin impressions is shown in Fig. 8-12. There are a smaller number of data points here because the original data for the firing pin impressions were measured on a different model of the confocal microscope with the camera having a different pixel spacing along the *y*-axis. Therefore we did not include those two images in this comparison. The clusters of diamonds and squares represent twelve correlations among four images each and the cluster of triangles represent 32 correlations between the two sets of four images.

We now discuss estimated uncertainties in the $ACCF_{max}$ values arising from sources of error in the topographic measurements. We expect that uncertainties in the relative distance parameter are based on similar considerations but do not discuss those here. The mean $ACCF_{max}$ values for these groupings and their standard deviations are shown in Table 8-1:

Table 8-1. Average correlation values and standard deviations for topography measurements of De Kinder Remington casings fired from Sig Sauer 535.

Components Being Compared	Breech Face Impressions $ACCF_{max} \pm 1$ std. dev.	Firing Pin Impressions $ACCF_{max} \pm 1$ std. dev.
Rem #3, Sig Sauer 535	0.920 ± 0.018	0.944 ± 0.010
Rem #7, Sig Sauer 535	0.947 ± 0.013	0.954 ± 0.011
Rem #7 vs.Rem#3	0.349 ± 0.030	0.647 ± 0.010

The results clearly indicate a high similarity between different topography measurements of the same casing but significant differences between measurements of the two different casings, even though those casings were fired by the same gun. The correlation values of 0.920, 0.944, 0.947, and 0.954 are both high and consistent with one another and have small standard deviations. However, they are slightly less than unity. This suggests that there are variations between topography images of the same object giving rise to an attenuation of the $ACCF_{max}$ value by 5 % to 8%. This variation arises from variation in the measured surface topography along all three coordinate axes, *x* or *y* or *z*. The fact that the standard deviations are so low suggests that the variation between images is consistent and random, probably with high spatial frequency components. By contrast the low scores, averaging 0.349 when comparing the two breech faces and 0.647 for the two firing pins, indicate large differences between the #3 and #7 surfaces, even though the casings were fired by the same gun.

We, therefore, conclude that the correlation value is biased below unity by about 5 % to 8 % by measurement-related differences and noise in the topography images, but the statistical uncertainty of this bias is only about 1.8 % (standard uncertainty). This source of uncertainty is much smaller than the changes in $ACCF_{max}$ due to the topography of the surface, which are the principal sources of the differences we aim to observe.

8.6.1 Uncertainty Budget for Breech Face Impressions

For the breech face impressions, the trimming process leads to a second source of variation and bias. Figure 8-13 shows the topography of the breech face impressions of Remington casings #3 and #7 from Sig Sauer #007. It reveals a prominent ridge around the firing pin impression. If such inner ridges are trimmed out of the breech face impressions to be matched, then the average correlation value is 0.349 for gun #535 as shown in Table 8-1. If, however, the inner ridges are included in the breech face impressions, the correlation score is more influenced by the prominent ridge, and the average score for the topography comparisons of the #3 casing with the

63

#7 casing increases by about 0.06 to 0.402 with a variation that is several times larger than the 0.01 levels we record in Table 8-1 above.
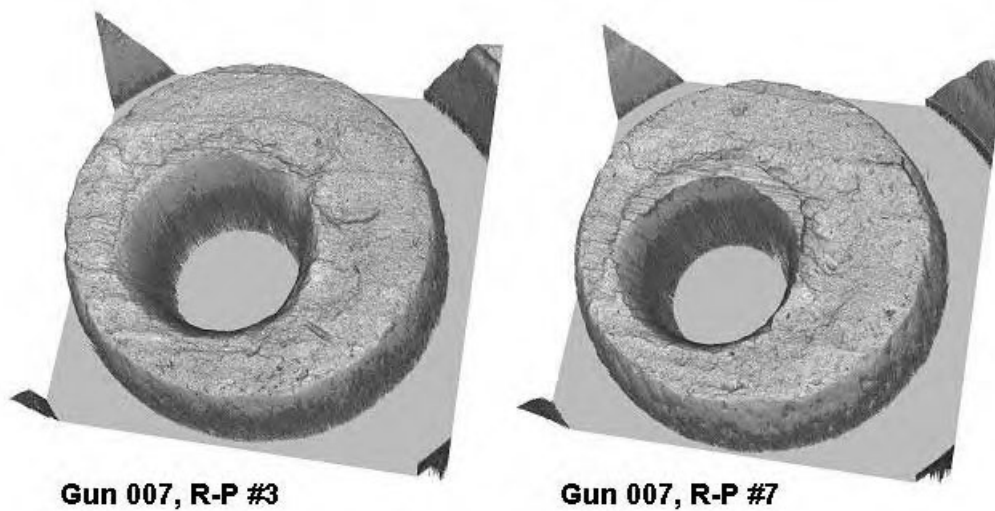

Figure 8-13. Topography images of the breech face impressions of Remington casings #3 and #7 fired in the Sig Sauer gun 007 showing the raised section around the inner radius.

As a result, one might include another factor of 0.06 as an added component of uncertainty for $ACCF_{max}$. However, including the inner ridge or other form deviations in the breech face topography is not regarded as good practice because this form component obscures the real individual characteristics of the surfaces present in the fine roughness structure. A good area-selection procedure should not include the ridge. The trimming procedure leading to Fig. 8-11 and the entries for breech face impressions in Table 8-1 correctly excluded the ridge from the topography images that were correlated. The average difference of 0.06 is therefore regarded as an upper limit of the variation in $ACCF_{max}$ that may arise from variation in the selection of areas to correlate for the breech face impression. This is a limitation of current algorithms. Further development could lead to automated algorithms for area selection that are more accurate.

Because the $ACCF_{max}$ cross-correlation parameter depends on relative differences between images, we believe that there are very few other significant sources of uncertainty arising from the measurement process, as long as a systematic protocol for measurement and the areas to be measured have been determined. If instrument-related errors occur in one of the images being compared, those errors produce a change in the $ACCF_{max}$ value. These types of changes should be captured by the Type A [46,47] statistical uncertainty discussed above.

The uncertainty budget for $ACCF_{max}$, for the breech face impressions therefore contains two components. The first (u_1) is the statistical type A uncertainty arising from instrument variations. We estimate this to be 0.018 ($k = 1$) by using the largest of the standard deviations shown in Table 8-1. The second source of uncertainty (u_2) arises from selection and variation of the measured areas. Using the value of 0.06 discussed above as an outer limit of error and assigning a uniform probability distribution to the error, we derive a Type B uncertainty [46,47] for u_2 equal to 0.034 ($k = 1$). Combining these two components quadratically, we arrive at a combined

64

standard uncertainty u_c ($k = 1$) for $ACCF_{max}$ equal to 0.038. In addition, we expect the $ACCF_{max}$ value to be consistently biased low by 0.065. The biases of 0.05 to 0.08 recorded in connection with Table 8-1 are consistent with that estimate and with the Type A uncertainty component discussed above.

8.6.2 Uncertainty Sources for Firing Pin Impressions

The data in Fig. 8-12 suggest that the uncertainties for firing pin impressions are smaller than the uncertainties for breech face impressions. However, correlation of the firing pin impressions is more susceptible to two sources of uncertainty, relocation error of the measured area and dropouts and outliers, than correlation of the breech face impressions. The measured area of the firing pin impressions is 800 μm × 800 μm, several times smaller than the 4 mm × 4 mm area of the breech face impressions. Modest changes in the measurement location between two matching images can produce relatively large offsets between the compared areas, thus reducing the overlap of matching features and leading to a decrease of the $ACCF_{max}$ value. Relocation error is especially likely when the firing pin impression is wide and flat, as seems to be the case for the Rem image in Fig. 6-1. In addition, because a large part of the field of view of the firing pin impression is on steeply sloped walls of the impression, outliers are more difficult to distinguish and eliminate while preserving the good data. Due to these effects, the uncertainty for calculating the $ACCF_{max}$ metric for firing pin impressions may be larger than the uncertainty for breech face impressions. In future work, we intend to estimate these effects by comparing matched pairs of images independently measured on each of the 70 De Kinder firing pin impressions.

8.7   Some Cross-Correlation Results

We now examine several cases for pair-wise correlation of the De Kinder firing pin impressions.

First, the same gun with different ammos can produce similar firing pin impression topographies producing a high $ACCF_{max}$ value. Figure 8-14 shows an example of a pair of casings fired from the same gun, Sig Sauer 007. The reference topography is obtained from a Remington casing and the compared topography from a Speer casing. The $ACCF_{max}$ value is about 95 %. The Sq values of the two topographies are quite similar, 1.43 μm and 1.47 μm. The topography difference D_s is 6.9 %.

The same gun with different ammos can also produce firing pin impressions with different forms yielding a low $ACCF_{max}$ value. Figure 8-15 shows an example from the Sig Sauer 139 gun. The reference topography was obtained from a CCI casing and the compared topography from a Federal casing. The $ACCF_{max}$ is about 46 %. Their Sq values are quite different, 0.46 μm and 0.86 μm, respectively. The topography difference D_s is large, about 222 %.

By contrast, different guns with different ammos can produce firing pin impressions with similar topographies leading to a high $ACCF_{max}$ value. Figure 8-16 shows the topographies of the firing pin impressions from gun 314 with CCI ammo and gun 215 with Winchester ammo. The $ACCF_{max}$ value is about 82 %. However, their Sq values are 1.31 μm and 2.55 μm, respectively, and the topography difference D_s is about 164 %. From this observation, it is clear that a high

$ACCF_{max}$ value does not necessarily confirm a match. The D_s parameter provides supplementary information that can improve the accuracy of proposed matches.
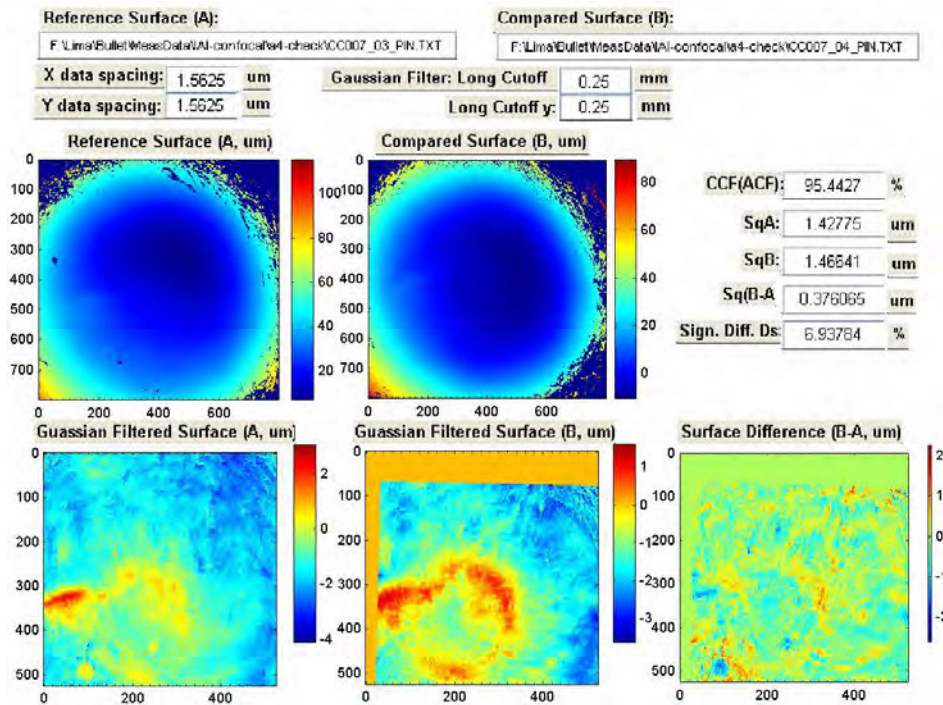

Figure 8-14. Comparison of the topography images of the firing pin impressions from a Remington casing (reference surface) and a Speer casing (compared surface) fired by the same Sig Sauer 007 gun. The value of $ACCF_{max}$ is approximately 95 %. All axes have units of μm.
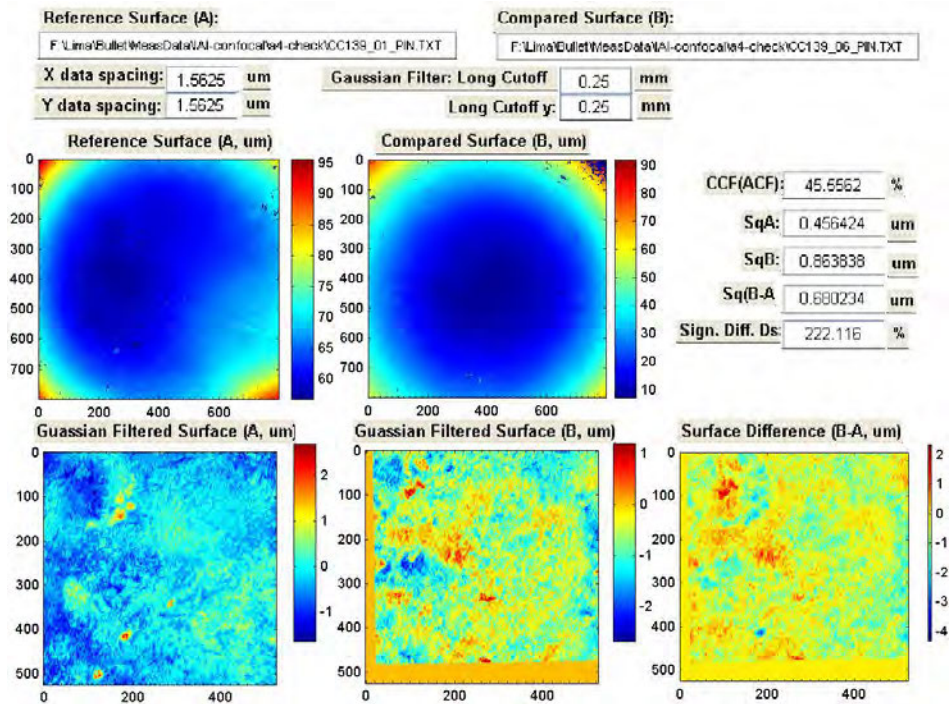
Figure 8-15. Comparison of the topography images of the firing pin impressions from a CCI casing (reference) and a Federal casing (compared) fired by the same Sig Sauer 139 gun. The value of $ACCF_{max}$ is approximately 46 %, indicating that the same gun with different ammos can also produce firing pin impressions with different forms. All axes have units of μm.
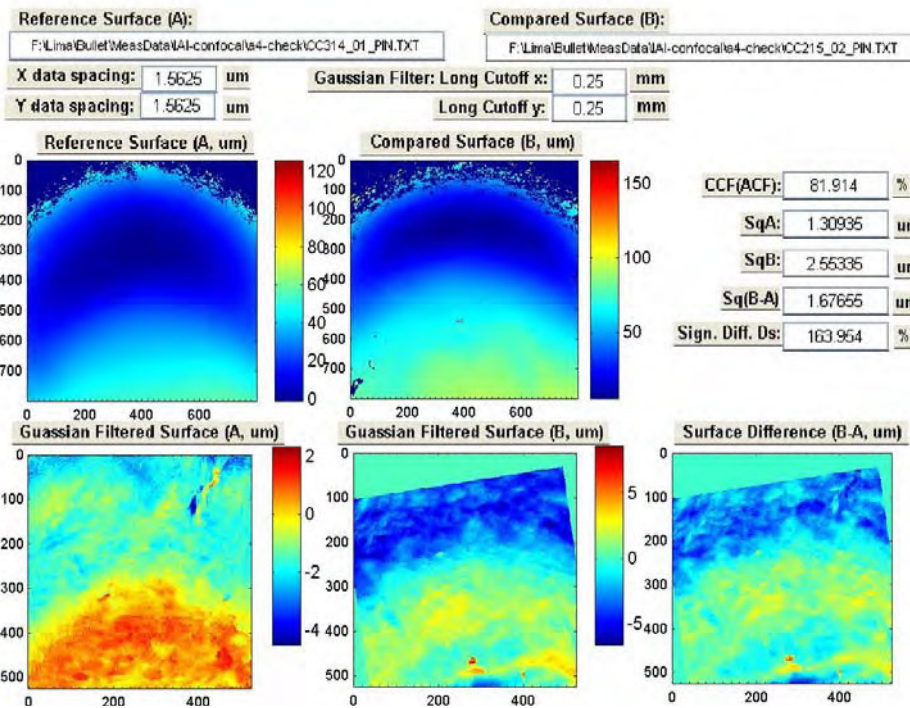

Figure 8-16. Topography images of firing pin impressions from Sig Sauer 314 gun with CCI ammo and Sig Sauer 215 gun with Winchester ammo.

9.   Statistical Analysis: General

9.1   Introduction

This section analyzes the available data using several methods in order to answer the following questions:

1. Are the firing pin and breech face impressions left by individual guns on casings distinct and repeatable enough to distinguish them from those of other guns?
2. How does a 3D topography technology, such as N-3D, perform relative to an I-2D system?
3. What factors (e.g. gun manufacturer, ammunition type, etc.) affect the performance of the systems?
4. Which are more helpful–firing pin or breech face data? Does the answer depend on the imaging system or the set of casings used?

The section begins with a descriptive recap of the experimental data obtained in this study. Then, using these data sets as test databases, the N-3D and I-2D imaging systems are tested using Top Ten searches of the type routinely output by the I-2D system. These tests give a clear picture of the performances of the systems over the various data sets.

Later sub-sections analyze the full round-robin correlation data obtained by the N-3D system. A probabilistic overlap metric *p* is introduced as a useful heuristic for comparing the empirical distributions of correlation scores. This metric is later used as an input for probability models that project how a system performs for very large databases.

9.2   Recap of Casings Databases

This study investigated two sets of casing data: "De Kinder" and "NBIDE". The De Kinder data are "historical," with the casings produced by De Kinder et al. [7] several years ago. NIST created the NBIDE database in 2005 as a part of this ballistics study.

9.2.1 De Kinder

What is referred to as the De Kinder casing set is actually a subset of a much larger experiment discussed in the paper by De Kinder et al. [7]. The casings described here involved 10 guns all of the same model (Sig Sauer 9 mm Model P226) and 7 ammunitions (cartridge types) of which two are repeats of the same brand (Remington):

1.   CCI
2.   Winchester
3.   Remington (Rem)
4.   Speer
5.   Wolf
6.   Federal
7.   Rem (a repeat)

A total of $10 \times 7 = 70$ test firings were used. The corresponding 70 casings were originally studied via the I-2D imaging method, and a report on the findings appears in Ref. 7. As part of the NIST study, these 70 De Kinder casings were also imaged for topography (3D) and analyzed via the N-3D method.

9.2.2 NBIDE

A central component in the NIST study was to determine the effect of gun type (manufacturer) on gun identifiability. As noted above, the De Kinder study was limited to a single gun type (Sig Sauer 9 mm). To go beyond this, a statistically designed experiment was developed as part of the NBIDE. The specifications for this experiment were as follows:

1. Number of gun types: 3 – Smith&Wesson, Ruger, Sig Sauer P226
2. Number of guns (total): 12
3. Number of ammo types: 4 – Remington, Winchester, PMC, Speer (extra)
4. Number of days: 3 – 48 firings per day
5. Total number of test firings: 144 = 12 guns × 4 ammos × 3 days
6. Total number of casings subsequently analyzed: 108 – only 3 ammos

The experiment was designed and conducted in accordance with rigorous statistical design principles and techniques. Within a given day, there were four sets of 12 gun firings. Across the three days, there were a total of 12 sets of 12 gun firings. In a latin square [48] fashion, for each time position (1 to 12) within a set, each gun was fired once and only once across the 12 sets. Each set of 12 firings used the same ammo. All four ammo types were used every day. The ordering of the ammos for the four sets within a day was balanced in an incomplete latin square fashion. Finally, after the 144 fired casings were collected and annotated, the casings were re-randomized for the first part of the analysis so as to assure that the analysis was done in a double blind fashion [21].

The advantage of the resulting NBIDE database is that it allows one to ascertain the existence and magnitude of gun type and ammo type on gun distinguishability. Because a given gun-ammo combination occurs three times across the experiment, this NBIDE database also allows for the assessment of how distinguishable or indistinguishable a given gun can be across the three firings. The limitation of the NBIDE database is that the conclusions are, strictly speaking, limited in scope to the three gun types utilized and the three ammo types. Also, the total number of firings (144) and analyzed casings (108) is still relatively small compared to the large sizes envisioned for a national ballistics database.

9.3   System Performance Analysis

When we compare a single reference casing to an existing database of casings, how do we determine if the comparison was a "success"? What is the criterion for "success"? In practice, an imaging/analysis system will yield a short (e.g., ten item) list of best-matching casings, which will then be subject to further examination by a human forensics expert. Thus for this single trial, the performance of an imaging/analysis system might be declared a "success" if the correct database casing, if existent, appears in the short list produced by the imaging/analysis system,

since frequently that list is what the human expert will limit his/her search to. The system will have "failed" if a matching database casing exists and the "TopTen" list does not include that casing.

The analysis of the performance of a system will thus be centered on the questions: Was it a "success" (for a given reference casing trial)? What percentage of the time was it a success (for a set of reference casings)?

For the two data sets at hand (De Kinder with 70 test firings and NBIDE with 108 test firings), we have many opportunities to achieve success. For the I-2D system, a Top Ten list is produced automatically via proprietary software. In the present study, we used the topography data, described in earlier sections, estimated the similarity of pairs of casings based on the cross-correlation function maximum ($ACCF_{max}$) described in Section 8, and formed Top Ten lists from the relative rankings.

We can use the 70 De Kinder test firings as an opportunity for 70 reference-casing trials. For each such reference casing, we use the remaining 69 test casings as our "database". Thus the casing from the first test firing was compared against 70 - 1 = 69 other casings from the remaining test firings. Out of those 69 casings, six came from the same gun as the reference firing, and 63 (= 9 guns × 7 ammos) came from other guns. Thus in an ideal world we would expect the system's Top Ten list to contain all six of those remaining casings. A less stringent criterion would expect five out of the six casings to appear in the Top Ten list, and so forth. A very weak criterion would expect at least one out of the six casings in the Top Ten list. A complete failure would yield none out of the six casings in the Top Ten list. Thus for a fixed inclusion criterion (6, 5, ..., 1), the imaging/analysis system will be judged as a "success" or a "failure".

In a similar fashion, we choose the casing from the second test firing and compare it to the 69 casings from the remaining firings. Again, for a fixed inclusion criterion, the imaging/analysis systems may be declared a success or a failure.

Repeating the process for all 70 casings yields a sequence of 70 successes or failures. The proportion of those 70 cases that were a success defines the performance of the imaging/analysis system for this particular casing database and a fixed inclusion (6,5,...,1) criterion.

For the NBIDE database, we have 12 guns, three ammos, and three days (repeats), and so the same sort of process would yield a comparison of the casing from the first test firing against the 108 - 1 = 107 casings from the remaining firings. Of those 107 casings, 8 (= 3 ammos × 3 days - 1) come from the same gun as the first test firing, and 99 (= 11 guns × 3 ammos × 3 days) come from different guns. Again in an ideal world, an excellent-performing imaging/analysis system would have all eight of those same-gun casings in the Top Ten list, a good performing system might have seven out of the eight in the Top Ten, and so forth, down to a weak performing system having only one out of the eight, and a poor-performing system having none out of the eight. Thus for a given reference casing and a fixed inclusion criterion (8, 7, ..., 1), the system may be judged as "successful" or "failing". Repeating the process for all 108 test-fire reference

70

casings will yield the percentage of the time the imaging/analysis system performed well; this will be our performance metric.

9.3.1 De Kinder Top Ten System Performance Analysis

In light of the above, the following listing shows the performance metric for the I-2D and N-3D imaging/analysis systems for the

Database = De Kinder (70 casings: 10 guns × 7 ammos)
Region = Firing pin impressions only
Total number of items in the system output list = 10 (i.e., Top Ten)
"Success" criterion: at least 1 of the 7-1 = 6 casings appears in the Top Ten
(a very weak criterion)

	$i \geq 1$	
	I2	N3
FP (%):	94	74

where I-2D is shortened to I2 here and N-3D is shortened to N3 and the casing region is designated by "FP" for firing pin.

This listing says that for the firing pin region of the De Kinder data, 94 % (= 66) of the 70 I-2D Top Ten lists were "successful"–containing one or more of the remaining six correct casings. It further says that 74 % (= 52) of the 70 N-3D "Top Ten" lists were "successful"—containing one or more of the remaining six correct casings. Thus for this particular case and (very weak) criterion, the I-2D system performed better.

The expanded listing, which contains stronger criteria, is as follows:

	$i \geq 1$		$i \geq 2$		$i \geq 3$		$i \geq 4$		$i \geq 5$		$i \geq 6$	
	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3
FP (%):	94	74	83	63	66	57	39	51	17	47	7	33

This listing shows that for the firing pin region of the De Kinder data, and for both imaging/analysis methods, as the inclusion criterion becomes more stringent, the success proportion decreases. For example, only 7 % (= 5) of the 70 I-2D Top Ten lists were fully "successful"—containing all six of the remaining six correct casings, and only 33 % (= 23) of the 70 N-3D Top Ten lists were "successful"—containing all six of the remaining six correct casings. Thus for this particular case and very strong criterion, the two imaging/analysis systems performed poorly, but with N-3D performing better than I-2D. Note that for the less stringent criteria, I-2D performs better than N-3D, but as the stringency increases, N-3D performs increasingly better relative to I-2D. The reason may be because N-3D performed very well on some guns and very poorly on others, while the I-2D performance tended to be between those two extremes.

416

The above analysis was for the firing pin region only. A similar analysis could be done for the breech face (BF) region only. Doing so yields the following augmented listing:

De Kinder/ Top Ten

	$i \geq 1$		$i \geq 2$		$i \geq 3$		$i \geq 4$		$i \geq 5$		$i \geq 6$	
	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3
FP (%):	94	74	83	63	66	57	39	51	17	47	7	33
BF (%):	67	90	30	74	4	57	0	37	0	19	0	6

This listing shows that for the breech face region of the De Kinder data, as the inclusion criterion becomes more stringent, the success proportion decreases for I-2D from 67 % to 0 % and for N-3D from 90 % to 6 %. For the most stringent criterion requiring all six of the six matching casings to be in the Top Ten list, none of the 70 I-2D Top Ten lists were "successful", and only 6 % (= 4) of the 70 N-3D Top Ten lists were "successful." Thus for this particular case and very strong criterion, the I-2D system failed and the N-3D system performed only slightly better.

General De Kinder/Top Ten findings (valid across both Firing Pin and Breech Face) that may be drawn from the above listing are as follows:

1.   Success proportions for both I-2D and N-3D decrease as stringency increases, a mathematical necessity.
2.   The most drastic decrease is I-2D/Breech Face.
3.   The least drastic decrease is N-3D/Firing Pin.
4.   I-2D does not do well for Breech Face.
5.   For Firing Pin, I-2D performs better than N-3D in the three least stringent cases, but N-3D performs better than I-2D in the three most stringent cases.
6.   For Breech Face, N-3D performs better than I-2D in all 6 out of 6 cases.
7.   For I-2D, Firing Pin is a better discriminator than Breech Face. For N-3D, Firing Pin is a better discriminator for the most stringent cases, but not for the least stringent cases.

9.3.2 De Kinder: Table of Top Ten Results

As a reference, the following table of Top Ten testing results gives some more information on how the systems performed with respect to the area imaged and the technology used.

Table 9-1. Comparison of the numbers of *correct* matches (out of the ten highest *scoring* matches) for the I-2D correlation metric and the N-3D $ACCF_{max}$ metric applied to firing pin and breech face impressions for 10 De Kinder guns . Rem sep denotes the second Remington casing filed separately.

Number of Correct Matches (Maximum 6)

Ref. Casing	Ammo	Firing Pin		Breech Face	
		I-2D	N-3D	I-2D	N-3D
Means		3.06	3.26	1.01	2.83
007-01	CCI	4	4	0	2
007-02	WIN	5	6	0	1
007-03	Rem	6	6	2	2
007-04	SPEER	6	6	1	0
007-05	WOLF	6	5	1	1
007-06	FC	6	5	2	1
007-07	Rem sep	6	6	1	4
009-01	CCI	3	5	0	2
009-02	WIN	0	1	0	0
009-03	Rem	2	5	0	4
009-04	SPEER	4	5	1	0
009-05	WOLF	4	5	0	1
009-06	FC	4	5	1	3
009-07	Rem sep	4	5	2	3
117-01	CCI	3	1	1	5
117-02	WIN	0	3	0	1
117-03	Rem	4	2	2	5
117-04	SPEER	5	4	1	3
117-05	WOLF	5	4	1	4
117-06	FC	5	3	0	4
117-07	Rem sep	3	0	1	6
139-01	CCI	3	1	0	4
139-02	WIN	3	0	1	1
139-03	Rem	4	0	2	3
139-04	SPEER	3	0	2	3
139-05	WOLF	2	0	0	2
139-06	FC	1	0	1	5
139-07	Rem sep	3	0	3	5
213-01	CCI	1	0	2	4
213-02	WIN	4	0	0	0
213-03	Rem	4	1	2	5
213-04	SPEER	3	0	0	3
213-05	WOLF	3	0	1	3

73

418

213-06	FC	1	0	0	1
213-07	Rem sep	0	0	2	5
215-01	CCI	1	3	1	6
215-02	WIN	2	0	0	4
215-03	Rem	3	0	1	6
215-04	SPEER	1	3	0	4
215-05	WOLF	1	1	2	3
215-06	FC	2	2	0	3
215-07	Rem sep	2	1	1	5
314-01	CCI	4	1	2	4
314-02	WIN	2	0	0	0
314-03	Rem	5	0	1	4
314-04	SPEER	2	2	2	2
314-05	WOLF	3	1	1	1
314-06	FC	3	2	1	0
314-07	Rem sep	3	0	1	4
375-01	CCI	3	6	3	1
375-02	WIN	3	6	0	2
375-03	Rem	4	6	2	2
375-04	SPEER	5	6	2	0
375-05	WOLF	4	6	1	2
375-06	FC	4	6	2	1
375-07	Rem sep	4	6	2	4
430-01	CCI	2	5	1	5
430-02	WIN	0	6	0	2
430-03	Rem	3	6	2	6
430-04	SPEER	3	6	1	1
430-05	WOLF	3	6	1	4
430-06	FC	2	6	1	3
430-07	Rem sep	5	6	1	5
535-01	CCI	1	5	0	3
535-02	WIN	4	6	1	2
535-03	Rem	1	6	3	3
535-04	SPEER	3	6	0	3
535-05	WOLF	2	6	0	2
535-06	FC	2	6	0	2
535-07	Rem sep	2	6	2	3

The table shows large performance differences between individual guns.

System Gun Differences: Firing Pin

For further analysis of the performance differences, we define for the firing pins:

Y = (# of N-3D correct matches) – (# of I-2D correct matches).

We find that: Median (Y) = 0, Mean(Y) = 0.2, Standard Deviation (Y) = 2.5.

This implies a slight advantage for N-3D. However, one wonders if that holds across the guns or whether it depends on the particular gun. Figure 9-1 charts the difference variable Y by reference gun.
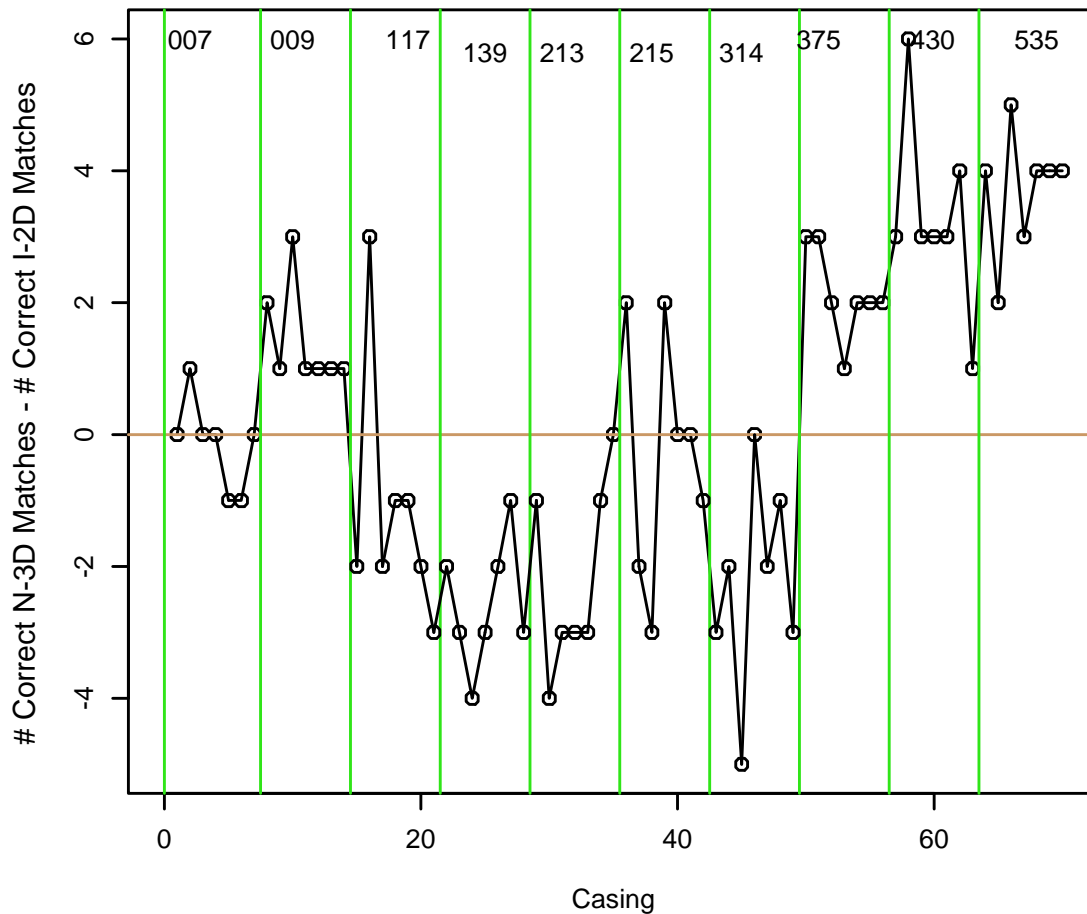

Figure 9-1. Graph of Y = (# of correct N-3D matches) – (# of correct I-2D matches) of De Kinder firing pin impressions arranged by reference gun. The points above the zero-line depict those casings where N-3D performs better than I-2D, while the points below the zero-line depict those where I-2D does better.

The performance differences are very gun-dependent. It is clear that the 3D and 2D methods are not making the same mistakes; the 3D method does better for guns 009, 375, 430, and 535, while the 2D method does better on 139, 213, and 314. Further inspection of the actual images for these casings may give clues as to what characteristics drive the differential performance of the two methodologies. A detailed analysis of differences between the guns is given in Sec. 10.

75

System Gun Differences: Breech Face

For further analysis of the performance differences, we define for the breech faces:

Y = (# of N-3D correct matches) – (# of I-2D correct matches). The summary statistics for Y are:

Median (Y) = 2.0, Mean (Y) = 1.9, Standard Deviation (Y) = 1.8.

This implies a clear advantage for the 3D method. Again, one wonders if that holds across the guns or whether it depends on the particular gun. Figure 9-2 charts the difference variable Y by reference gun.
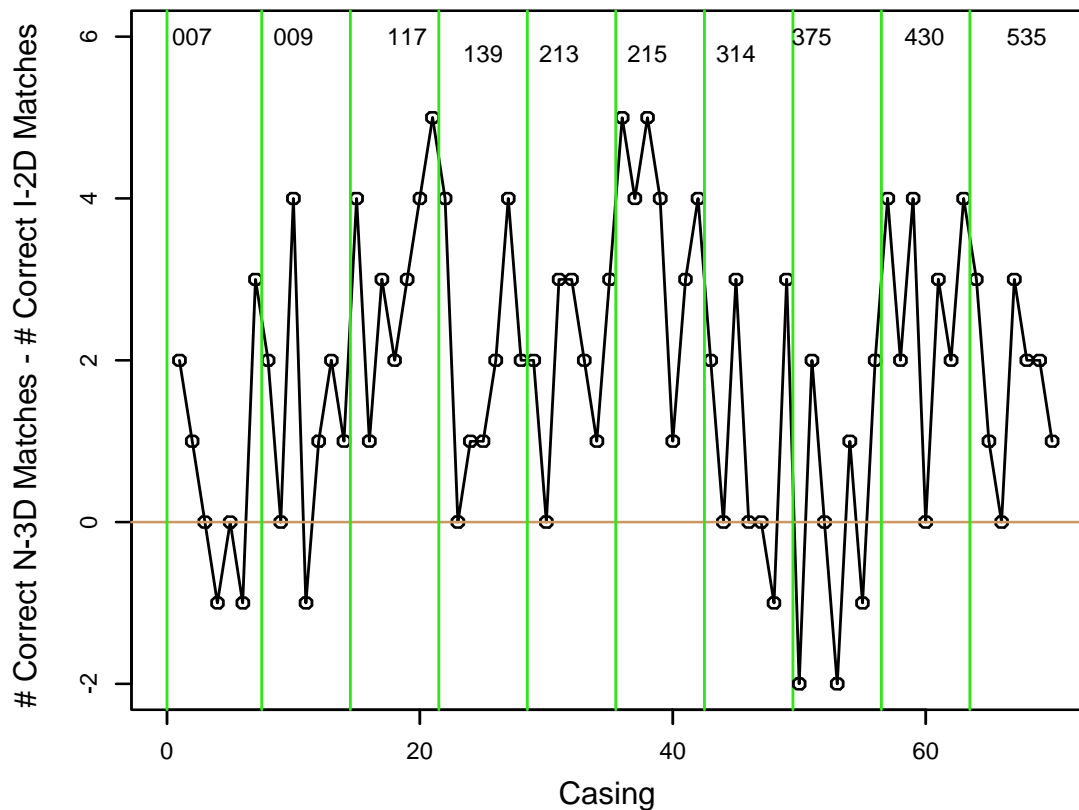

Figure 9-2. Graph of Y = (# of correct N-3D matches) – (# of correct I-2D matches) of De Kinder breech face impressions arranged by reference gun. The points above the zero-line depict those casings where N-3D performs better than I-2D, while the points below the zero-line depict those where I-2D does better.

The 3D method is doing somewhat better than the 2D method across the board except for gun 375. Again, a detailed analysis of the differences between the guns is given in Section 10.

9.3.3 NBIDE: Top Ten System Performance Analysis

In addition to the De Kinder data, the NBIDE study also created and analyzed its own casings database (NBIDE). Whereas the De Kinder data set had one gun type (Sig Sauer), ten guns, and seven ammo types (including a replicated Remington ammo) for a total of 70 (= 10 × 7) test firings, the NBIDE experiment had three gun types (Smith&Wesson, Ruger, and Sig Sauer), four physical guns of each type, for a total of 12 guns, three ammo types (Remington, Winchester, and PMC), and three repeat days, for a total of 108 (= 3 × 4 × 3 × 3) test firings. Carrying out a

76

similar analysis as that described above for the De Kinder data, we compare a given test fired casing with the remaining 107 test fired casings and compute an appropriate similarity metric (e.g., the $ACCF_{max}$) to yield a total correlation data base consisting of $108 \times 107 = 11\,556$ comparisons. For a given test fired casing, there will be eight (3 ammos × 3 days – 1) other compared casings that "match" (come from the same physical gun), and a total of 99 (= 11 guns × 3 ammos × 3 days) compared casings that do not match.

In carrying out a Top Ten analysis as above, we may thus define our comparison to be a "success" if (as before) one or more of the matching casings appear in the Top Ten list, or if two or more do, or if three or more do, all the way up to all 8 matching cases showing up in the Top Ten list. Summing up these "successes" over the 108 reference casings (and converting them to percentages out of the 108), we thus get the following listing:

NBIDE/ Top Ten

	i≥1		i≥2		i≥3		i≥4		i≥5		i≥6		i≥7		i≥8	
	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3	I2	N3
FP (%):	98	100	92	97	71	93	53	83	38	69	18	56	2	41	0	23
BF (%):	100	100	99	100	94	100	86	100	73	100	56	100	35	100	12	94

From this listing we conclude:

1. In stark contrast to the De Kinder data, Breech Face was the better discriminator over Firing Pin for both I-2D and N-3D.
2. For Breech Face, both I-2D and N-3D performed much better on NBIDE than on De Kinder.
3. Success proportions for N-3D and I-2D decrease as stringency increases—more dramatically for Firing Pin, less so for Breech Face.
4. For Breech Face, N-3D was perfect (= 100%) for 7 out of the 8 criteria.
5. Across Firing Pin and Breech face, N-3D performed better than I-2D.

9.3.4 NBIDE: Table of Results

The following table of Top Ten testing results gives detailed information on how the systems performed with respect to the area imaged and the technology system used for the NBIDE casings.

7340d2d7-67ae-4b31-9c9d-0419dd510c7a   20220314-16442

Table 9-2. Comparison of the numbers of *correct* matches out of the ten highest *scoring* matches for N-3D and I-2D applied to both firing pin and breech face. Note that the ammunition codes for the first column are [1=Winchester, 2=Remington, 3=PMC].

Number of Correct Matches in Top Ten (Maximum 8)

Gun – Ammo – RR#	Firing Pin		Breech Face	
	I-2D	N-3D	I-2D	N-3D
Means	3.72	5.63	5.57	7.94
Ruger 41 - 1 - RR 78	4	7	7	8
Ruger 41 - 1 – RR 102	4	6	6	8
Ruger 41 - 1 - RR 111	5	7	7	8
Ruger 41 - 2 - RR 45	1	5	3	8
Ruger 41 - 2 - RR 94	0	8	6	8
Ruger 41 - 2 - RR 134	2	7	1	8
Ruger 41 - 3 - RR 118	5	6	2	8
Ruger 41 - 3 - RR 129	5	7	2	8
Ruger 41 - 3 - RR 142	5	7	2	8
Ruger 42 - 1 - RR 28	6	8	7	8
Ruger 42 - 1 - RR 43	4	8	3	8
Ruger 42 - 1 - RR 75	4	8	5	8
Ruger 42 - 2 - RR 2	3	8	6	8
Ruger 42 - 2 - RR 35	6	8	5	8
Ruger 42 - 2 - RR 50	4	8	6	8
Ruger 42 - 3 - RR 16	5	8	5	8
Ruger 42 - 3 - RR 54	5	8	7	8
Ruger 42 - 3 - RR 72	5	8	6	8
Ruger 46 - 1 - RR 95	3	6	5	8
Ruger 46 - 1 - RR 120	3	5	5	8
Ruger 46 - 1 - RR 125	1	3	5	8
Ruger 46 - 2 - RR 1	2	4	6	8
Ruger 46 - 2 - RR 67	2	5	3	8
Ruger 46 - 2 - RR 82	4	5	5	8
Ruger 46 - 3 - RR 19	3	6	5	8
Ruger 46 - 3 - RR 53	3	6	5	8
Ruger 46 - 3 - RR 136	3	4	4	8
Ruger 48 - 1 - RR 31	7	8	7	8
Ruger 48 - 1 - RR 80	6	6	7	8
Ruger 48 - 1 - RR 96	7	8	6	8
Ruger 48 - 2 - RR 22	0	8	5	8
Ruger 48 - 2 - RR 130	2	3	6	7
Ruger 48 - 2 - RR 138	5	4	8	8
Ruger 48 - 3 - RR 49	5	8	6	8
Ruger 48 - 3 - RR 55	6	8	6	8
Ruger 48 - 3 - RR 139	6	7	8	8

Sig Sauer 30 - 1 - RR 40	1	4	4	8
Sig Sauer 30 - 1 - RR 60	1	4	5	8
Sig Sauer 30 - 1 - RR 89	2	4	6	8
Sig Sauer 30 - 2 - RR 8	2	4	3	8
Sig Sauer 30 - 2 - RR 10	3	5	3	8
Sig Sauer 30 - 2 - RR 17	2	1	5	8
Sig Sauer 30 - 3 - RR 21	2	4	6	8
Sig Sauer 30 - 3 - RR 30	3	4	4	8
Sig Sauer 30 - 3 - RR 135	2	6	5	8
Sig Sauer 31 - 1 - RR 27	4	2	4	8
Sig Sauer 31 - 1 - RR 48	4	5	7	8
Sig Sauer 31 - 1 - RR 114	1	1	4	8
Sig Sauer 31 - 2 - RR 15	3	3	7	8
Sig Sauer 31 - 2 - RR 65	1	2	2	8
Sig Sauer 31 - 2 - RR 92	2	2	7	8
Sig Sauer 31 - 3 - RR 20	3	3	6	8
Sig Sauer 31 - 3 - RR 62	1	1	5	8
Sig Sauer 31 - 3 - RR 119	5	5	6	8
Sig Sauer 32 - 1 - RR 87	5	7	7	7
Sig Sauer 32 - 1 - RR 90	3	5	7	8
Sig Sauer 32 - 1 - RR 91	6	6	8	8
Sig Sauer 32 - 2 - RR 12	5	7	7	8
Sig Sauer 32 - 2 - RR 25	5	7	7	8
Sig Sauer 32 - 2 - RR 115	3	6	7	8
Sig Sauer 32 - 3 - RR 42	3	8	3	8
Sig Sauer 32 - 3 - RR 56	5	8	7	8
Sig Sauer 32 - 3 - RR 100	5	8	7	8
Sig Sauer 33 - 1 - RR 23	6	6	4	8
Sig Sauer 33 - 1 - RR 66	5	6	5	8
Sig Sauer 33 - 1 - RR 99	4	5	4	7
Sig Sauer 33 - 2 - RR 32	4	2	3	8
Sig Sauer 33 - 2 - RR 34	4	2	3	8
Sig Sauer 33 - 2 - RR 141	2	6	3	8
Sig Sauer 33 - 3 - RR 61	4	7	4	8
Sig Sauer 33 - 3 - RR 79	5	6	4	8
Sig Sauer 33 - 3 - RR 128	4	6	4	8
S&W 305 - 1 - RR 57	6	8	8	8
S&W 305 - 1 - RR 64	6	7	6	7
S&W 305 - 1 - RR 97	6	7	8	8
S&W 305 - 2 - RR 24	6	8	7	7
S&W 305 - 2 - RR 103	2	7	5	8
S&W 305 - 2 - RR 137	2	8	5	8
S&W 305 - 3 - RR 4	6	8	7	8
S&W 305 - 3 - RR 5	6	8	8	8
S&W 305 - 3 - RR 59	6	8	8	8

S&W 306 - 1 - RR 7	4	6	8	8
S&W 306 - 1 - RR 26	2	4	7	8
S&W 306 - 1 - RR 71	3	3	8	8
S&W 306 - 2 - RR 106	2	4	8	8
S&W 306 - 2 - RR 121	2	3	7	8
S&W 306 - 2 - RR 131	2	3	7	8
S&W 306 - 3 - RR 13	3	6	8	8
S&W 306 - 3 - RR 41	3	7	8	8
S&W 306 - 3 - RR 143	2	3	8	8
S&W 314 - 1 - RR 85	3	7	4	7
S&W 314 - 1 - RR 112	6	8	5	7
S&W 314 - 1 - RR 127	5	7	6	8
S&W 314 - 2 - RR 36	6	6	5	8
S&W 314 - 2 - RR 39	6	7	6	8
S&W 314 - 2 - RR 116	2	4	6	8
S&W 314 - 3 - RR 6	5	7	6	8
S&W 314 - 3 - RR 29	5	7	7	8
S&W 314 - 3 - RR 117	5	5	6	8
S&W 401 - 1 - RR 3	2	5	6	8
S&W 401 - 1 - RR 46	3	4	4	8
S&W 401 - 1 - RR 63	2	5	7	8
S&W 401 - 2 - RR 9	5	5	6	8
S&W 401 - 2 - RR 84	3	3	2	8
S&W 401 - 2 - RR 110	4	3	7	8
S&W 401 - 3 - RR 44	4	5	6	8
S&W 401 - 3 - RR 51	3	4	7	8
S&W 401 - 3 - RR 76	3	4	4	8

The N-3D Breech Face results are much better than the other results. As an example of gun differences, all methods seemed to work relatively well with Ruger 42.

9.3.5 Findings: System Performance Analysis

A summary of some of the results of the previous sections is given in the following listing.

Proportion of Correct Matches in Top Ten Listings

I-2D De Kinder FP	3.06/6=0.51
I-2D De Kinder BF	1.01/6=0.17
N-3D De Kinder FP	3.26/6=0.54
N-3D De Kinder BF	2.83/6=0.47
I-2D NBIDE FP	3.72/8=0.46
I-2D NBIDE BF	5.57/8=0.70
N-3D NBIDE FP	5.63/8=0.70
N-3D NBIDE BF	7.94/8=0.99

Based on these results and others in the previous six sections, we thus extract the following observations:

1. For both the De Kinder and the NBIDE data, N-3D generally performed better on the average than I-2D.
2. For the De Kinder data, Firing Pin was usually a better discriminator than Breech Face.
3. For the NBIDE data, Breech Face was a better discriminator than Firing Pin.
4. The worst discriminator case was De Kinder Breech Face.
5. The best discriminator (near perfect) was N-3D on NBIDE Breech Face.

9.4   Matching and Non-matching Distributions of Correlation Scores

The previous sections on system performance analysis have demonstrated performance variations between systems and between imaged regions. For instance, N-3D performed better on the NBIDE breech face impressions than anything else, meaning that the topographic breech face images of casings fired from the same gun almost always correlated more highly with each other than with breech face images of casings fired from different guns. In contrast, for the De Kinder breech face impressions, a casing would often correlate more highly with certain casings fired from other guns than with casings fired from the same guns. In this section, we describe a probabilistic model for such findings.

A pair of casings fired from the same gun is called a "match," and a pair of casings fired from different guns is called a "non-match." Therefore, for each particular casing in the De Kinder set, there are six other casings that produce a match with that casing, and 63 that produce a non-match with that casing. The correlations from the six matching pairs should ideally be considerably higher than the 63 non-matching correlations. Similarly, for each casing in the NBIDE set, there are eight matching correlations, which should be higher than the 99 non-

matching correlations. How well the different technologies succeed in differentiating matches and non-matches will be the topic of this section.

Supposing that a ballistics database contains a casing from the same gun that fired the reference casing, this database search will yield consistently good results only if a pair of objects that should match each other (i.e., fired from same gun) will correlate much more highly than a pair that should not match each other (i.e., fired from different guns). However, as has been seen from the Top Ten results from the previous sections, there can be considerable variability in behavior between guns and even between casings fired from the same gun. For instance, it is possible that two casings fired from different guns may correlate highly with each other, but two other casings fired from the same gun may not correlate as well. Thus, the marks left by guns and the correlations found between images are not deterministic but have random components.

A probabilistic interpretation of these variations is to envision correlation scores of pairs of casings fired from the same gun as being generated by one distribution (which we call the matching distribution), and correlations of casings fired from different guns as following another distribution (the non-matching distribution). These distributions can be purely empirical in nature, rather than having specific parametric forms. A large degree of overlap between these two distributions will result in a significant number of false matches occurring during a database search for a match. If there is almost perfect separation, as shown in Fig. 9-3, then there will be minimal mistakes. Note that for most of this section, the $ACCF_{max}$ values will be scaled so that the maximum (perfect correlation) is 100 %.
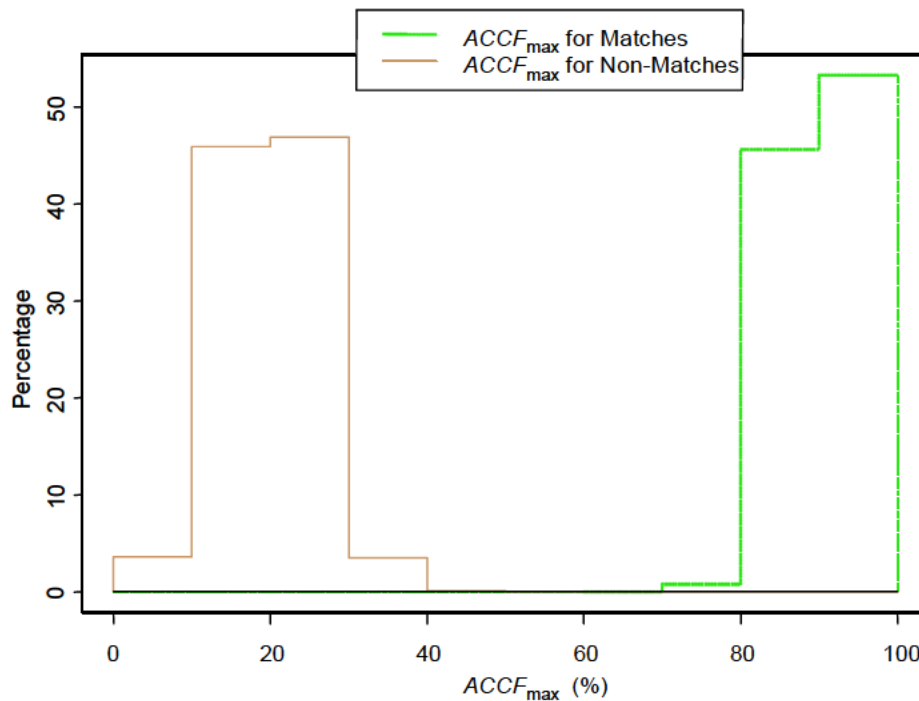
Figure 9-3. Idealized histograms of a matching distribution and a non-matching distribution of cross-correlation scores of ballistics surface topography.

82

The N-3D data contain the full round-robin of correlations between each pair of casings in the De Kinder 70-casing set, and also between each pair of casings in the NBIDE set. A complete round-robin correlation data set is usually not available, and indeed would not be feasible for much larger data sets because of the computational burden. These data provide an opportunity to explore the respective variabilities of matching and non-matching correlation scores and their dependence on underlying factors such as gun model.

For example, for each of the 70 De Kinder casings, the topography image of its firing pin impression was correlated with that of each of the 69 other casings. It is as if each casing were a reference casing found at a crime scene and were compared with 69 casings in a database (i.e. 69 *compared* casings). For arbitrary casings A and B, we denote $ACCF_{max}(A,B)$ as the maximum *ACCF* value between the reference casing A and the compared casing B. Each such casing pair, A and B, has two associated cross-correlations, since $ACCF_{max}(A,B)$ and $ACCF_{max}(B,A)$ are not necessarily identical (although they are usually very close). Thus, for the 70 casings, there are a total of 70×69 correlations = 4830 correlations that will serve as our basic data set for the De Kinder firing pin analysis. Note that $ACCF_{max}(A,A)$ is presumed to be 100 %.

It is natural to display such round-robin correlation matrix data as a color matrix. For instance, if the 70 De Kinder casings are labeled from 1 to 70, the correlation matrix is depicted as a 70×70 pixel chart, where the color of the pixel in, say, row 31 and column 45, is indicative of the size of the $ACCF_{max}$ between casing 31 as the reference and casing 45 as the compared casing.

Figure 9-4 displays what a fictional hypothetical $ACCF_{max}$ matrix for the De Kinder guns would look like with ideal discrimination properties. Here the casings in the matrix are ordered by reference gun, with the white lines partitioning different guns but grouping together the casings fired by the same gun. Those correlations depicted by the pixels inside the boxes with white numbers are $ACCF_{max}$ values of those casing pairs where both the reference casing and compared casing are fired from the same gun. Hence, those pixels depict the matching scores. The pixels outside the numbered boxes depict the non-matching scores.

For the hypothetical idealized case represented by Fig. 9-4, the matching scores are depicted by the orange and reddish pixels inside the numbered boxes; hence the matching scores are almost all above 70 %. A casing is presumed to correlate perfectly with itself, leading to the dark red line of pixels along the diagonal having $ACCF_{max} = 100$ %. The non-matching scores are represented by the bluish pixels outside the numbered boxes; thus, the matching scores are almost all below 25 %. There appears to be no overlap between matching and non-matching scores, which precludes mistakes by Top Ten selection procedures as discussed in previous sections.
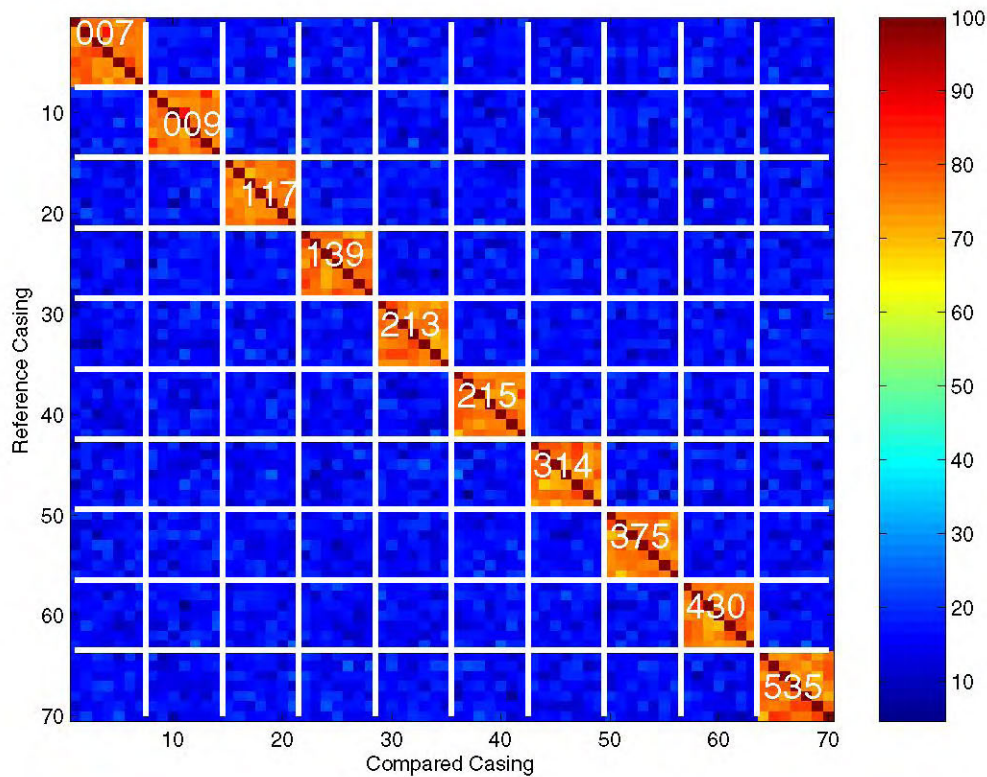
83

Figure 9-4. $ACCF_{max}$ matrix for the 70 De Kinder casings with ideal discrimination between guns. The color of the pixel in row I and column J indicates the value of the $ACCF_{max}$ (in %) between casing I as the reference casing and casing J as the compared casing. A casing correlates perfectly with itself. The casings in the matrix are ordered by reference gun. The white lines partition the casings into groups that are fired by the same reference gun or compared gun. Those correlations depicted by the pixels inside the boxes with white numbers are $ACCF_{max}$ values of those casing pairs where both the reference casing and compared casing are fired from the same gun. Hence, those pixels depict the matching scores. The pixels outside the numbered boxes depict the non-matching scores.

9.5   Overlap Metric p

There are many possible ways to quantify the degree of separation, or distance, between the matching and non-matching $ACCF_{max}$ distributions. One way we use here is to compute an *overlap metric p,* which is the probability that the $ACCF_{max}$ value of a randomly chosen member of the non-matching distribution is larger than a randomly chosen member from the matching distribution.

Since matching scores should ideally be near 100 % and non-matching correlations should ideally be near 0, the probability that a non-matching score exceeds a matching score should be at or near zero. If the two distributions were the same, then p would be 0.5. Such an overlap metric is commonly used in other fields; and it is referred to in the psychometric literature as the Probability of Superiority [49,50]. It is also related to the area under a Receiving Operating Characteristic (ROC) Curve [51] in the statistics literature.

A simple way to estimate *p* is to look at all pair-wise comparisons between single observations from each of the matching and non-matching samples and calculate the proportion in which the non-matching observation score is higher than the matching observation score. Such calculations are similar to those used in the Mann-Whitney Test [52]. Unfortunately, it is very difficult to accurately estimate very small *p* using sample sizes as small as those used in this study, leading to too many estimates of zero for *p*. Another way to estimate *p* is to fit a continuous distribution, such as the normal distribution, to the empirical $ACCF_{max}$ distributions. This ameliorates the problem by using tail values of the continuous distribution to provide estimates of very small probabilities. We will do this especially in those cases where there is little overlap, leading to very small estimates of *p*.

In the following sub-sections, we use the *p* and average *p* of $ACCF_{max}$ values grouped by gun as a descriptive method to compare the guns and the imaging methods. Since the behavior of casings fired from even the same gun can differ (as will be seen in some of the $ACCF_{max}$ matrices below), the *p*'s estimated from individual casings will also be explored. Since the sample sizes for individual casings are smaller, fitting the matching and non-matching distributions using parametric distributions makes sense here. Again, the average and median of the casing *p*'s are displayed for descriptive and comparison purposes. How they will be used and combined to predict the performance of the $ACCF_{max}$ measures in Top Ten lists (similar to those produced by I-2D) is a more complicated procedure that will be described in the later sub-section on binomial models.

The calculation of *p* for different levels of grouping can be visualized in terms of comparisons made within the $ACCF_{max}$ matrix. The hypothetical $ACCF_{max}$ matrix in Fig. 9-4 will be used for an intuitive view of the overlap calculations. We discuss three types of groupings below.

Single *p*
Suppose the same matching and non-matching distributions can be used for all casings and guns. In terms of the $ACCF_{max}$ matrix in Fig. 9-4, the sample of scores inside all of the numbered boxes constitutes the matching scores, and all those outside the numbered boxes are the non-matching scores. It turns out for this idealized case that all the matching scores are large in magnitude (orange-red colors), while all the non-matching scores are small (bluish color). Unfortunately, there usually will not be this much separation between matching and non-matching scores.

Gun specific *p*
There can be different matching and non-matching distributions for each gun, resulting in a different *p* for each gun. Referring back to Fig. 9-4, the matching scores for gun 007 are those represented by the pixels inside the white-bordered box numbered "007"; the corresponding non-matching scores are those depicted by the pixels contained in the nine other white-bordered boxes in the first row (the same level as the box labeled "007"). For gun 007, there are 7×6=42 matching scores (not counting the self-correlations on the main diagonal) and 7×7×9=441 non-matching scores.

Casing specific *p*

There can be different matching and non-matching distributions for each casing, resulting in a different *p* for each casing. Referring back to Fig. 9-4, suppose that the pixels on the first row of the matrix depict the $ACCF_{max}$ scores when casing 1 is the reference casing. Thus casing 1 is compared with six other casings fired from the same gun (gun 007), and with $7 \times 9 = 63$ casings fired from different guns. Having only six matching scores can make the estimation of *p* problematic, especially if *p* is very small. For this reason, estimating the matching and non-matching distributions here can be useful, although estimating the parameters of a distribution using only 6 observations is problematic as well.

9.6   Data Analysis of Correlation Distributions

For each of the N-3D data sets (De Kinder Firing Pin, De Kinder Breech Face, NBIDE Firing Pin, NBIDE Breech Face), we thus present the data on the round-robin correlations in several stages: first, with a color depiction of the entire data set using a color $ACCF_{max}$ matrix, then by examination of the matching and non-matching empirical distributions at three levels of grouping: 1) overall, 2) by gun, and 3) by casing..

9.6.1 De Kinder Firing Pin Correlation Analysis

This subsection contains analysis of the firing pin image data of the De Kinder casings. Recall that the De Kinder set consists of 70 casings fired from ten guns, with seven casings fired from each gun. Thus, given any particular casing, the other 69 (= 70 – 1) casings include six (= 7 – 1) casings fired from the same gun, and 63 (= 70 – 7) casings that were fired from different guns. Therefore, for each particular casing in the De Kinder set, there are six other casings that produce a match with that casing, and 63 that produce a non-match with that casing. The correlations from the six matching pairs should ideally be considerably higher than the 63 non-matching correlations.

Recall the hypothetical $ACCF_{max}$ matrix of Fig. 9-4 in which there was no overlap between the matching and non-matching scores. Is such clear separation between matches and non-matches present with actual data? Figure 9-5 contains the color depiction of the actual $ACCF_{max}$ matrix for the De Kinder firing pin topography images.
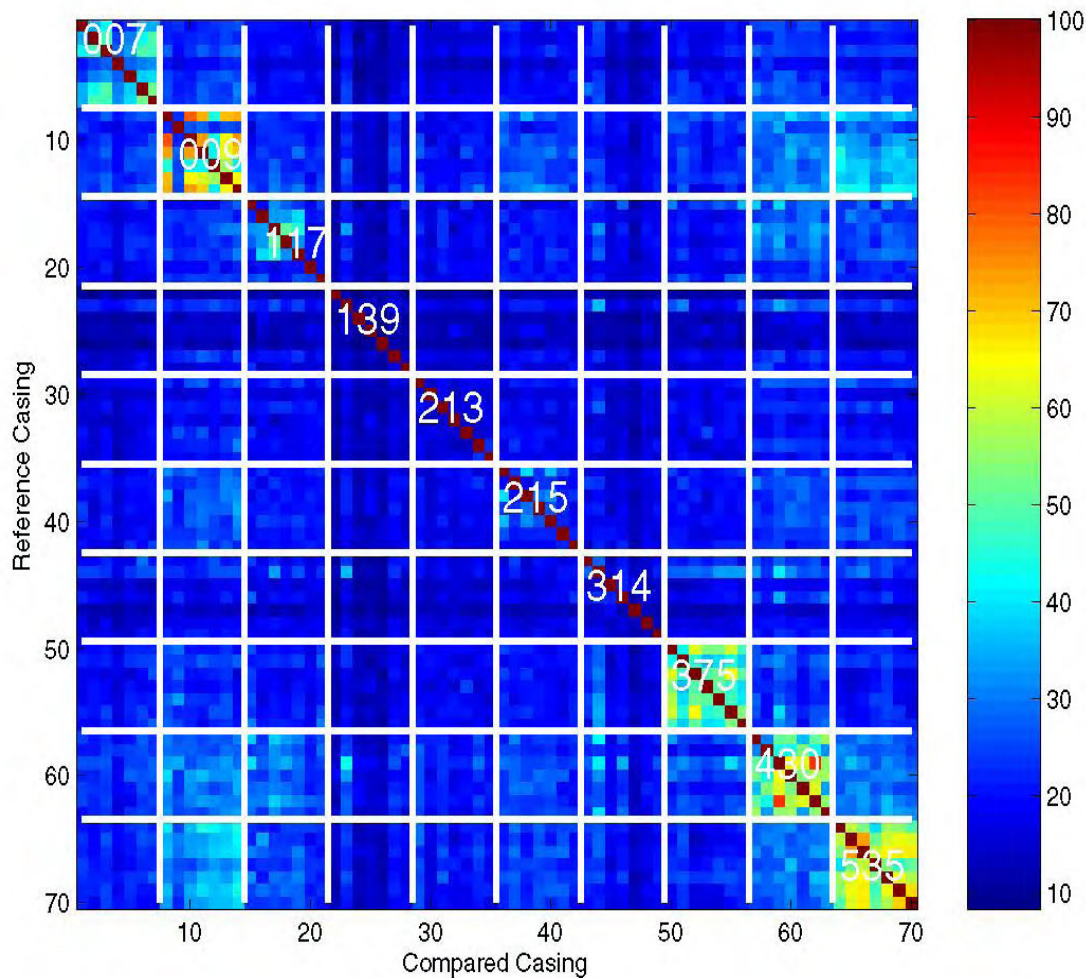
Figure 9-5. $ACCF_{max}$ matrix of the De Kinder firing pin data. The color of the pixel in row I and column J indicates the value of the $ACCF_{max}$ (in %) between casing I as the reference casing and casing J as the compared casing. The pixels are ordered by gun ID, and within gun by ammunition in the following order: CCI, Winchester, Speer, Wolf, Federal, Remington, Remington.

Looking at the color $ACCF_{max}$ matrix for the actual data shows that separation between matching and non-matching scores is not close to the ideal situation depicted in the hypothetical $ACCF_{max}$ matrix of Fig. 9-4. There are many casings that should correlate highly with each other that do not, especially for guns 139, 213, 215, and 314. Guns 007, 009, 375, 430 and 535 have much better (but not perfect) separation. One can see many other patterns from the $ACCF_{max}$ matrix. For instance, the Winchester casing from gun 009 does not correlate with the other casings from that gun. The highest non-match scores are between the casings from guns 009 and 535. The scores generated by gun 139 seem particularly low for both matches and non-matches.

Let's try to summarize the data depicted in Fig. 9-5 in useful groupings. How clearly separated are the distributions of the matching and non-matching correlation scores? Is it close to the ideal

situation of Fig. 9-3? Figure 9-6 contains a histogram of the matching scores and a histogram of the non-matching scores in a plot analogous to Fig. 9-3.
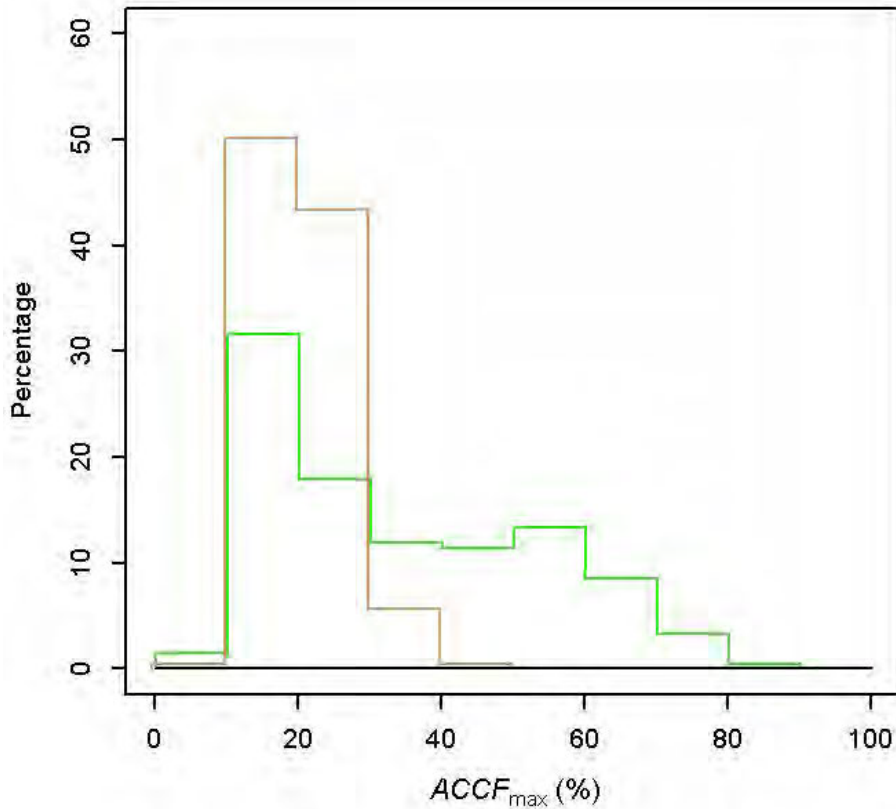

Figure 9-6. De Kinder firing pin correlations: The green lines depict a histogram of the matching scores, while the brown lines depict a histogram of the non-matching scores.

Instead of the clear separation between distributions of Fig. 9-3, there is a large degree of overlap, suggesting that there would be many misidentifications using these correlation data.

The patterns in Fig. 9-5 as well as the multimodal nature of the matching scores in Fig. 9-6 strongly suggest that the matching and non-matching distributions may be gun-dependent. Figure 9-7 breaks down the results by reference gun, with the overlap metric p given for each grouping.
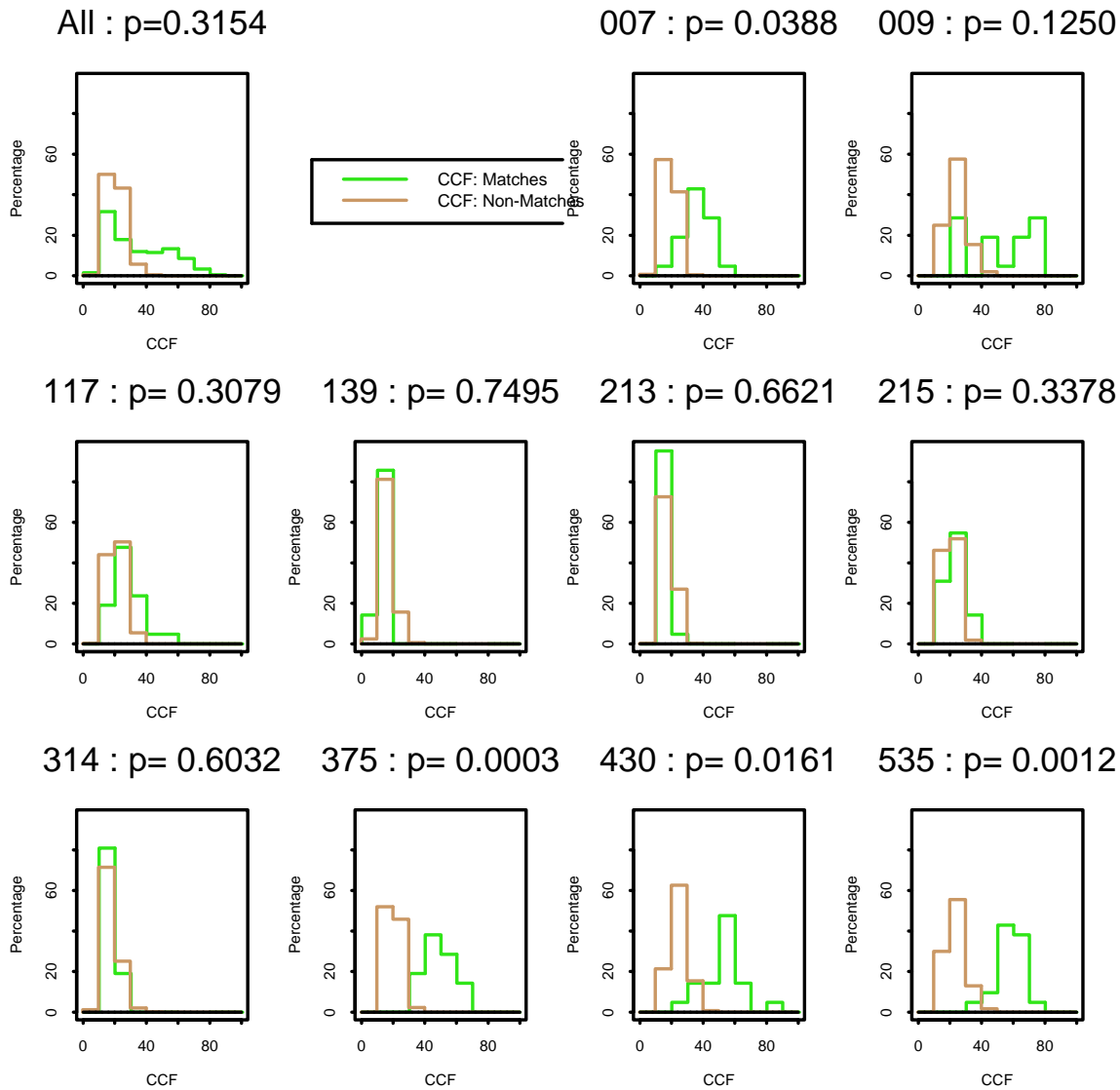
Figure 9-7. De Kinder firing pin correlations: Above each of the individual plots is a heading with the ID of the reference gun and the overlap metric *p* estimated for that group. In each plot, the green lines depict a histogram of the matching scores, while the brown lines depict a histogram of the non-matching scores. The horizontal scale is the same as Fig. 9-6 (%). The symbol CCF stands for $ACCF_{max}$.

Figure 9-7 shows that while the guns have similar non-matching score distributions, they have differing matching distributions. Guns 139, 213 and 314 have match scores that are even more concentrated in the lower end of the scale than their non-match scores. Other guns show much better, though not perfect, separation.

Table 9-3 lists the overlap metric information for the firing pin data of individual gun groupings ordered by size of overlap metric. Small *p* values would indicate better discrimination between matching and non-matching distributions.

Table 9-3. Overlap metric *p* for the topographic signatures of the firing pin impressions of 10 De Kinder guns.

Gun ID	*p*
375	0.0003
535	0.0012
430	0.0161
007	0.0388
009	0.1250
117	0.3079
215	0.3378
314	0.6032
213	0.6621
139	0.7495
Mean	0.2842

Despite all guns being of the same Sig Sauer model, there are large differences in the overlaps produced by the different guns, with guns 139, 213 and 314 producing so much overlap that they are even worse than random chance. While the other guns have much better separation, the level of wrong matches for them are likely still too high for satisfactory performance in a very large database.

It makes sense to further refine overlap metric results by individual casings rather than just by guns. We will continue to use the overlap metric applied to gun scores as a convenient descriptor of gun properties, but we will also calculate *p* for individual casings to examine possible performance in very large databases. As an example, a look back at the $ACCF_{max}$ matrix in Fig. 9-5 reveals some casing-specific patterns. For instance, note that inside the boxes labeled "009" and "007", the matching scores are larger than the non-match scores except for the scores associated with particular pixels. Figure 9-8 breaks down the data by casing, by dividing the data into groups by reference casing. Each of the smaller plots depicts for each casing, its correlations with the 6 casings fired from the same gun (matches, lower triangles), and the correlations with the 63 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID and ammunition manufacturer of the reference casing. The main value of these plots (called "strip plots") is to see the degree of overlap between the six triangles depicting the matching scores with the mass of points representing the non-matching scores. Does the overlap vary greatly among casings fired from the same gun (which are on the same row of the diagram)? Again, ideally there should be clear separation between the matching and non-matching scores.

From Fig. 9-8, it can be seen that guns 375 and 535 produce the greatest separation. Note that because the triangles in Fig. 9-8 have non-zero width, some of the casing plots may give the impression of somewhat more overlap than actually exists.
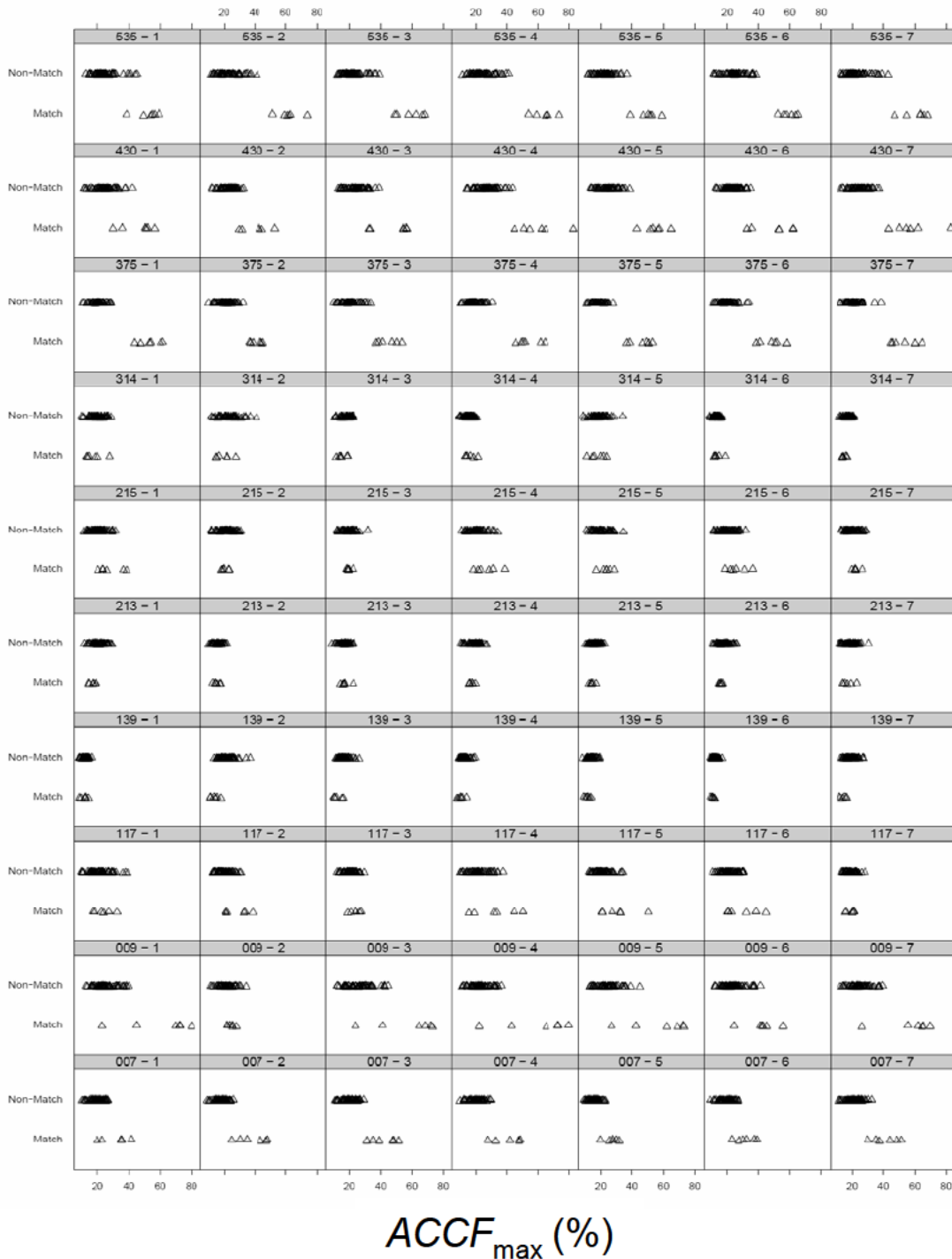
Figure 9-8. De Kinder Firing Pin: Correlations for individual casings. The above figure plots for each casing its correlations with the six casings fired from the same gun (matches, lower triangles), and the correlations with the 63 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID and ammunition manufacturer (1-CCI, 2-Win, 3-Rem, 4-Speer, 5-Wolf, 6-Fed, 7-Rem) of the reference casing.

For each of the 70 casings, an overlap metric for the matching and non-matching correlation scores produced by that casing can be estimated by looking at all pair-wise comparisons between

91

matching and non-matching correlations for it. The histogram in Figure 9-9 shows the empirical distribution of those overlap metrics. How do these estimates compare with estimates by gun?

The histogram in Fig. 9-9 shows that while some casings produce small overlaps, most produce substantial overlaps between matching and non-matching correlation scores that would lead to mistakes in a large database search.
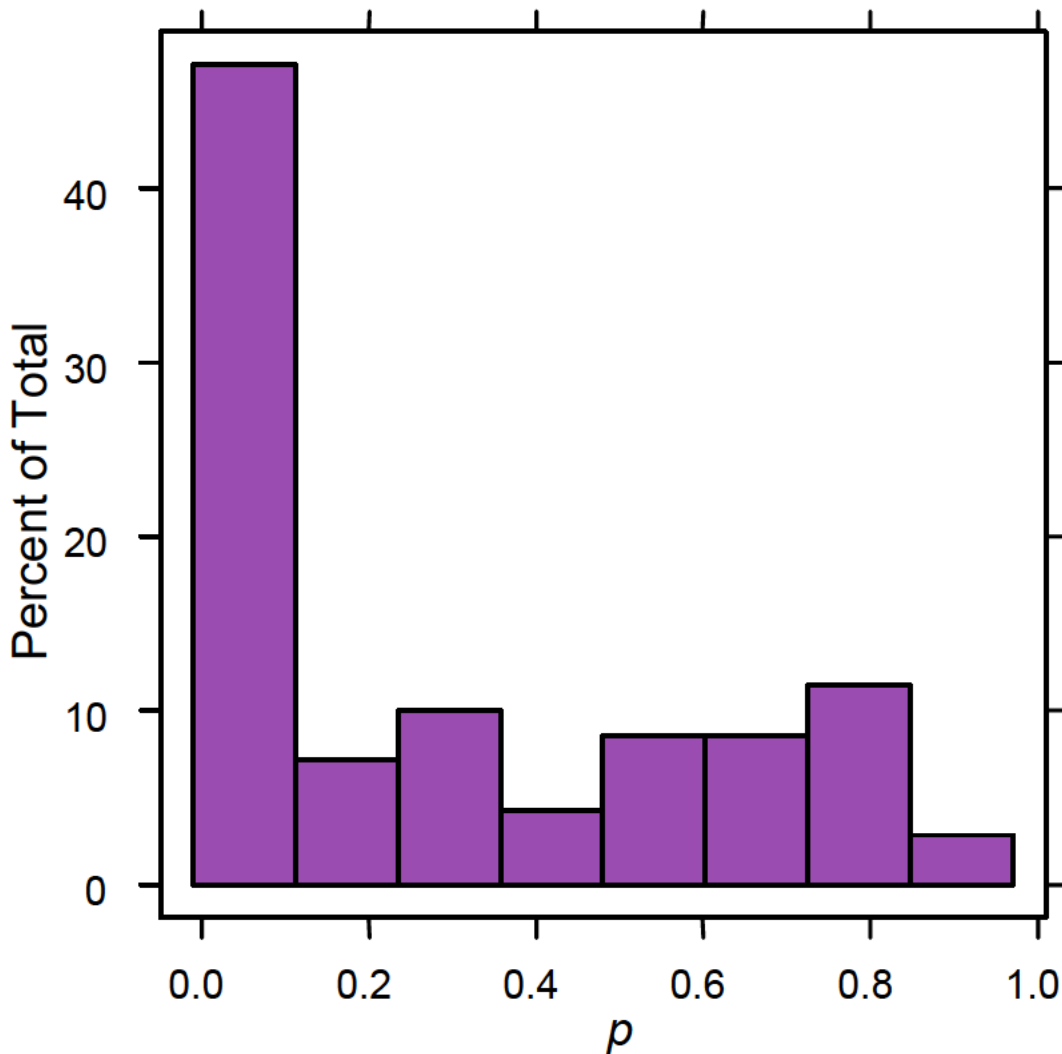
Figure 9-9. Histogram of the overlap metric p estimated for each of the 70 De Kinder casings using topographic signatures of the firing pin impressions and pair-wise comparison methods. The mean p is 0.29 and the median is 0.17. While 24 % of the estimates are zero, there are a considerable proportion of large estimates over 0.3.

9.6.2 De Kinder Breech Face Correlation Analysis

Figure 9-10 contains the $ACCF_{max}$ matrix for the topographic breech face analyses for the De Kinder casings. How does the pattern of separation and overlap between matching and non-matching scores compare with the firing pin data?
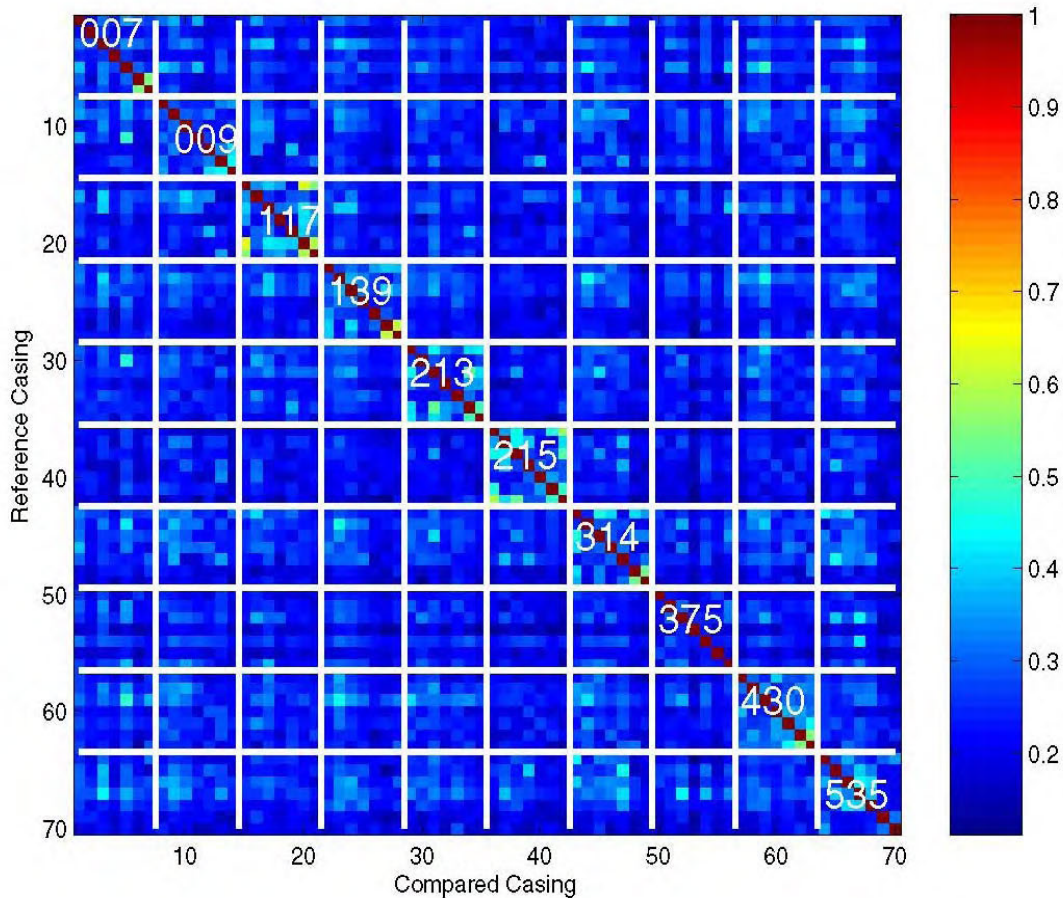
92

Figure 9-10. *ACCF*$_{max}$ matrix of the De Kinder breech face data. The color of the pixel in row *I* and column *J* indicates the value of the *ACCF*$_{max}$ between casing *I* as the reference casing and casing *J* as the compared casing. The pixels are ordered by gun ID, and within gun by ammunition in the following order: CCI, Winchester, Speer, Wolf, Federal, Remington, Remington.

It appears that most *ACCF*$_{max}$ values, including both matching scores and non-matching scores are in the 20 % to 40 % range, indicating more overlap and less separation than was seen with the firing pins. The green-colored pixels in the lower right corner of most boxes on the diagonal show that the breech faces from the two Remington casings are correlating more highly with each other than with the other casings fired from the same guns.

Figure 9-11 depicts histograms of the matching and non-matching scores for the De Kinder breech face impressions. What is the degree of overlap or separation between the two distributions?
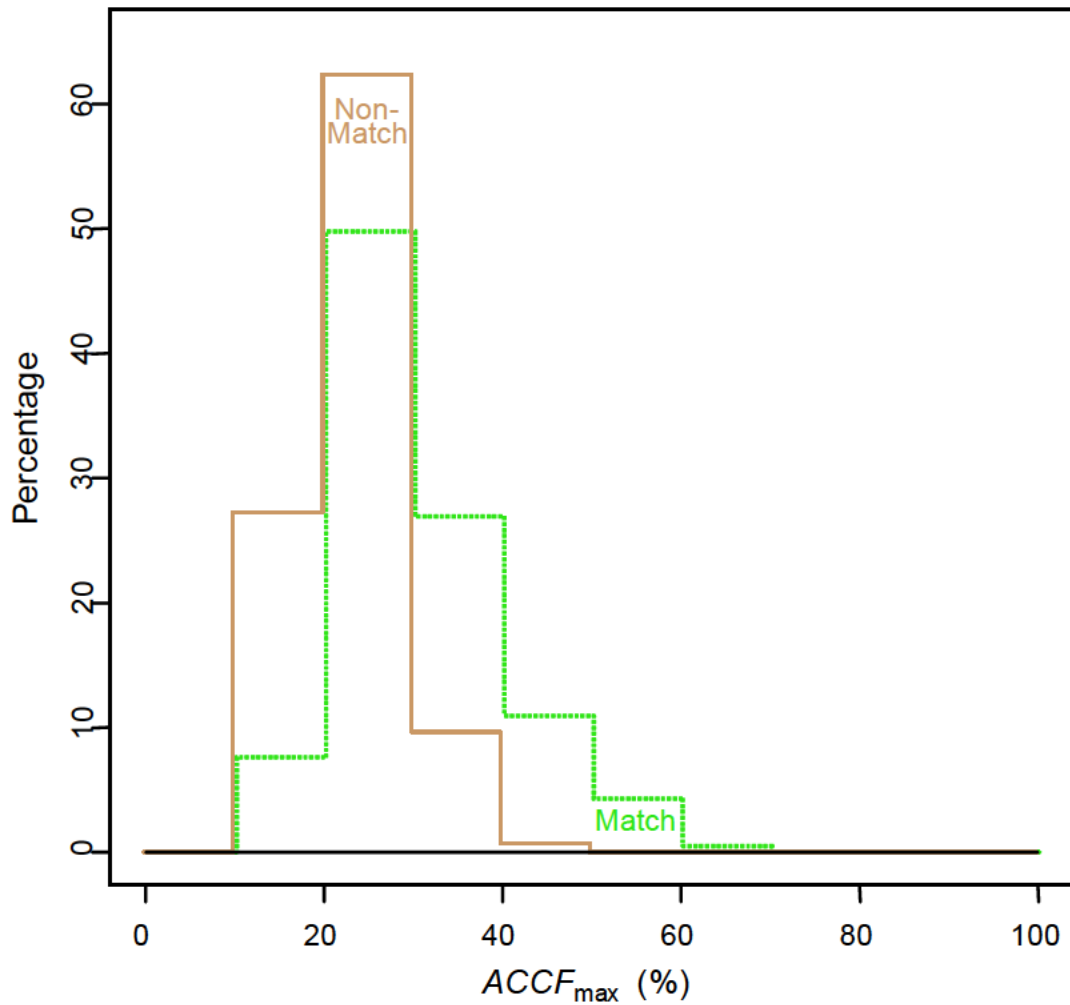
93

Figure 9-11. De Kinder breech face correlations: The green dashed lines depict a histogram of the matching scores, while the brown solid lines depict a histogram of the non-matching scores.

Figure 9-11 shows that there is a considerable degree of overlap between the matching and non-matching scores, which is not ideal. Figure 9-12 breaks down these results by reference gun, with the overlap metric p given for each grouping. For the firing pin impressions, there were considerable performance variations between guns. Does the same hold true for breech faces?

94
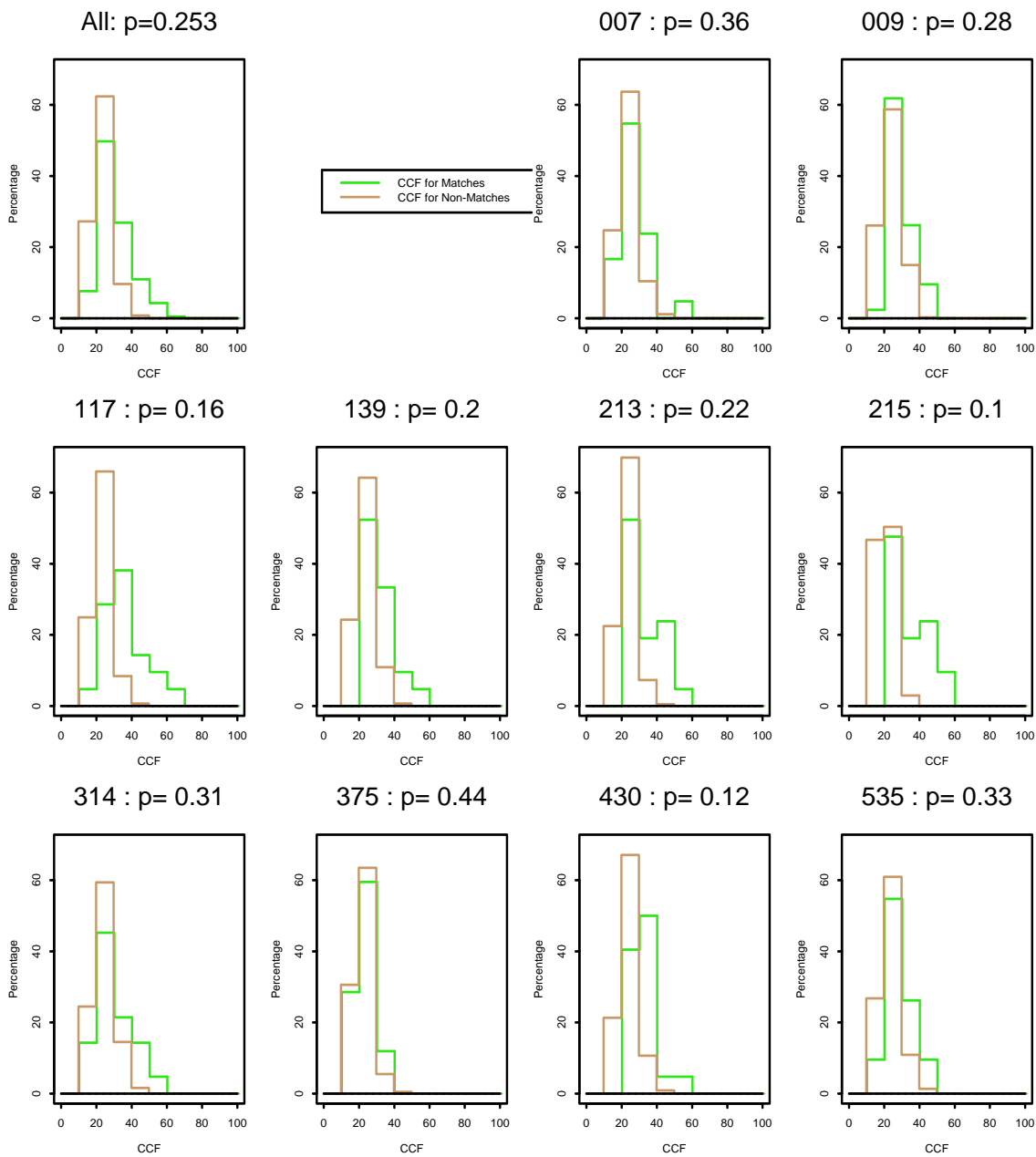
Figure 9-12. De Kinder breech face correlations: Above each of the individual plots is a heading with the ID of the reference gun and the overlap metric *p* estimated for that group. In each plot, the green lines depict a histogram of the matching scores, while the brown lines depict a histogram of the non-matching scores. The symbol CCF stands for $ACCF_{max}$ The horizontal scale for $ACCF_{max}$ is the same as Fig. 9-6 (%).

No gun has good separation between the matching and non-matching distributions. Also for the De Kinder breech face distributions in Fig. 9-12, we find less variation between the guns than we did for the De Kinder firing pin data. All the non-matching distributions look similar.

95

Table 9-4 contains the overlap metric statistics for the breech face data, ordered by performance of the gun. Small *p* values indicate better separation between matching and non-matching distributions. None of the guns have a large degree of separation. Even the gun with the most separation (215) still would produce many errors in a large database search scenario.

Table 9-4. Overlap metric *p* for the topographic signatures of the breech face impressions of 10 De Kinder guns.

Gun ID	*p*
215	0.098
430	0.124
117	0.158
139	0.201
213	0.219
009	0.283
314	0.314
535	0.332
007	0.358
375	0.442
Mean	0.253

Figure 9-13 breaks down the data more finely by dividing the data into groups by reference casing. Each of the smaller plots depicts, for each casing, its correlations with the six casings fired from the same gun (matches, lower triangles), and the correlations with the 63 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID and ammunition of the reference casing. Although there is some variation in the matching distributions, there is a consistently high degree of overlap or lack of separation between matching and non-matching casings for nearly all reference casings. From Fig. 9-13, it can be seen that only one casing, 215-3, has good separation between matching and non-matching scores, and even for that case, there is a very small margin between the smallest matching score and the largest non-matching score. Some of the other casings from gun 215 are among the next best performers in terms of separation between matching and non-matching scores.

For each of the 70 casings, an overlap metric for the matching and non-matching correlation scores produced by that casing can be estimated by looking at all pair-wise comparisons between matching and non-matching correlations for each casing. The histogram in Fig. 9-14 shows the empirical distribution of those overlap metrics. While some casings produce small overlaps, most produce substantial overlaps between matching and non-matching correlation scores that would lead to mistakes in a large database search.

96
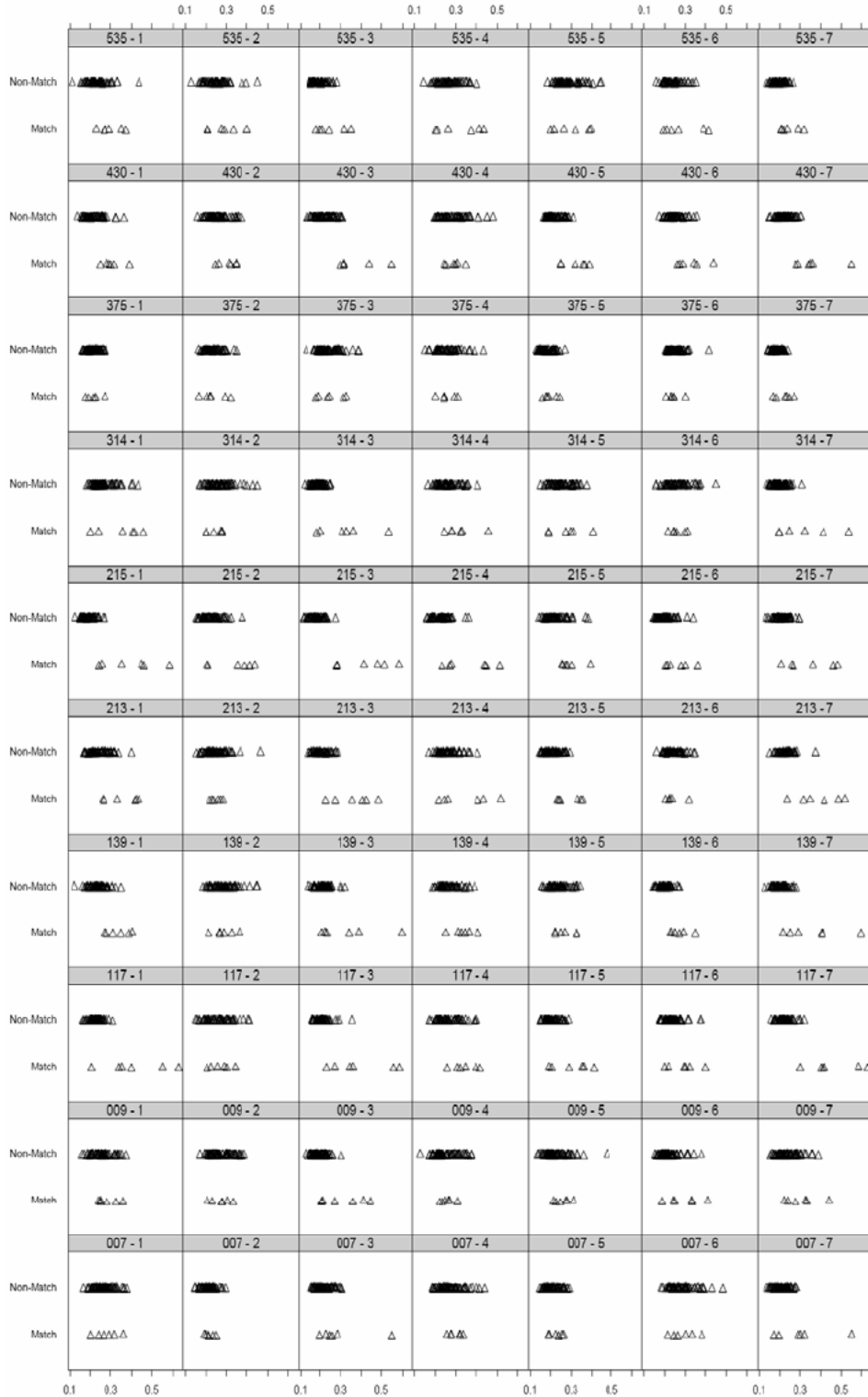
Figure 9-13. De Kinder Breech Face: The above figure plots, for each casing, its correlations with the six casings fired from the same gun (matches, lower triangles), and the correlations with the 63 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID and ammunition manufacturer (1-CCI, 2-Win, 3-Rem, 4-Speer, 5-Wolf, 6-Fed, 7-Rem) of the reference casing.
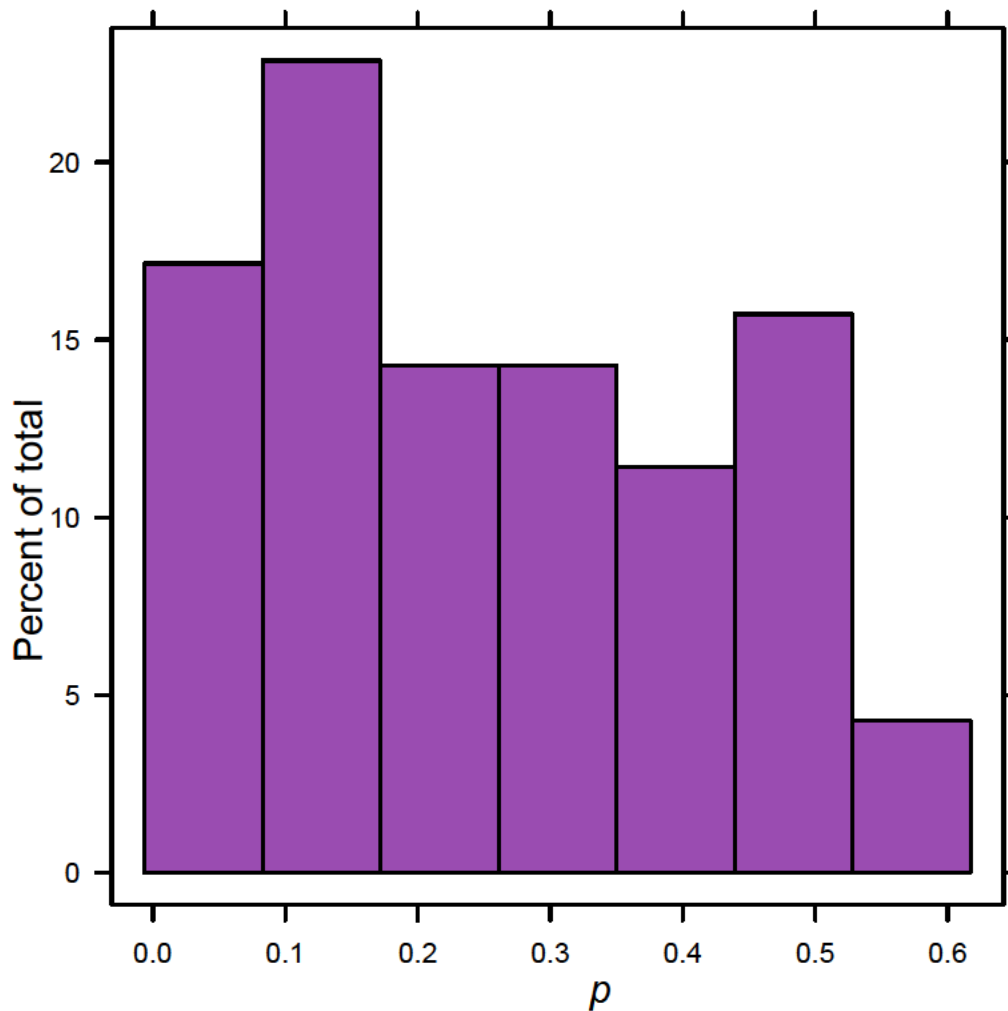
Figure 9-14. Histogram of the overlap metric p estimated for each of the 70 De Kinder casings using topographic signatures of the breech face impressions and pair-wise comparison methods. The mean is 0.26 and the median is 0.23.

9.6.3 De Kinder: Combining Firing Pin and Breech Face Results

A combination of the correlation metrics from both regions of the casings, e.g., firing pin and breech face, should perform better than the use of one region alone. There has been similar research on multi-modal biometrics, i.e. combining several fingerprints or combining face and fingerprint algorithms [53]. There are many possible methods of combining metrics, but we will just look at two of the simplest here. In the De Kinder article [7], casings that make the Top Ten of *either* I-2D breech face or I-2D firing pin correlations are included in the combined list. Such a combined list obviously must do at least as well as either metric alone, although it results in potentially twice as many candidate casings requiring manual examination. Most other combination schemes require more information than just Top Ten lists.

For example, if the $ACCF_{max}$ values are available, they can be combined in a multitude of ways. The simplest way is first to normalize the metrics, then add them together. Here, since the

$ACCF_{max}$ values are bounded by a perfect score of 100 %, we can produce a combined correlation metric by merely averaging the breech face and firing pin $ACCF_{max}$ values for each casing pair. Of course, the average and sum produce the same results, so here we call it the Sum Method, which is more prevalent in the literature. Unlike the 'make either list' method above, the Sum Method can perform worse for a particular casing than either or both of the constituent methods. Table 9-5 shows the results of the two combined metrics for the De Kinder casings.

Of course, using both metrics results in more, or at least no fewer, correct matches. The number of correct matches using both metrics cannot be more than the sum of correct matches for either, and it is equal to the sum only if there is no overlap. Thus, for I-2D, the strong measure (Firing Pin) cannot be helped much here by the weak method (Breech Face), especially since firing pin impressions caught all 6 correct matches for several casings. The 3D breech face impressions were more of a help to the 3D firing pin impressions.

Table 9-5. Results combining breech face impressions and firing pin impressions for 70 De Kinder casings. Columns 3 and 4 – number of correct matches appearing in the Top Ten List of either the breech face or the firing pin impressions. Column 5 – number of correct matches appearing in the Top Ten list for the sum of $ACCF_{max}$ values for the breech face and firing pin impressions.

Number of Correct matches (Max 6)

Ref.Casing	Ammo	I-2D Either List	N-3D Either List	N-3D Sum
Means		3.39	4.77	4.23
007-01	CCI	4	5	4
007-02	WIN	5	6	5
007-03	R-P	6	6	6
007-04	SPEER	6	6	5
007-05	WOLF	6	5	6
007-06	FC	6	6	4
007-07	R-P.sep	6	6	5
009-01	CCI	3	5	5
009-02	WIN	0	1	0
009-03	R-P	2	5	5
009-04	SPEER	4	5	5
009-05	WOLF	4	5	5
009-06	FC	5	6	5
009-07	R-P.sep	5	5	5
117-01	CCI	3	5	4
117-02	WIN	0	4	3
117-03	R-P	4	6	6
117-04	SPEER	5	6	3
117-05	WOLF	6	6	5

99

117-06	FC	5	5	3
117-07	R-P.sep	3	6	6
139-01	CCI	3	4	3
139-02	WIN	3	1	0
139-03	R-P	4	3	3
139-04	SPEER	4	3	2
139-05	WOLF	2	2	1
139-06	FC	2	4	3
139-07	R-P.sep	3	5	3
213-01	CCI	2	4	3
213-02	WIN	4	0	0
213-03	R-P	5	5	4
213-04	SPEER	3	3	3
213-05	WOLF	3	3	3
213-06	FC	1	1	1
213-07	R-P.sep	2	5	4
215-01	CCI	2	6	6
215-02	WIN	2	4	4
215-03	R-P	3	6	6
215-04	SPEER	1	6	5
215-05	WOLF	2	3	3
215-06	FC	2	5	4
215-07	R-P.sep	2	5	3
314-01	CCI	4	5	3
314-02	WIN	2	0	0
314-03	R-P	5	4	4
314-04	SPEER	3	4	2
314-05	WOLF	4	2	1
314-06	FC	4	2	0
314-07	R-P.sep	4	4	3
375-01	CCI	4	6	6
375-02	WIN	3	6	5
375-03	R-P	4	6	5
375-04	SPEER	5	6	6
375-05	WOLF	5	6	6
375-06	FC	4	6	6
375-07	R-P.sep	4	6	6
430-01	CCI	3	6	6
430-02	WIN	0	6	6
430-03	R-P	5	6	6
430-04	SPEER	3	6	6
430-05	WOLF	3	6	6

445

20220314-16465

430-06	FC	3	6	6
430-07	R-P.sep	5	6	6
535-01	CCI	1	5	6
535-02	WIN	4	6	6
535-03	R-P	3	6	6
535-04	SPEER	3	6	6
535-05	WOLF	2	6	6
535-06	FC	2	6	6
535-07	R-P.sep	2	6	6

Table 9-6 shows the overlap metric statistics for the sum method of N-3D firing pin and breech face impressions.

Table 9-6. Overlap metric p for the sum of the $ACCF_{max}$ values for topographic signatures of the breech face impressions and firing pin impressions of 10 De Kinder guns, ordered by value of overlap metric.

Gun ID	p
535	0.0013
375	0.0056
430	0.0082
007	0.0706
117	0.0820
215	0.0843
009	0.0981
213	0.3303
139	0.3513
314	0.3660
Mean	0.140

The mean p was 0.284 for N-3D firing pins and 0.253 for N-3D breech faces, so the sum measure does somewhat better than either of the two measures. A comparison of the results of Table 9-1 and Table 9-5 shows that this is consistent with the marginal improvement of the mean number of correct matches in the N-3D top ten lists from 2.8 (breech face) and 3.3 (firing pin) to 4.2 (sum).

Although all the $ACCF_{max}$ values are on nominally the same scales, the firing pin $ACCF_{max}$ values are more spread out from 0-100 and thus tend to dominate the sum. This domination can be lessened by a different normalization scheme, e.g., using a z-score that divides by the standard deviation. However, that normalization is more difficult to apply because it requires knowledge of the population or the sample of scores to obtain the standard deviations.

Although it may be true that the use of more and more measures to be combined may always improve performance, there may also be diminishing returns, where each added measure adds only a marginal amount of improvement that may not be worth the extra effort. This will be especially true if one measure dominates or performs better than the others, or if there is dependence between the measures. If one measure performs extremely well by itself, combining it with much weaker measures may not help much.

9.6.4 NBIDE Firing Pin Correlation Analysis

Recall that the NBIDE set contains 108 casings fired from 12 guns, with nine casings fired from each gun. Thus, for any particular casing in the NBIDE set, there are eight (=9 – 1) casings fired from the same gun and 99 (=108 – 9) casings fired from different guns. Therefore, for each particular casing in the NBIDE set, there are eight other casings that produce a match with that casing, and 99 casings that produce a non-match with that casing. The correlations from the eight matching pairs should ideally be considerably higher than the correlations from the 99 non-matching pairs.

Figure 9-15 depicts the $ACCF_{max}$ matrix of the NBIDE firing pin data, where the casings are ordered by gun (Ruger 41, Ruger 42, Ruger 46, Ruger 48, Sig Sauer 30, Sig Sauer 31, Sig Sauer 32, Sig Sauer 33, S&W 305, S&W 306, S&W 314, S&W 401) and within gun by ammunition (1-Win, 2-Rem, 3-PMC), and within ammunition by repetition number (RR#). Recall that there are four guns of each brand; each gun fires three shots of each of the three ammunition brands for a total of $4{\times}3{\times}3{\times}3 = 108$ firings.

Recall the $ACCF_{max}$ matrices for the De Kinder set (Figs. 9-5 and 9-10) and remember how greatly they differed from the "ideal" fictional case of Fig. 9-4. How do the results for the NBIDE firing pin images in Fig. 9-15 compare? It can be seen that most of the non-matching correlations are quite small in magnitude (bluish). Some of the guns have matching correlations that are higher in magnitude (greenish or orange). It looks more like the pattern of the ideal $ACCF_{max}$ matrix in Fig. 9-4, though still far from ideal.

Figure 9-16 depicts histograms of the matching and non-matching scores for the NBIDE firing pin N-3D data. It shows that there is still a considerable degree of overlap between the matching and non-matching scores, which is not ideal.
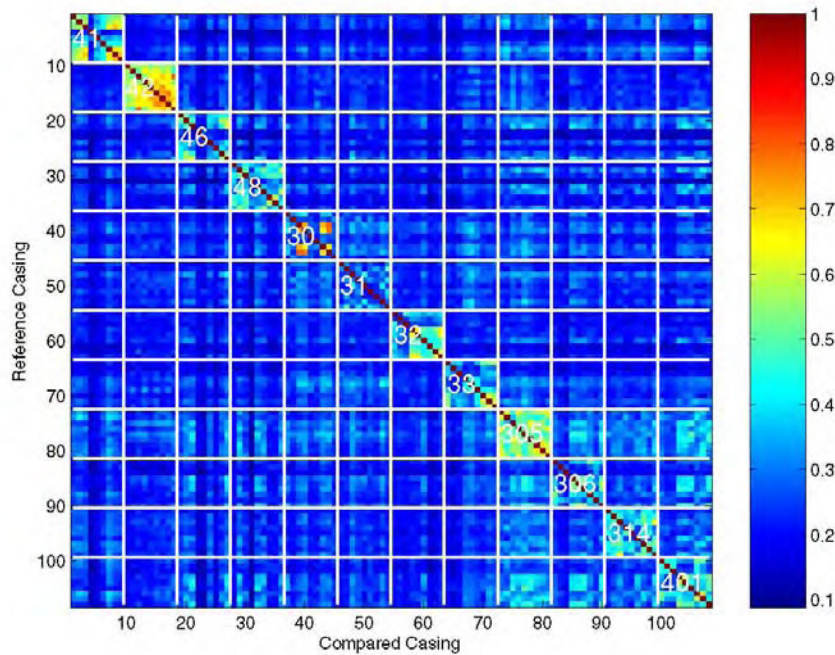
Figure 9-15. $ACCF_{max}$ matrix of the NBIDE firing pin N-3D data. The color of the pixel in row I and column J indicates the value of the $ACCF_{max}$ between casing I as the reference casing and casing J as the compared casing. The pixels are ordered by gun ID (Ruger 41, Ruger 42, Ruger 46, Ruger 48, Sig Sauer 30, Sig Sauer 31, Sig Sauer 32, Sig Sauer 33, S&W 305, S&W 306, S&W 314, S&W 401) with Rugers in the upper left and S&W's in the lower right), and within gun by ammunition (1-Win, 2-Rem, 3-PMC), and within ammunition by RR#.
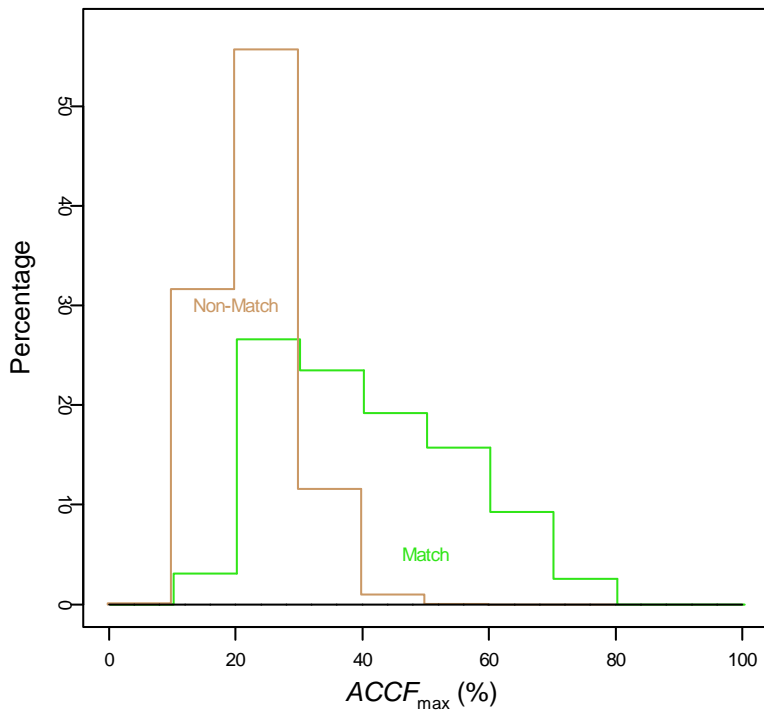
Figure 9-16. N-3D data for NBIDE firing pin impressions: The green lines depict a histogram of the matching scores, while the brown lines depict a histogram of the non-matching scores.

103

Figure 9-17 breaks down these results by reference gun, with the overlap metric *p* given for each grouping. For the De Kinder firing pin impressions, there was considerable performance variation between guns. Does the same hold true for the NBIDE firing pin impressions?
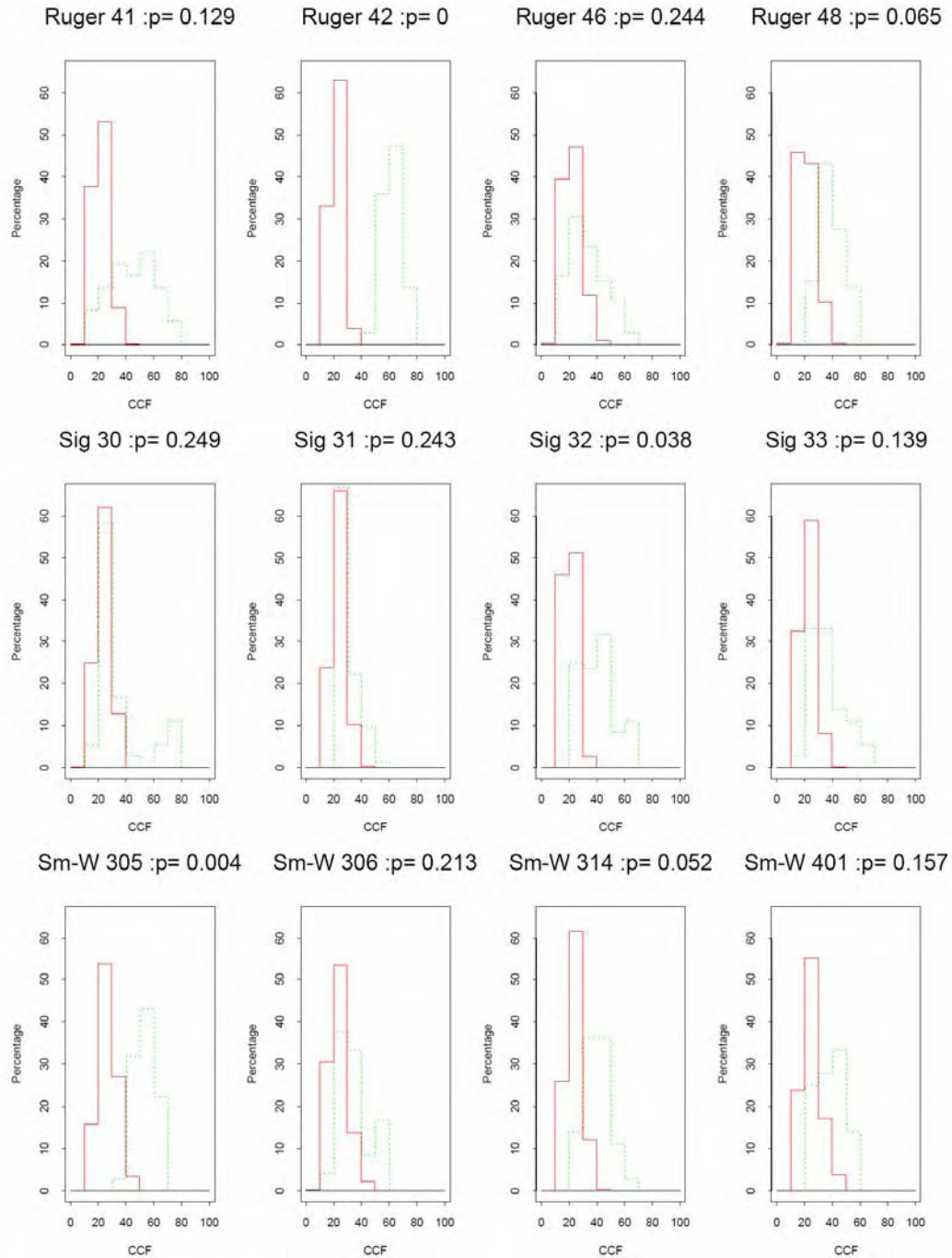
Figure 9-17. NBIDE firing pin N-3D data: Above each of the individual plots is a heading with the ID of the reference gun and the overlap metric *p* estimated for that group. In each plot, the green dotted lines depict a histogram of the matching scores, while the brown solid lines depict a histogram of the non-matching scores. The horizontal scale for $ACCF_{max}$ is the same as Fig. 9-6 (%).

Figure 9-17 shows considerable differences between guns. While the non-matching distributions appear similar for each gun, the matching distributions are quite different for each. There is variability within the gun brands: each brand (depicted by row) has one gun that has much more separation than the others of the same brand (Ruger 42, Sig Sauer 32, and S&W 305); and each brand has at least one gun with large overlap and p metric greater than 0.2.

Table 9-7 contains the overlap metric statistics for the firing pin data ordered by performance of the gun as estimated by the overlap metric.

Table 9-7. Overlap metric p for the $ACCF_{max}$ values for topographic signatures of the firing pin impressions of 12 NBIDE guns.

Gun ID	p
Ruger 42	0
S&W 305	0.004
Sig Sauer 32	0.038
S&W 314	0.052
Ruger 48	0.065
Ruger 41	0.129
Sig Sauer 33	0.139
S&W 401	0.157
S&W 306	0.213
Sig Sauer 31	0.243
Ruger 46	0.244
Sig Sauer 30	0.249
Mean	0.128

There are considerable differences within guns of the same brand.

Figure 9-18 breaks down the data even more finely, by dividing the data into groups by reference casing. Each of the smaller plots depicts for each casing, its correlations with the eight casings fired from the same gun (matches, lower triangles), and the correlations with the 99 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID #, ammunition #, and RR # of the reference casing. How does the degree of overlap or separation between matching and non-matching vary among casings fired from the same gun, which are shown on the same row in the diagram?

From Fig. 9-18, it can be seen that Ruger 41 is not consistent because the row for this gun contains some casings that possess separation between matching and non-matching scores and other casings that have considerable overlaps. For the other guns, there is a lesser but still considerable variability within casings fired from the same gun.
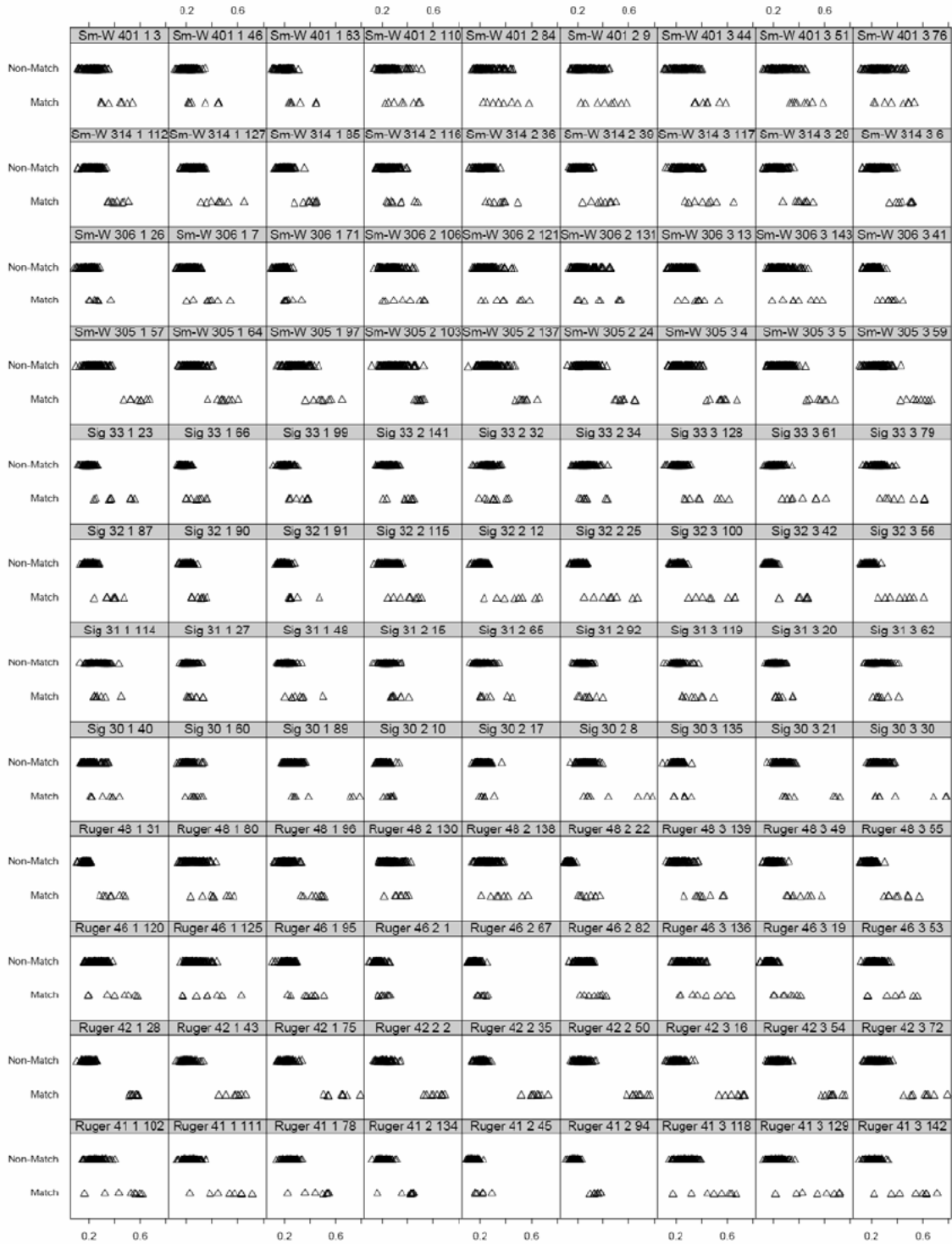
Figure 9-18. NBIDE Firing Pin N-3D data: Correlations for each casing with the eight casings fired from the same gun (matches, lower triangles), and with the 99 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID and ammunition manufacturer (1-Win, 2-Rem, 3-PMC), and RR# of the reference casing

For each of the 108 casings, an overlap metric for the matching and non-matching correlation scores produced by that casing can be estimated by looking at all pair-wise comparisons between matching and non-matching correlations for each casing. The histogram in Fig. 9-19 shows the empirical distribution of those overlap metrics. How do these estimates compare with the estimates by gun?

From Fig. 9-19, it can be seen that the NBIDE firing pins produce greater separation between matching and non-matching distributions than do the De Kinder firing pins, but most still have a degree of overlap that will produce mistakes in a large database scenario.
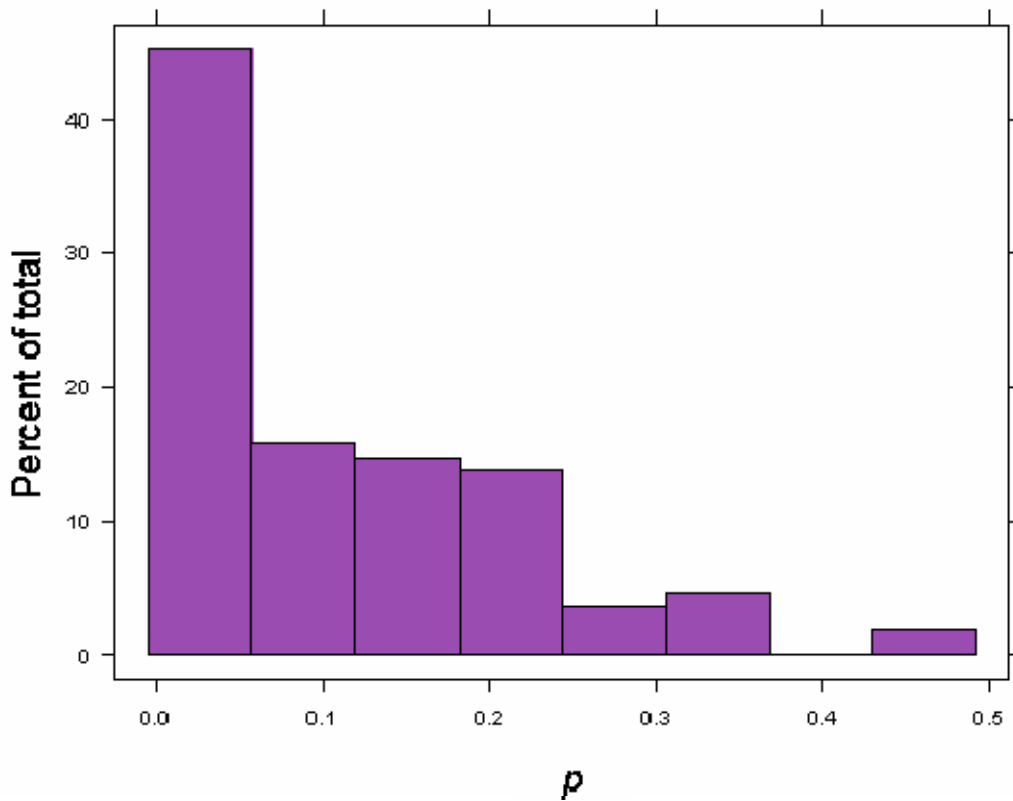
Figure 9-19. NBIDE Firing Pin N-3D data: Histogram of the overlap metric p estimated for each of the 108 NBIDE casings using pair-wise comparison methods. The mean value of p is 0.11 and the median value of p is 0.08. About 75% of the values are larger than 0.01.

9.6.5 NBIDE Breech Face Correlation Analysis

Figure 9-20 depicts the $ACCF_{max}$ matrix of the NBIDE breech face impression data, where the casings are ordered by gun (Ruger 41, Ruger 42, Ruger 46, Ruger 48, Sig Sauer 30, Sig Sauer 31, Sig Sauer 32, Sig Sauer 33, S&W 305, S&W 306, S&W 314, S&W 401), and within gun by ammunition ((1-Win, 2-Rem, 3-PMC), and within ammunition by RR#. Recall that there are four guns of each brand; each gun fires three shots each of each of the three ammunition brands for a total of $4 \times 3 \times 3 \times 3 = 108$ firings.

107

Recall the $ACCF_{max}$ matrices for the De Kinder set (Fig. 9-5 and 9-10) and for the NBIDE firing pin impressions (Fig. 9-15) in the previous section. How do the results for the NBIDE breech face N-3D images in Fig. 9-20 compare?
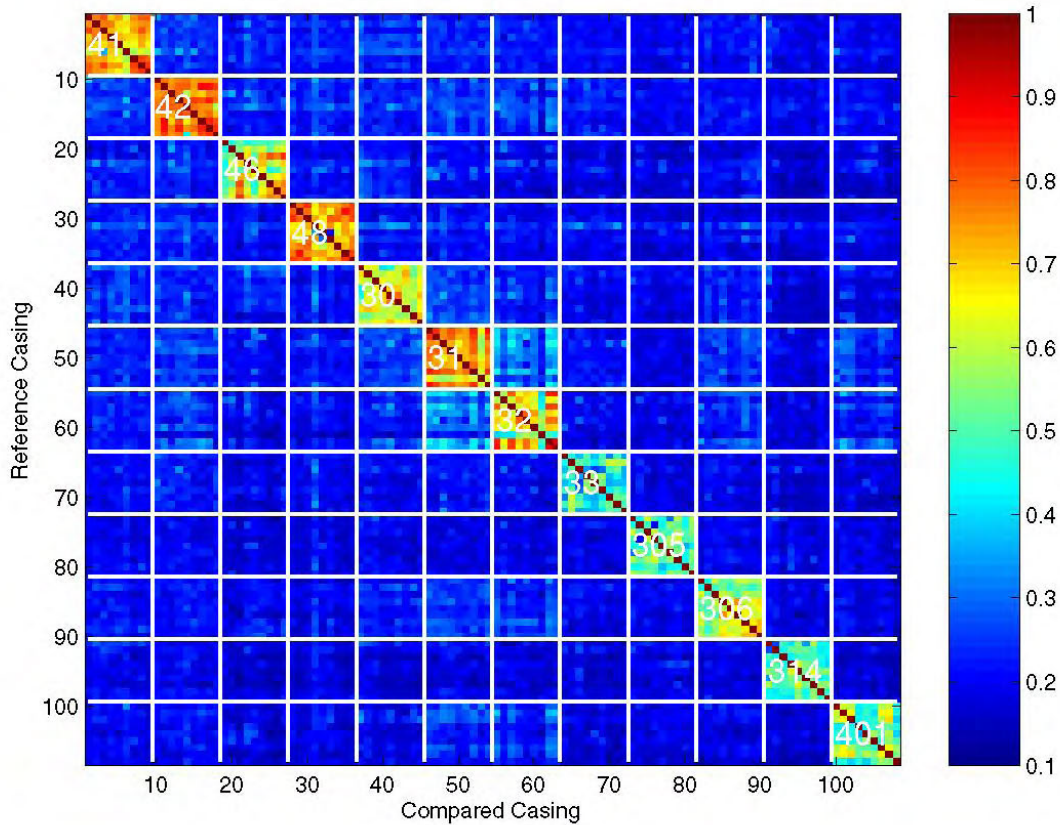

Figure 9-20. $ACCF_{max}$ matrix of the NBIDE Breech Face N-3D data. The color of the pixel in row *I* and column *J* indicates the value of the $ACCF_{max}$ between casing *I* as the reference casing and casing *J* as the compared casing. The pixels are ordered by gun ID (Ruger 41, Ruger 42, Ruger 46, Ruger 48, Sig Sauer 30, Sig Sauer 31, Sig Sauer 32, Sig Sauer 33, S&W 305, S&W 306, S&W 314, S&W 401, with Rugers at upper left and S&W's in lower right), and within gun by ammunition ((1-Win, 2-Rem, 3-PMC), and within ammunition by RR#.

The $ACCF_{max}$ matrix in Fig. 9-20 is much closer to the "ideal" $ACCF_{max}$ matrix in Fig. 9-4 than anything else seen so far.

Figure 9-21 depicts histograms of the matching and non-matching scores for the NBIDE breech faces. Figure 9-21 reveals a much greater degree of separation between matching and non-matching scores than has been seen in the other data sets.
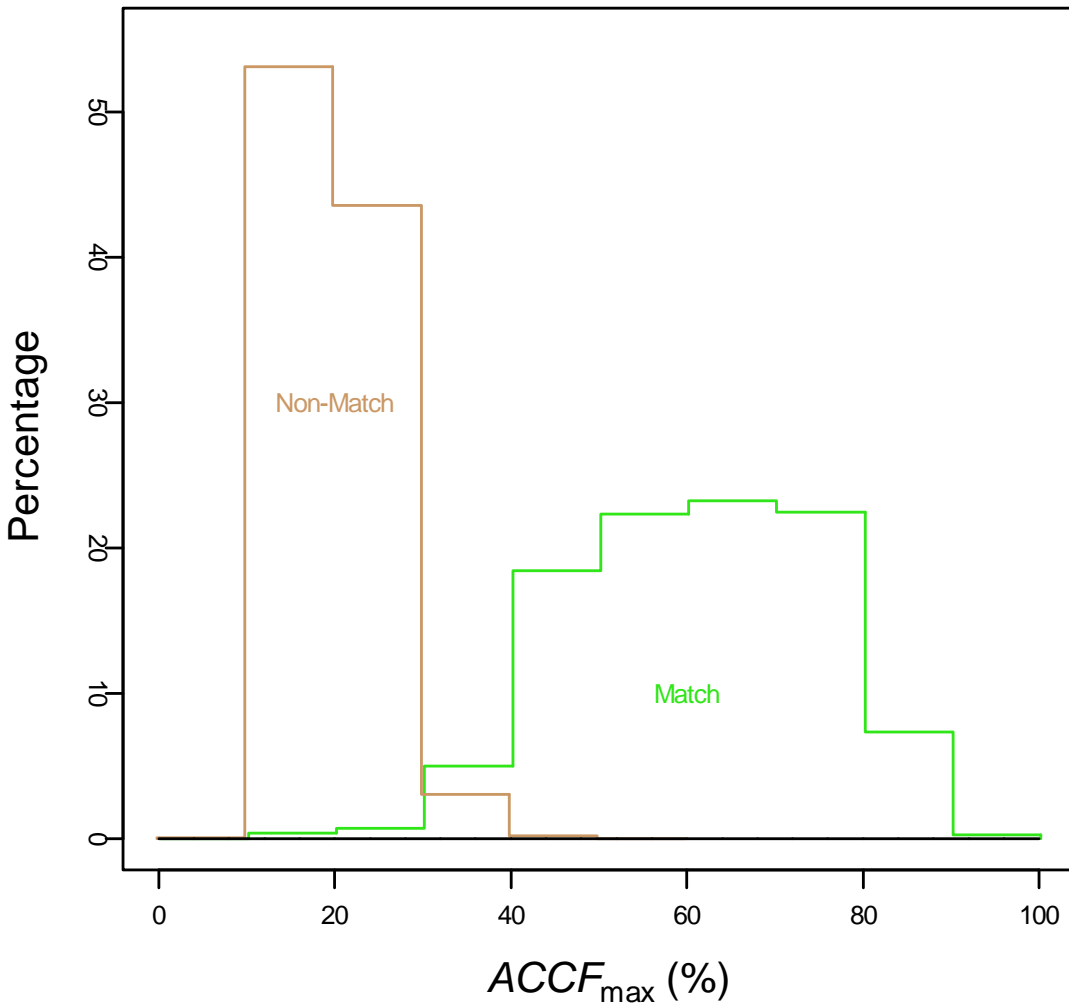
108

Figure 9-21. NBIDE Breech Face: The green lines depict a histogram of the matching scores, while the brown lines depict a histogram of the non-matching scores.

Figure 9-22 breaks down these results by reference gun, with the overlap metric p given for each grouping. For the NBIDE firing pin impressions, there was considerable variation between guns. By comparison, Fig. 9-22 shows better separation between matching and non-matching distributions for all the guns. However, it should be noted than even for some cases where there is little or no overlap, there may not be the wide separation between distributions that we find ideal; examples would be Sig Sauer 33 and S&W 401.

Table 9-8 contains the overlap metric statistics for the breech face data, ordered by performance of the gun as estimated by the overlap metric, and calculated from all pair-wise comparisons. How do the guns perform?
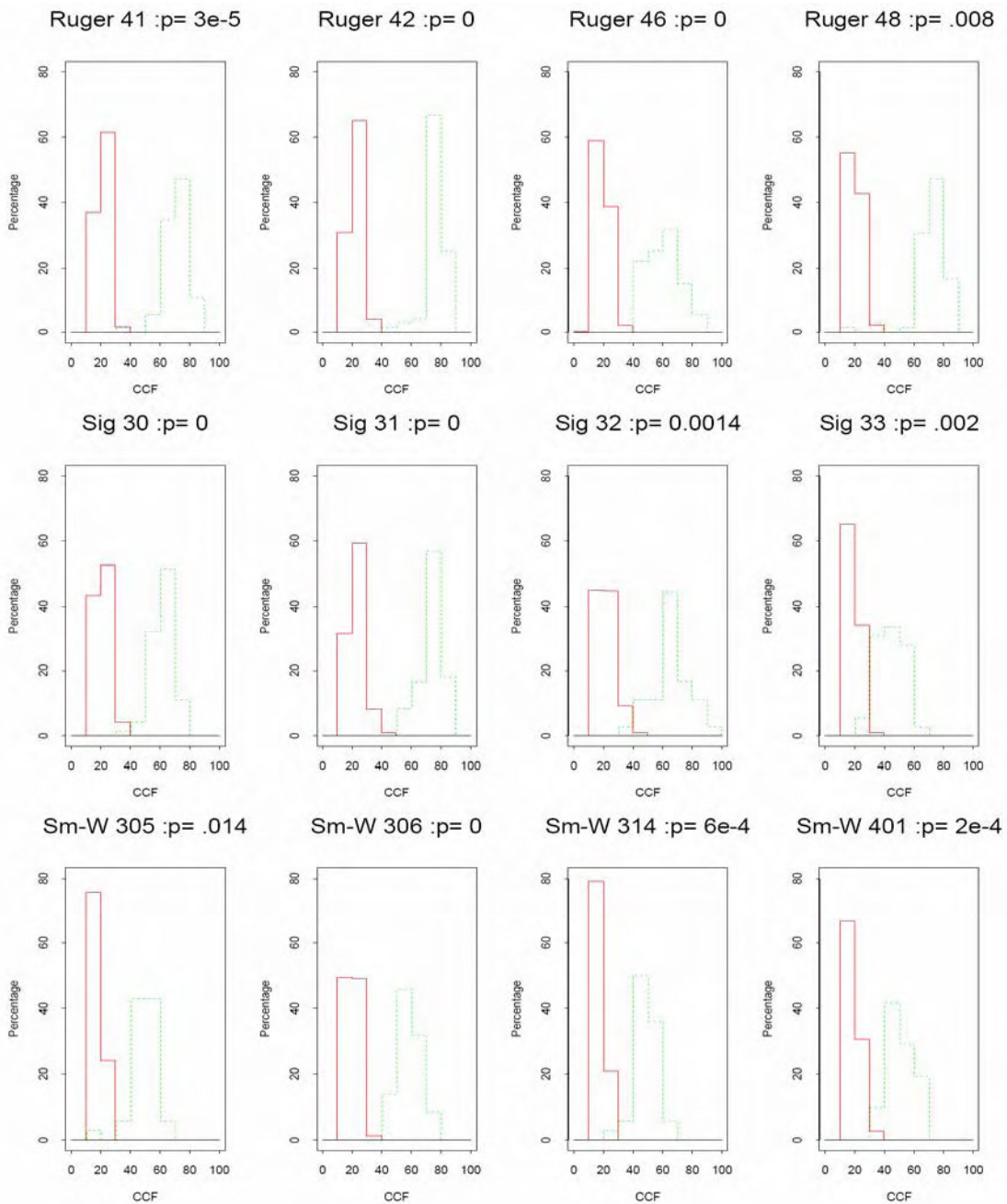
109

Figure 9-22. NBIDE Breech Face N-3D data: Above each of the individual plots is a heading with the ID of the reference gun and the overlap metric *p* derived for that group. In each plot, the green dotted line depicts a histogram of the matching scores, while the red solid line depicts a histogram of the non-matching scores. The symbol CCF stands for $ACCF_{max}$. The horizontal scale for $ACCF_{max}$ is the same as Fig. 9-6 (%).

110

Table 9-8. Overlap metric *p* for topographic signatures of the breech face impressions of 12 NBIDE guns

Gun ID	*p*
Ruger 42	0
Ruger 46	0
Sig Sauer 30	0
Sig Sauer 31	0
S&W 306	0
Ruger 41	0.00003
S&W 401	0.00017
S&W 314	0.00056
Sig Sauer 32	0.00142
Sig Sauer 33	0.00192
Ruger 48	0.00809
S&W 305	0.01353
Mean	0.00214

While some of the guns seem to have no overlap, note that as stated earlier, the estimation of very small probabilities by pair-wise comparisons can be problematic.

Figure 9-23 breaks down the data more finely by dividing the data into groups by reference casing. Each of the smaller plots depicts for each casing its correlations with the eight casings fired from the same gun (matches, lower triangles), and the correlations with the 99 casings fired from other guns (non-matches, upper triangles). The label above each plot indicates the gun ID, ammunition, and RR number of the reference casing.

From Fig. 9-23 it can be seen that for most of the guns the degree of overlap or separation between matching and non-matching casings varies significantly among casings fired from the same guns, which are shown on the same row in the diagram. For some casings (e.g. Ruger 42-1-28), there is considerable space between matching and non-matching scores. For others (e.g., those from Ruger 46), the matching and non-matching scores also do not overlap but may come close enough to cause overlapping observations for very large sample sizes.

The histogram in Fig. 9-24 charts the estimates of *p* of each of the 108 casings using the all pair-wise comparisons methods. How many are essentially zero and how high do the estimates go?
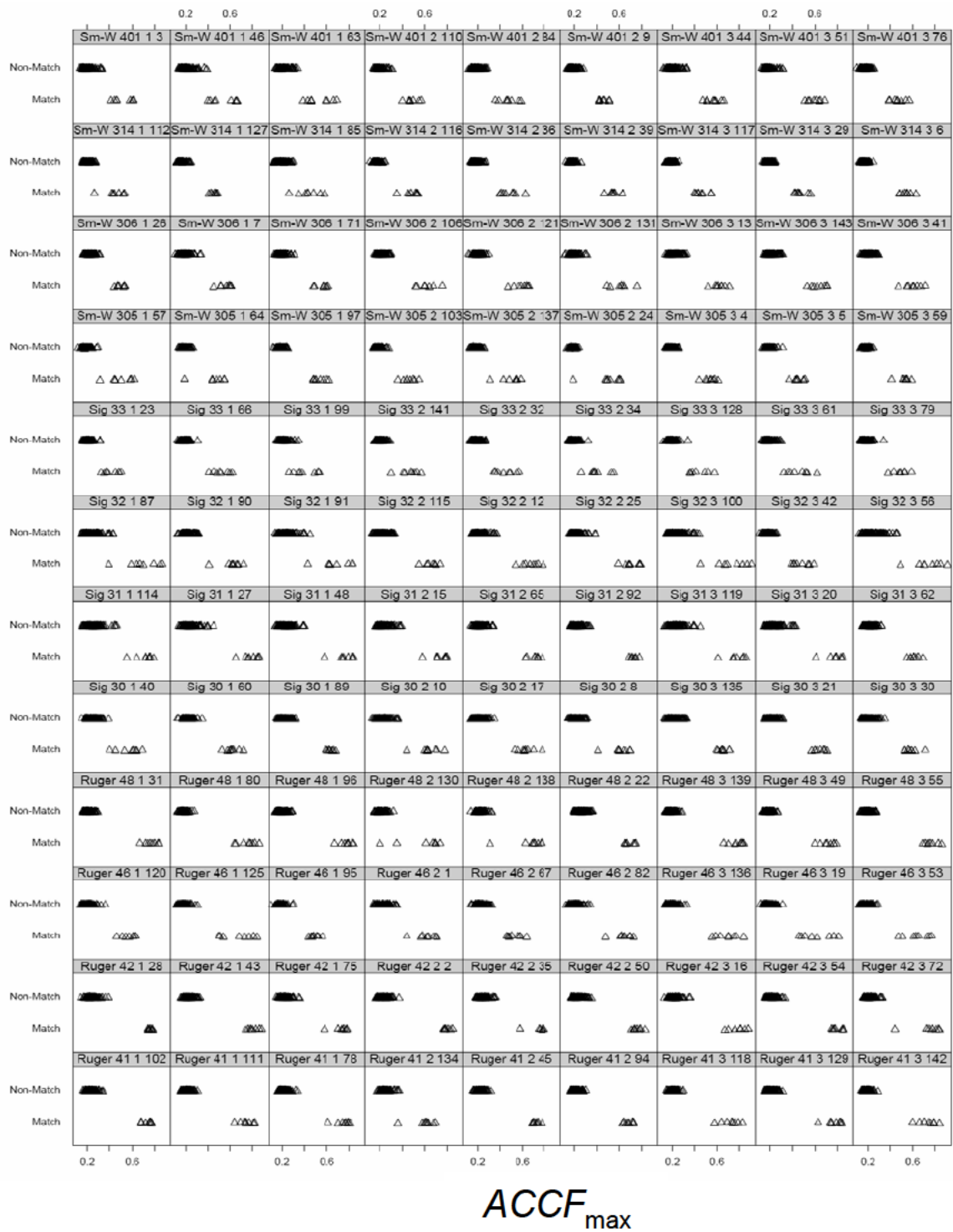
Figure 9-23. NBIDE breech face N-3D data: Correlations, for individual casings with the eight casings fired from the same gun (matches, lower triangles), and with the 99 casings fired from other guns (nonmatches, upper triangles). The label above each plot indicates the gun ID, ammunition (1-Win, 2-Rem, 3-PMC), and RR# of the reference casing.
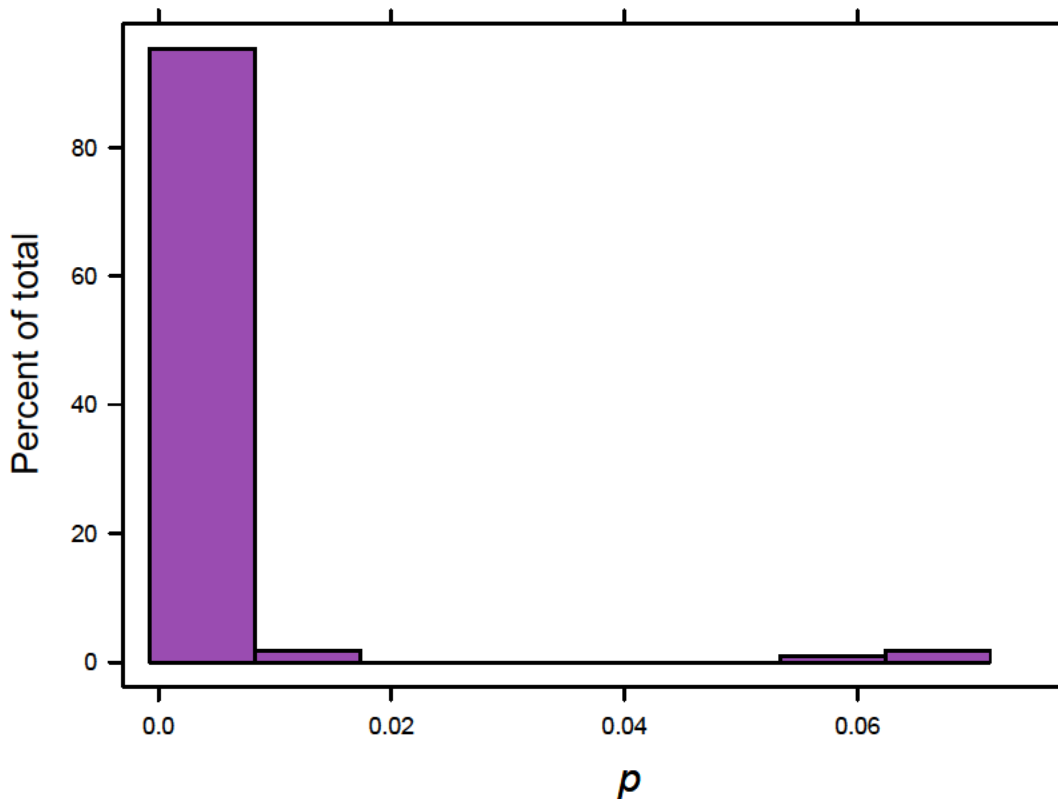
Figure 9-24. NBIDE breech face N-3D data: Histogram of the overlap metric p estimated for each of the 108 casings using pair-wise comparison methods. The mean value of p is 0.0022, and the median p is zero. In fact, 97 of the 108 estimates are exactly 0 (no overlap). The other eleven estimates range from 0.0013 to 0.0707.

The estimates of $p=0$ indicate no overlap between matching and non-matching distributions. It may be problematic to estimate very small probabilities when each casing has only eight observations in its matching sample and 99 observations in the non-matching samples. If a larger number of samples were available, there would likely be more overlaps and non-zero estimates.

Fitting continuous distributions to the samples involved may ameliorate the difficulty associated with low probability events and relatively small sample sizes. To assess this, we fit normal distributions to each of the matching and non-matching samples of correlation scores and derive the estimated means and variances of each sample of correlations. The effect of using different distributions is a topic for further investigation. Using this normal approximation, we obtain a new group of 108 estimates of p. Figure 9-25 contains a histogram of these estimates. Is it very different than that for the pair-wise comparison estimates shown in Fig. 9-24?
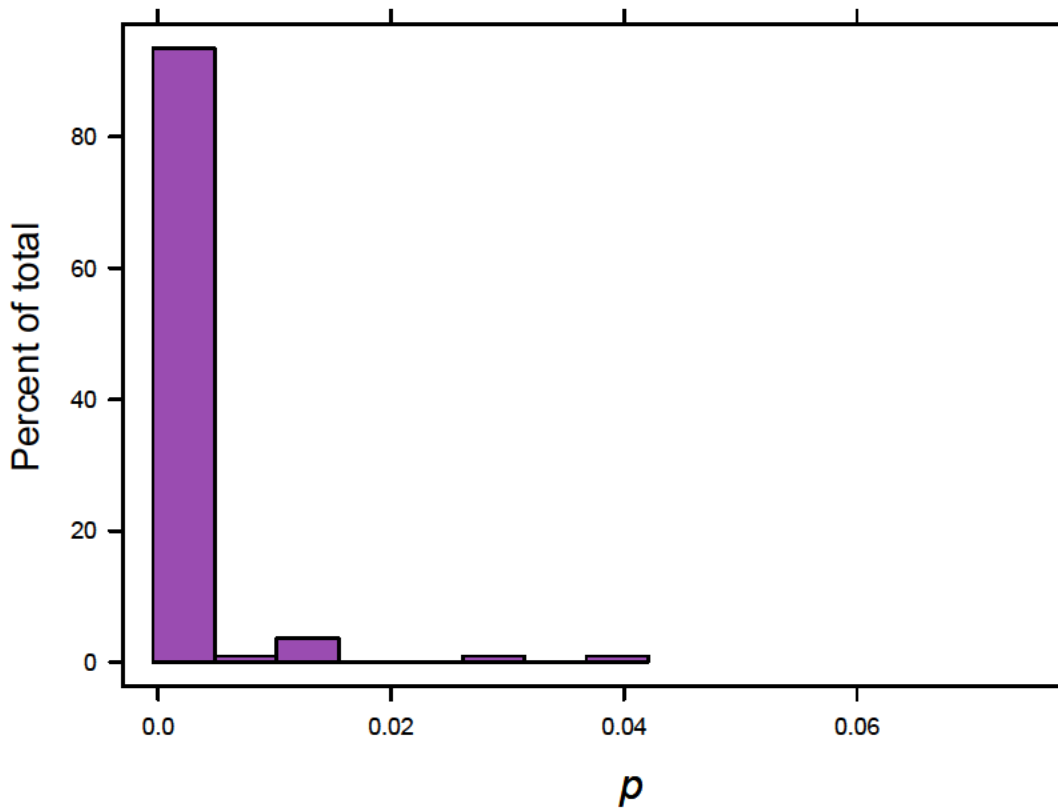
Figure 9-25. NBIDE breech face N-3D data: Histogram of the overlap metric p estimated for each of the 108 casings using normal distribution estimates. The mean value of p is 0.0016 and the median p is 2.8×10^{-7}.

A comparison of the histograms shown in Fig. 9-25 and Fig. 9-24 suggests that the two estimation methods produce very similar results; however, there are differences between the two that do not show up in the histograms. The primary difference is that the normal distribution method should enable a finer resolution in estimating very small probabilities. This difference is shown in the following listing, which breaks down the distribution of the estimates of p. It shows that 55 % of the estimates are less than or equal to 1×10^{-6}, while 94 % are less than or equal to 0.01.

Proportion of p estimates ≤	0	10^{-6}	10^{-5}	10^{-4}	10^{-3}	0.01	0.1
	0.11	0.55	0.63	0.72	0.81	0.94	1

The listing shows that most estimates of p for the individual casings are not exactly 0, but are close to 0. This is more useful since p needs to be extremely small for that casing not to call up many mismatches from an extremely large database. This is discussed further in Sec. 9.10 on probability models.

114