

ANSI/ASB Standard 062, First Edition
2021

Standard for Topography Comparison Software for Toolmark Analysis



ASB
ACADEMY
STANDARDS BOARD



Standard for Topography Comparison Software for Toolmark Analysis

ASB Approved May 2020

ANSI Approved October 2021



410 North 21st Street
Colorado Springs, CO 80904

This document may be downloaded from: www.asbstandardsboard.org

This document is provided by the AAFS Standards Board. You are permitted to print and download the document and extracts from the document for your own use, provided that:

- *you do not modify this document or its related graphics in any way;*
- *you do not use any illustrations or any graphics separately from any accompanying text; and,*
- *you include an acknowledgement alongside the copied material noting the AAFS Standards Board as the copyright holder and publisher.*

You expressly agree not to reproduce, duplicate, copy, sell, resell, or exploit for any commercial purposes, this document or any portion of it. You may create a hyperlink to www.asbstandardsboard.org to allow persons to download their individual, free copy of this document. Your hyperlink must not portray AAFS, the AAFS Standards Board, this document, our agents, associates and affiliates in an offensive manner, or be misleading or false. You may not use our trademarks as part of your link without our written agreement for you to do so.

The AAFS Standards Board retains the sole right to submit this document to any other forum for any purpose.

Certain commercial entities, equipment or materials may be identified in this document to describe a procedure or concept adequately. Such identification is not intended to imply recommendations or endorsement by the AAFS or the AAFS Standards Board, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Foreword

This standard was revised, prepared, and finalized by the Firearms and Toolmarks Consensus Body of the AAFS Standards Board. The draft of this standard was developed by the Firearms and Toolmarks Subcommittee of the Organization of Scientific Area Committees (OSAC) for Forensic Science, and determined this document is part of a series of documents jointly submitted to include:

- ANSI/ASB Standard 061, *Firearms and Toolmarks 3D Measurement Systems and Measurement Quality Control*
- ANSI/ASB Standard 062, *Standard for Topography Comparison Software for Toolmark Analysis*
- ANSI/ Standard 063, *Implementation of 3D Technologies in Forensic Firearm and Toolmark Laboratories*

The purpose of these standards is to ensure the production of reliable data and statistically based conclusions and is applicable to all forensic science service providers that provide conclusions regarding toolmark-related evidence. The documents establish performance expectations for new technologies while allowing legacy systems to coexist in the lab. The hardware document specifically refers to 3D scanning hardware and does not apply to legacy 2D type systems. The software document specifies three categories (levels) of software. Legacy systems are Category 0 whereas systems which provide validated statistical measures are Category 2. The implementation document outlines the necessary steps to ensure the proper implementation of 3D technologies.

Topography analysis and comparison software provides a means of evaluating the similarities and differences between high resolution surface topographies. The aim of such analysis is a quantifiable measure of toolmark topography comparisons to assist an examiner in reaching a conclusion or an assessment of the weight of the evidence regarding common origin.

The ANSI/ASB Standard 062, *Standard for Topography Comparison Software for Toolmark Analysis* specifies the minimum requirements for computer software intended to compare 2D and/or 3D digital representations of toolmarks. It covers necessary conditions for consistent and interpretable comparisons. Software that complies with the specifications of this document can be used for topography analysis and comparison.

The AAFS Standards Board (ASB) is an ANSI-accredited Standards Developing Organization with the purpose of providing accessible, high quality science-based consensus forensic standards. The ASB is a wholly owned subsidiary of the American Academy of Forensic Sciences (AAFS), established in 2015 and accredited by the American National Standards Institute (ANSI) in 2016. The ASB consists of Consensus Bodies (CB), which are open to all materially interested and affected individuals, companies, and organizations; a Board of Directors; and Staff.

All hyperlinks and web addresses shown in this document are current as of the publication date of this standard.

Keywords: *3D, topography, measurement, algorithm, quality assurance, firearms, toolmarks, identification.*

Table of Contents

1	Scope	1
2	Normative References	1
3	Terms and Definitions	1
4	Requirements	4
4.1	Data and Format Requirements (Mandatory)	4
4.2	Software Categories (Mandatory)	5
4.3	Software Validation (Mandatory)	7
4.4.	User Training (Mandatory)	8
4.5	Statistical Comparison Models and Interpretation (Mandatory)	8
4.6	Criteria for Interpretation (Mandatory)	9
	Annex A (informative) Sample Development Validation Study of Category 2 Software	10

Standard for Topography Comparison Software for Toolmark Analysis

1 Scope

This document specifies the minimum requirements for computer software intended to compare 2D and/or 3D digital representations of toolmarks. It covers necessary conditions for consistent and interpretable comparisons.

2 Normative References

The following references are documents that are indispensable for the application of the standard. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ANSI/ASB Standard 061, *Firearms and Toolmarks 3D Measurement Systems and Measurement Quality Control*, First Edition, 2021^a

ANSI/ASB Standard 063, *Implementation of 3D Technologies in Forensic Firearm and Toolmark Laboratories*, First Edition, 2021^a

ISO 25178-72, *Geometrical product specifications (GPS)—Surface texture: Areal—Part 72: XML file format x3p*^b

3 Terms and Definitions

3.1

comparison algorithm

A series of computational steps which seeks to assess both the level of geometric similarity (similarity of toolmarks) and the degree of certainty that the observed similarity results from a common origin. A comparison algorithm makes use of a scoring function or similarity score.

3.2

comparison software

Software that implements a comparison algorithm and may also include database, search, and visualization functionality.

3.3

comma-separated value file

CSV

A simple file format for tabular data where individual values are separated by a comma (or other designated delimiter).

3.4

false positive rate

Within this document the false positive rate is defined as the number (or percentage) of Known Non-Matches which are incorrectly determined to be an Identification.

^a Available from: <https://www.asbstandardsboard.org/published-documents/>

^b Available from: <https://www.iso.org/standard/62310.html>

3.5**frequency**

The rate at which an event occurs.

3.6**heightmap**

A three-dimensional topographic data set consisting of surface points (x, y, z) where each dimension is a coordinate measured in standard units (e.g., micrometers). See 3.14 topography data.

3.7**interpretable scoring function**

A Category 1 scoring function. The score is explainable and has quantifiable meaning. See 4.2.5.

3.8**noncompliant 3D measurement**

A 3D Topographic measurement made using hardware that does not comply with ANSI/ASB Standard 061, *Firearm and Toolmarks 3D Measurement Systems and Measurement Quality Control*. The accuracy and precision of these measurements may be uncertain and untraceable.

3.9**positive predictive value**

Within this document the positive predictive value is defined as the fraction of identifications that correspond to true known matches, $TP/(TP+FP)$ where TP (True Positives) is the number of correctly identified Known Matches and FP (False Positives) is the number of Known Non-Matches incorrectly determined to be an Identification.

3.10**probability**

A quantified measure between zero and one indicating how probable or likely it is that an event will or has occurred. In the frequentist interpretation, probability is based on the rates at which events occur. In the Bayesian interpretation, probability reflects a degree of belief. On this scale, zero indicates impossibility and one indicates absolute certainty.

3.11**rank-score only scoring function**

A Category 0 scoring function. See 4.2.4. A rank-score only scoring function is either non-quantified, inconsistent, or non-explainable.

3.12**recall rate**

The recall rate is also known as true positive rate. It is defined as the number (or percentage) of toolmark specimens containing a Known-Match in the test set for which a Known-Match is correctly determined.

3.13**scoring function (or similarity score)**

The mathematical core of a comparison algorithm which takes two input topographies and generates one or more outputs that quantify the comparison based on geometric similarity. Scoring functions are either rank-score only, interpretable, or statistically validated.

3.14**specificity (also called the true negative rate)**

Measures the proportion of actual negatives that are correctly identified as such.

3.15**statistically validated scoring function**

A Category 2 scoring function. See 4.2.6.

3.16**topography data**

A measurement of an object's surface geometry. Topographic data may be one, two, or three dimensional.

3.17**topography data: 1D data**

1D topography data is also known as a Linear Profile. A linear profile with n points shall be represented as a function of a single coordinate where at each point along a single axis (x_i) (where $i=1,\dots,n$) there is a measured height (z_i). Both x_i and z_i are measured in standard units (e.g., micrometers). An example of a linear profile is a cross-section through a striated toolmark (e.g., bullet land area); where x_i is a spatial position measured in micrometers and z_i is the corresponding height of the striation profile measured in micrometers.

3.18**topography data: 2D data**

2D topography data is also known as a Planar Image. An n -by- m planar image I shall be represented as a function of two coordinates where at each point (x_i, y_j) (where $i=1,\dots,n; j=1,\dots,m$) there is a measured surface color or intensity denoted $I(x_i, y_j)$. An example of 2D topographic data is an image taken through a comparison microscope; where each point $I(x_i, y_j)$ is the RGB (red, green, blue) color value measured at the specified (x_i, y_j) position. The measured color or intensity is a function of the surface geometry and the environmental conditions (e.g., light position). Although reference scales may be included in the collected image, the points (x_i, y_j) may or may not be measured in standard units (e.g., micrometers).

3.19**topography data: 3D data**

3D topography data is also known as a Heightmap. An n -by- m heightmap H shall be represented as a function of two coordinates where at each point (x_i, y_j) (where $i=1,\dots,n; j=1,\dots,m$) there is a measured surface height $z_{ij} = H(x_i, y_j)$. All three coordinates x_i, y_j , and z_{ij} are measured in standard units (e.g., micrometers). The surface H is a 1-to-1 representation of the actual object. An example of 3D topographic data is a primer surface measured using a confocal microscope; where each point $H(x_i, y_j)$ represents the surface height (in micrometers) measured at the specified (x_i, y_j) position. 3D data captured on hardware compliant with ANSI/ASB Standard 061, *Firearm and Toolmarks 3D Measurement Systems and Measurement Quality Control* can be used in virtual comparison microscopy.

3.20**XML 3D Surface Profile****X3P**

X3P is an open file format for the exchange of three dimensional surface topography data in standard units. Details are specified in ISO 25178-72.

4 Requirements

4.1 Data and Format Requirements (Mandatory)

4.1.1 General Information

4.1.1.1 Comparison software shall accept as input one or more of the following topographies: a Linear Profile (1D), a Planar Image (2D), or a Heightmap (3D).

4.1.1.2 Comparison software should assist the forensic examiner in a number of ways. The comparison algorithm should be able to compare one object to another (e.g., individual comparison), one object to a set of many objects (e.g., database search), or a set of objects to another set of objects (e.g., expanded database search). The resulting similarity scores shall be interpreted in the manner described for the corresponding software Category described in 4.2.

4.1.2 1D Data

4.1.2.1 Data Format: For the representation and exchange of 1D Topographic Data, comparison software shall support either X3P or CSV.

4.1.2.2 Data Source: Linear profiles are typically obtained from a 1D profilometer or by processing (e.g., cross-section) 3D topographic data.

4.1.2.3 Data Representation: 1D Topographic Data shall be measured in standard units of length (e.g., micrometers).

4.1.3 2D Data

4.1.3.1 Data Format: For the representation and exchange of 2D Topographic Data, comparison software shall support any commonly established lossless image file format (e.g., TIFF, PNG).

4.1.3.2 Data Source: Planar Images are typically acquired using a traditional digital camera attached to a comparison microscope or similar imaging optics.

4.1.3.3 Data Representation: 2D Topographic Data shall be measured on a grid of fixed dimension sample points. The interpoint spacing may or may not be measured in standard units of length (e.g., micrometers).

4.1.4 3D Data

4.1.4.1 Data Format: For the representation and exchange of 3D Topographic Data, comparison software shall support the X3P file format. The X3P data shall follow the specifications of ISO 25178-72 with the following addition. Record 2 (Metadata), which is an optional data record in the ISO document, shall be a required record for toolmark analysis. Three fields (Version, Probing Type: Identification, and Comment) may be left blank, but all Record 2 fields specified in the ISO document shall be included in the X3P file. It is recommended that data from cartridge cases and bullets also include an optional Record X which specifies toolmark metadata. If included, Record X shall have a <VendorSpecificID> field equal to <http://www.openfmc.org/firearm> and shall include the fields specified by the OpenFMC group. The X3P may also include vendor specific records.

4.1.4.2 Data Source: Heightmaps shall be measured on imaging hardware compliant with ANSI/ASB Standard 061, *Standard for 3D Measurement Systems and Measurement Quality Control for Firearm and Toolmark Analysis* and with procedures compliant with ANSI/ASB Standard 063, *Standard for Implementation of 3D Technologies in Forensic Laboratories for Firearm and Toolmark Analysis*. Measurements obtained using hardware that does not comply with ANSI/ASB Standard 061, *Standard for 3D Measurement Systems and Measurement Quality Control for Firearm and Toolmark Analysis* shall be considered Noncompliant 3D Measurements and the value of their interpretation may be limited. Noncompliant 3D measurements shall only be used with Category 0 software. Noncompliant 3D measurements shall not be used with Category 1 or Category 2 software. Noncompliant 3D measurements shall not be exchanged between labs for the purposes of firearms and toolmark examination or for casework comparisons.

4.1.4.3 Data Representation: 3D Topography Data shall be measured on a grid of fixed dimension sample points. The interpoint spacing and heights (x , y , and z) shall all be measured in standard units.

4.2 Software Categories (Mandatory)

4.2.1 General Information

4.2.1.1 Comparison Software seeks to provide, in an explainable manner, a numeric measure (e.g., degree of certainty) quantifying the geometric support for common origin.

4.2.1.2 Comparison Software (and its associated Scoring Function) shall be evaluated before its use by forensic science service providers. Software evaluation shall advance through a series of Categories. When a software satisfies the requirements of a category it can be used for the purposes and in the manner described for the corresponding category. Software shall not be used for the purposes or in the manner of categories whose requirements have not been satisfied.

4.2.1.3 In accordance with ANSI/ASB Standard 063, *Standard for Implementation of 3D Technologies in Forensic Laboratories for Firearm and Toolmark Analysis*, all software, regardless of Category, shall be validated prior to use. Therefore, developmental and deployment validation studies must be completed.

4.2.2 Category 0: Rank-Scores Only Scoring Function

4.2.2.1 Criteria for Category 0: Any comparison software may attain a Category 0 designation by completing developmental and deployment validation studies (see 4.3). Any measurement hardware may be used with Category 0 software. That is, measurement hardware does not need to comply with ANSI/ASB Standard 061, *Standard for 3D Measurement Systems and Measurement Quality Control for Firearm and Toolmark Analysis* to be used with Category 0 software. Category 0 software is considered Rank-Scores Only.

4.2.2.2 Use of Category 0 Software: Rank-Scores Only software shall only be used as part of a hit finding or sorting process. The numeric score (e.g., similarity score or any other numeric measure of match quality) shall not be mentioned in reports or discussion. That is, the scores shall not be used in any way secondary to ranking. For example, one shall not state that the pair of cartridge cases has a comparison score of X, which is very high. One may only state that the pair of cartridge cases ranked in the top Y of the database search (where Y may be a number, e.g., 10, 20). If asked about the significance of this statement the only acceptable response shall be that there is no

statistical confidence established for numeric results for a rank-scores only, non-statistically validated scoring function.

4.2.3 Category 1: Interpretable Scoring Function

4.2.3.1 Criteria for Category 1: Category 1 software shall only utilize scan data collected using hardware compliant with ANSI/ASB Standard 061, *Standard for 3D Measurement Systems and Measurement Quality Control for Firearm and Toolmark Analysis*. To satisfy Category 1 and be considered an Interpretable Scoring Function, the scoring function and software shall complete developmental and deployment validation studies (see 4.3). In addition, for a single sample-to-sample comparison an Interpretable Scoring Function shall satisfy the following criteria.

4.2.3.1.1 It shall output a single quantified numeric value.

4.2.3.1.2 It shall be reported on a consistent scale. A score of X obtained in two different searches shall mean the same thing and shall carry the same degree of confidence, probability, statistical weight, or likelihood. For example, if cartridge cases C1 and C2 have a similarity score of X and if cartridge cases C3 and C4 have a similarity score of X then the confidence that C1 and C2 have common origin shall be equal to the confidence that C3 and C4 have common origin.

4.2.3.1.3 It shall be explainable, such that a firearms examiner can describe the general principles on which it works.

4.2.3.2 Scoring functions not meeting the Interpretable Scoring Function criteria shall be considered Rank-Score Only Scoring Functions (Category 0). These Rank-Score Only Scoring Functions are either non-quantified, inconsistent, or non-explainable.

4.2.3.3 Use of Category 1 Software: Comparison Software utilizing an Interpretable Scoring function may be used for all uses indicated for Category 0. In addition, the numeric score (e.g., similarity score or any other numeric measure of match quality) may be described in reports or discussion. For example, one may state that a pair of cartridge cases has a comparison score of X, and one may describe the frequency of seeing that score in their experience. If asked about the significance of this statement for software at Category 1, the only acceptable response shall be that there is no statistical confidence established for any results for a non-statistically validated scoring function. Interpretable Scoring functions can be used to develop a statistical model of confidence (Category 2).

4.2.4 Category 2: Statistically Validated Scoring Function

4.2.4.1 Criteria for Category 2: To satisfy Category 2 and be considered a Statistically Validated Scoring Function, software shall meet the criteria for an Interpretable Scoring function and have completed extensive developmental and deployment validation studies (Section 4.3) to demonstrate the statistical performance on a large representative test set. In addition, Category 2 software shall only utilize scan data collected using hardware compliant with ANSI/ASB Standard 061, *Standard for 3D Measurement Systems and Measurement Quality Control for Firearm and Toolmark Analysis*.

4.2.4.2 Use of Category 2 Software: Comparison Software completing the validation requirements of Category 2 may be used for all uses indicated for all Categories. Software at Category 2 has established statistical confidence for match results. The statistical significance or support of the

numeric score (e.g., similarity score or any other numeric measure of match quality) may be described in reports or discussion. One may present statements such as, “the pair of cartridge cases has a similarity score of X. At a threshold of X, the false match probability is Y”. If asked about the significance of this statement for software at Category 2, an acceptable answer shall cite relevant validation studies and their relevance to the question at hand. For example, “A 2017 study by W included V test fires from firearms with relevant manufacturing techniques, ammunition types, and materials. The study demonstrated that a score of X indicates a false match probability of Y.”

4.3 Software Validation (Mandatory)

4.3.1 General

In accordance with ANSI/ Standard 063, *Implementation of 3D Technologies in Forensic Firearm and Toolmark Laboratories*, all toolmark analysis software, regardless of Category and regardless of toolmark type being examined, shall be validated prior to use.

The developmental validation shall be conducted by an organization with appropriate knowledge and/or expertise. The deployment validation is a smaller follow-on evaluation and shall be conducted by a lab prior to their implementation of a new technology. A deployment validation shall only be completed after a successful development validation. Validation studies shall be documented as described in ANSI/ASB Standard 063, *Implementation of 3D Technologies in Forensic Firearm and Toolmark Laboratories*.

Personnel responsible for deployment validation shall at a minimum have a bachelor's degree or equivalent degree with a natural science-based or applied science-based major field of study. To the extent possible, a single individual should conduct all components of the deployment validation to minimize sources of operator uncertainty.

Comparison and Analysis software is not subject to traditional Ongoing Performance Checks as a single piece/version of software does not fall out of calibration. Software upgrades that affect scoring function functionality (such as a major version upgrade) shall require deployment validation.

4.3.2 Software at Categories 0 and 1: User certification shall be utilized as described in 4.4. In addition, development and deployment validation must demonstrate:

- a) the limitations of the procedure;
- b) the conditions under which reliable results can be obtained;
- c) critical aspects of the procedure that shall be controlled and monitored;
- d) the ability of the procedure to meet the needs of the given application.

4.3.3 Software at Category 2: User certification shall be utilized as described in 4.4. The developmental and deployment validation for Category 2 software shall include a large statistical study aimed at demonstrating the statistical significance of the scoring function.

4.3.3.1 Developmental Validation: Developmental validation studies shall be robust and report on the tools for which the study was explicitly designed and on which the results are explicitly applicable (e.g., for firearms toolmarks report on the firearm manufacturing techniques,

ammunition calibers, ammunition types, ammunition materials, etc.). To guard against overfitting and improve the likelihood that performance generalizes across actual casework data, the validation set shall include both a sufficiently broad selection of makes and models as well as a sufficiently broad selection of substrate types (e.g., for firearms toolmarks substrate types include ammunition materials). The validation set should therefore reflect the range of toolmark types expected to be seen in casework. The larger and more inclusive the validation set the stronger the validation result. Studies shall be sufficiently large to determine statistical performance to an appropriate degree of certainty. See Annex A1 for an example study set. All developmental validation studies shall estimate error rates and uncertainties on these error rates using established statistical methods.

The developmental validation report shall include overall performance measures indicating at least the Recall Rate, False Positive Rate, Specificity, and Positive Predictive Value.

4.3.3.2 Deployment Validation: The deployment validation shall be more modest in size than the developmental validation. For example, a deployment validation may involve running a small sample of proficiency tests. In completion of the validation study the following aspects shall be documented:

- a) the limitations of the validation study;
- b) the conditions under which reliable results can be obtained;
- c) critical aspects of the procedure that shall be controlled and monitored;
- d) the ability of the procedure to meet the needs of the given application.

4.3.3.3 Both development and deployment validation studies shall use test sets that are independent from the initial sets used to build the scoring function or statistical model.

4.3.3.4 Future developments in manufacturing techniques may require additional studies to augment the results of the development and deployment validation studies.

4.4 User Training (Mandatory)

A training process or program shall be established by the authors and/or distributors of all comparison software. The training program shall include an assessment module. The laboratory management shall ensure the initial and continued competence of all users of comparison software within the laboratory. Details of this process are provided in ANSI/ASB Standard 063, *Standard for Implementation of 3D Technologies in Forensic Firearm and Toolmark Laboratories*.

4.5 Statistical Comparison Models and Interpretation (Mandatory)

4.5.1 Statistical comparison models shall output a statistically grounded metric indicative of whether or not two toolmarks have a common origin or support for common origin.

4.5.2 Statistical comparison models shall be constructed only from Interpretable Scoring Functions satisfying Category 1.

4.5.3 Statistical comparison models may employ either a generative or an empirical approach.

4.5.3.1 A generative model estimates the probability of obtaining a similarity score by simulating the toolmark generation and subsequent matching process.

4.5.3.2 An empirical model uses a representative set of relevant population data to estimate the confidence, probability, or frequency of obtaining specific similarity score(s) for a known match or a known non-match. With respect to firearm forensics, the population data shall include a sufficiently broad selection of firearm makes and models as well as a sufficiently broad selection of ammunition types, dependent on the case at hand. An empirical model measures the observed similarity scores over the dataset's known matches and known non-matches. Several statistical methods may be used to compute match or non-match confidence, probability, or frequency.

4.5.4 Statistical comparison models shall be validated prior to their use.

4.6 Criteria for Interpretation (Mandatory)

4.6.1 Scoring functions and statistical comparison models at Category 2 may be used by an examiner to provide a statistical assessment regarding common origin or weight of evidence.

4.6.2 Conclusions regarding common origin or statements of weight of evidence shall follow the standard operating procedure of the laboratory. Interpretable Scoring Functions and statistical comparison models provide one way for stating the degree of certainty or weight of evidence for common origin and for supporting the finding based on the underlying comparative model.

Annex A **(informative)**

Sample Validation Study Detail

Sample Development Validation Study of Category 2 Software

A 2017 test set for 9mm Luger firearms would include approximately 4800 test fires from at least sixty (60) different firearm manufacturers and two hundred firearm models. The set would include all commonly used manufacturing mark types (e.g. milled, broached, EDM, and filed). The set would include multiple test fires per firearm involving at least six (6) different brands of ammunition. Bullet samples should contain jacket materials that have been forged or chemically bonded that contain lead, copper, steel, and brass metals. Cartridge case sets should contain cartridge cases with both brass and nickel primers. Cartridge case sets should contain cartridge cases with brass, steel, and nickel cases. Different calibers may or may not require different test sets.



ASB
ACADEMY
STANDARDS BOARD

Academy Standards Board
410 North 21st Street
Colorado Springs, CO 80904

www.asbstandardsboard.org