Spatial Data Accuracy As Defined by the GSDM

Earl F. Burkholder

ABSTRACT: The concept of a global spatial data model (GSDM) which combines horizontal and vertical spatial data into a single three-dimensional database is defined and described in Burkholder (1997a). This paper describes how the GSDM accommodates spatial data accuracy, providing specific equations for computing various accuracies. A careful distinction is made between describing GSDM features and issues of implementation. As a consistent set of equations and relationships, the GSDM can be used immediately by anyone. Greater benefits will, however, accrue as the concepts gain acceptance and the GSDM procedures are standardized and adopted by the spatial data user community at large (implementation).

Introduction[1]

The topic of accuracy is central to efficient use of spatial data. Previously, an answer to the question "Accuracy with respect to what?" was largely implied by the context in which the question was asked. However, with the evolution of modern measurement technologies, the use of spatial data in geographic information systems, and widespread use of sophisticated data processing capability, the issue of spatial data accuracy is increasingly relevant and the question begs an answer. Traditionally, spatial data accuracy has been considered in terms of 2-dimensional horizontal data (latitude/longitude) and separately in terms of 1-dimensional vertical data (elevations). But, spatial measurements are made in a 3-dimensional environment and, except for human perception and use, such separation of horizontal and vertical concepts is largely unwarranted. Recently, the concept of a global spatial data model (GSDM) which combines horizontal and vertical spatial data into a single three-dimensional (3-D) database has been defined (Burkholder 1997a). When considered within the framework of the GSDM, concepts of spatial data accuracy can be viewed with greater clarity.

The global spatial data model (GSDM) is defined in a report, "Definition of a Three-Dimensional Spatial Data Model for Southeastern Wisconsin" (Burkholder 1997a) and described in more detail in Burkholder (1997b, 1997c, 1998a, and 1998b). The GSDM consists of a functional model of geometrical equations and a stochastic model which defines error propagation procedures. The functional model equations are based on the Earth-centered, Earth-fixed (ECEF) rectangular geocentric coordinate system defined by the Defense Mapping Agency (DMA 1991), and the stochastic model utilizes formal error propagation techniques as described in Chapter 4 of Mikhail (1976) and Chapter 5 of Wolf and Ghilani (1997).

Earl F. Burkholder, PS, PE, is President of Global COGO, Inc., Las Cruces, new Mexico 88003. E-mail: <GlobalCOGO@zianet.com>.

Used with a BURKORD™ 3-D database which stores the geocentric coordinates of each point, the covariance matrix for each point, and correlations between points, the GSDM defines simple, rigorous, efficient procedures for storing, manipulating, and using geospatial data. Among others, an important feature of the GSDM is that it accommodates any level of accuracy and provides proven statistical tools which can be used to answer the question, "Accuracy with respect to what?" That means a user is able to make better decisions by knowing what assumptions (mathematical conditions) are associated with the use of such terms as "network" accuracy and "local" accuracy.

The GSDM Covariance Matrices

The functional component of the GSDM consists of geometrical equations which are used to manipulate X/Y/Z geocentric coordinates defining the spatial position of each point. The stochastic component of the GSDM is an application of the laws of variance/covariance error propagation and utilizes the following matrix formulation:

$$\Sigma_{YY}=J_{YX}\Sigma_{XX}J'_{YX} \qquad (1)$$

where:

Σ_{YY} = covariance matrix of computed result

[1] Editors Note: The use of trademarked terms in this paper does not constitute an endorsement of the product or trademark.

Σ_{XX} = covariance matrix of variables used in computation; and

J_{YX} = Jacobian matrix of partial derivatives of the result with respect to the variables.

The GSDM uses two covariance matrices for each point; the geocentric covariance matrix and the local covariance matrix. The following symbols and matrices are used in the stochastic model:

$\sigma_X^2\ \sigma_Y^2\ \sigma_Z^2$ = variances of geocentric coordinates for a point;

$\sigma_{XY}\ \sigma_{XZ}\ \sigma_{YZ}$ = covariances of geocentric coordinates for a point;

$\sigma_e^2\ \sigma_n^2\ \sigma_u^2$ = variances of a point in the local reference frame;

$\sigma_{en}\ \sigma_{eu}\ \sigma_{nu}$ = covariances of a point in the local reference frame;

$\sigma_{\Delta X}^2\ \sigma_{\Delta Y}^2\ \sigma_{\Delta Z}^2$ = variances of geocentric coordi nate differences;

$\sigma_{\Delta X\Delta Y}\ \sigma_{\Delta X\Delta Z}\ \sigma_{\Delta Y\Delta Z}$ = covariances of geocentric coordinate differences;

$\sigma_{\Delta e}^2\ \sigma_{\Delta n}^2\ \sigma_{\Delta u}^2$ = variances of coordinate differences in local frame;

$\sigma_{\Delta e\Delta n}\ \sigma_{\Delta e\Delta u}\ \sigma_{\Delta n\Delta u}$ = covariances of coordinate differences in local frame;

$\sigma_s^2\ \sigma_\alpha^2$ = variances of local horizontal distance and azimuth;

$\sigma_{s\alpha}$ = covariance of local horizontal distance with azimuth;

$\sigma_{X1X2}\ \sigma_{Y1Y2}$ = elements of Point 1 - Point 2 correlation matrix.

Geocentric Covariance Matrix

$$\Sigma_{X/Y/Z} = \begin{bmatrix} \sigma_X^2 & \sigma_{X/Y} & \sigma_{X/Z} \\ \sigma_{X/Y} & \sigma_Y^2 & \sigma_{Y/Z} \\ \sigma_{X/Z} & \sigma_{Y/Z} & \sigma_Z^2 \end{bmatrix} \quad (2)$$

Local Covariance Matrix

$$\Sigma_{e/n/u} = \begin{bmatrix} \sigma_e^2 & \sigma_{e/n} & \sigma_{e/u} \\ \sigma_{e/n} & \sigma_n^2 & \sigma_{n/u} \\ \sigma_{e/u} & \sigma_{n/u} & \sigma_u^2 \end{bmatrix} \quad (3)$$

Notes about the geocentric and local covariance matrices:

- Each covariance matrix is symmetric 3 x 3. Six numbers are required to store upper (or lower) triangular values.
- Units in each covariance matrix are length squared. The off-diagonal elements represent correlations, diagonal elements are called variances, and standard deviations are computed as the square root of the variances.
- Each covariance matrix (with its unique orientation) represents the accuracy of a point with respect to a defined reference frame (or to whatever control is held fixed by the user) and is designated datum accuracy.

The local covariance matrix and the geocentric covariance matrix are related to each other mathematically by a rotation matrix for the latitude/ longitude position of a point computed from its X/Y/Z coordinates (Burkholder 1993). The rotation matrix is:

$$R = \begin{bmatrix} -\sin\lambda & \cos\lambda & 0 \\ -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi \end{bmatrix} \quad (4)$$

and the relationship between the covariance matrices is:

$$\Sigma_{e/n/u} = R\,\Sigma_{X/Y/Z}\,R' \quad (5)$$

$$\Sigma_{X/Y/Z} = R'\,\Sigma_{e/n/u}\,R \quad (6)$$

With regard to the rotation matrix in equation (4), longitude is counted 0° to 360° east from the Greenwich Meridian, west longitude is a negative value, and latitude is counted positive north of the equator, negative south of the equator.

The GSDM 3-D Inverse

Given that point 1 is defined by $X_1/Y_1/Z_1$ and point 2 by $X_2/Y_2/Z_2$, the matrix formulations of the 3-D geocentric coordinate inverse and covariance error propagation are:

$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ X_2 \\ Y_2 \\ Z_2 \end{bmatrix} \quad (7)$$

$$\Sigma_\Delta = J_1\,\Sigma_{1\text{-}2}\,J_1' \quad (8)$$

The Jacobian matrix in equation (7) and the general covariance error propagation procedure in equation (8) are used to find the overall geocentric inverse covariance matrix:

$$
\Sigma_\Delta = \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix}
\begin{bmatrix} \sigma^2_{X_1} & \sigma_{X_1Y_1} & \sigma_{X_1Z_1} \\ \sigma_{X_1Y_1} & \sigma^2_{Y_1} & \sigma_{Y_1Z_1} \\ \sigma_{X_1Z_1} & \sigma_{Y_1Z_1} & \sigma^2_{Z_1} \end{bmatrix} &
\begin{bmatrix} \sigma_{X_1X_2} & \sigma_{X_1Y_2} & \sigma_{X_1Z_2} \\ \sigma_{Y_1X_2} & \sigma_{Y_1Y_2} & \sigma_{Y_1Z_2} \\ \sigma_{Z_1X_2} & \sigma_{Z_1Y_2} & \sigma_{Z_1Z_2} \end{bmatrix} \\
\begin{bmatrix} \sigma_{X_1X_2} & \sigma_{Y_1X_2} & \sigma_{Z_1X_2} \\ \sigma_{X_1Y_2} & \sigma_{Y_1Y_2} & \sigma_{Z_1Y_2} \\ \sigma_{X_1Z_2} & \sigma_{Y_1Z_2} & \sigma_{Z_1Z_2} \end{bmatrix} &
\begin{bmatrix} \sigma^2_{X_2} & \sigma_{X_2Y_2} & \sigma_{X_2Z_2} \\ \sigma_{Y_2X_2} & \sigma^2_{Y_2} & \sigma_{Y_2Z_2} \\ \sigma_{Z_2X_2} & \sigma_{Z_2Y_2} & \sigma^2_{Z_2} \end{bmatrix}
\end{bmatrix}
\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \textbf{(9)}
$$

Note: Correlation between points 1 and 2 is described by the off-diagonal sub-matrices.

Equation (9) is used to define various accuracies based upon user choices. The matrix operation in equation (9) can be used to compute:

- **Local accuracy**, if the full covariance matrix is employed;
- **Network accuracy**, if the correlation between points 1 and 2 is taken to be zero; and
- **P.O.B. accuracy**, if the covariance matrix of point 2 is the only one used.

Implementation issues related to the various accuracies defined by equation (9) are discussed later. For the sake of completeness, the remaining inverse computations for local coordinate differences, directions, distance, and associated standard deviations are given below. The matrix formulation for computing local coordinate differences from geocentric coordinate differences is (Burkholder 1993):

$$
\begin{bmatrix} \Delta e \\ \Delta n \\ \Delta u \end{bmatrix} = \begin{bmatrix} -\sin\lambda & \cos\lambda & 0 \\ -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad (10)
$$

ϕ= geodetic latitude
λ= geodetic longitude (at point 1).

The Jacobian (rotation) matrix in equation (10) is used with the general error propagation formulation to compute the covariance matrix of local coordinate differences as:

$$
\Sigma_{3D\text{-}INV} = \begin{bmatrix} \sigma^2_{\Delta e} & \sigma_{\Delta e\Delta n} & \sigma_{\Delta e\Delta u} \\ \sigma_{\Delta e\Delta n} & \sigma^2_{\Delta n} & \sigma_{\Delta n\Delta u} \\ \sigma_{\Delta e\Delta u} & \sigma_{\Delta n\Delta u} & \sigma^2_{\Delta u} \end{bmatrix} = J_2\, \Sigma_\Delta\, J_2^t = J_2 \begin{bmatrix} \sigma^2_{\Delta X} & \sigma_{\Delta X\Delta Y} & \sigma_{\Delta X\Delta Z} \\ \sigma_{\Delta X\Delta Y} & \sigma^2_{\Delta Y} & \sigma_{\Delta Y\Delta Z} \\ \sigma_{\Delta X\Delta Z} & \sigma_{\Delta Y\Delta Z} & \sigma^2_{\Delta Z} \end{bmatrix} J_2^t \quad \textbf{(11)}
$$

The functional model equations for a 2-D local tangent plane horizontal distance (Burkholder 1991) and 3-D azimuth (Burkholder 1997d) are:

$$
S = HD(1) = \sqrt{\Delta e^2 + \Delta n^2} \quad (12)
$$

$$
\alpha = \tan^{-1}\left(\frac{\Delta e}{\Delta n}\right) \quad (13)
$$

And the Jacobian matrix of those partial derivatives is:

$$
J_3 = \begin{bmatrix} \frac{\partial S}{\partial \Delta e} & \frac{\partial S}{\partial \Delta n} & \frac{\partial S}{\partial \Delta u} \\ \frac{\partial \alpha}{\partial \Delta e} & \frac{\partial \alpha}{\partial \Delta n} & \frac{\partial \alpha}{\partial \Delta u} \end{bmatrix} = \begin{bmatrix} \frac{\Delta e}{S} & \frac{\Delta n}{S} & 0 \\ \frac{\Delta n}{S^2} & \frac{-\Delta e}{S^2} & 0 \end{bmatrix} \quad (14)
$$

Finally, using the covariance propagation formulation, the 2-D results are:

$$
\Sigma_{2D\text{-}INV} = \begin{bmatrix} \sigma^2_S & \sigma_{S\alpha} \\ \sigma_{S\alpha} & \sigma^2_\alpha \end{bmatrix} = J_3 \Sigma_{3D\text{-}INV} J_3^t =
$$

$$
= J3 \begin{bmatrix} \sigma^2_{\Delta e} & \sigma_{\Delta e\Delta n} & \sigma_{\Delta e\Delta u} \\ \sigma_{\Delta e\Delta n} & \sigma^2_{\Delta n} & \sigma_{\Delta n\Delta u} \\ \sigma_{\Delta e\Delta u} & \sigma_{\Delta n\Delta u} & \sigma^2_{\Delta u} \end{bmatrix} J_3^t \quad (15)
$$

Take square root of diagonal elements to get standard deviations and convert radians to seconds using 206,264.8062471 seconds per radian.

Comments on Features of the GSDM

Using the GSDM raises many questions. It is not possible to anticipate or address them all in one short article, but comments on several features are:

- The GSDM really includes two covariance matrices for each point. A BURKORD™ data base stores values of the geocentric covariance matrix along with the geocentric X/Y/Z coordinates for each point. The local covariance matrix for each point relates more specifically to human perception of local directions and is computed on demand. Alternatively, when defining the positional tolerance of a point, the local standard deviations (or covariance values) may be known and input. In that case the geocentric covariance matrix is computed from the local covariance matrix. The following points focus on the relationship between the local and geocentric covariance matrices:

- The geocentric covariance matrix and the local covariance matrix will both have zero correlations if and only if the standard deviations are the same for all components (spherical uncertainty). Because there is no requirement that one spatial data component must have the same standard deviation as another, one of the two covariance matrices may have no correlations while the other one does. Routinely, both covariance matrices will have correlation values.
- Equation (9) computes the geocentric covariance matrix of the vector between points based on the stored geocentric covariance matrix of each point and the stored point-pair correlations (if they exist).
- Equation (11) computes the local covariance matrix of the vector between points. This is why "horizontal" and "vertical" data can be simultaneously stored in the same 3-D data base. Correlations in the geocentric covariance matrix preserve the integrity of spatial data in the local component directions.

- Reliable covariance and point-pair correlation data are required. Where do they come from? An over-simplified answer is that a user can assign standard deviations (reasonable or otherwise) to the east/north/up components of any point when defining or storing the 3-D position of the point. A better answer is that variances (and covariances) for each point and correlations between points are obtained through error propagation computations of each spatial data measurement contributing to the network. A least squares adjustment of a GPS network is one example. Options are:
  - No covariance (or correlation) data are available in which case the default values for standard deviations are all zero and the positional data (X/Y/Z coordinates) are used as being exact.
- Covariance data are stored, but point-pair correlations are not available because (1) they were computed and not stored or (2) they were not computed. Regardless of the reason, local accuracy computes to be the same as network accuracy if point-pair correlation values are zero or if point-pair correlations are not used.
  - Covariance matrix values for each point and point-pair correlation data are both available. In this case datum accuracy is as defined by the user and the GSDM provides a choice of computing (or using) local accuracy, network accuracy, or P.O.B. accuracy.

Implementation Issues

The mathematical procedures for manipulating spatial data are unambiguous, but the manner in which equation (9) is used and the reference frame to which the GSDM is attached both make a difference when answering the question, "Accuracy with respect to what?"

Given that it is each user's responsibility to verify that legitimate spatial data are being used, and given that stored information depends on quality control measures imposed during the data collection, reduction, adjustment, and verification, the GSDM specifically relies upon, defines, supports, and provides four different kinds of accuracy. They are, in summary:

- **Datum accuracy** of a single point is with respect to a defined reference frame or to whatever control the user decides to hold fixed. Although datum accuracy can be expressed in either the geocentric or a local system, a BURKORD™ data base stores datum accuracy in the geocentric reference frame covariance matrix. Datum accuracy in the local reference frame is computed upon demand using equation (5).
- **Local accuracy** of one point with respect to another is based upon a full covariance matrix in equation (9) which accommodates statistical correlation between endpoints. This option provides answers, the quality of which is determined primarily by the quality of the "connecting measurement" and is largely unaffected by the datum accuracy of the endpoints. While mathematical procedures for storing, tracking, and using the correlations necessary to realize local accuracy answers are well defined, computational procedures for doing so need to be documented better.
- **Network accuracy** of one point with respect to another is based upon the collective datum accuracy of both endpoints and assumes the two point positions are statistically independent. That means the correlation sub-matrices in equation (9) are either zero or assumed to be zero—even if

they are not. Network accuracy is probably the most useful of the various described accuracies because point-pair correlations needed for local accuracy computations are not routinely stored in most spatial data bases. A BURKORD™ data base accommodates storage of point-pair correlations.

- **P.O.B. accuracy** of any forepoint with respect to any user-selected standpoint assumes the two points are statistically independent and that the standpoint position is errorless. In terms of equation (9), the datum accuracies of point 2 are the only covariance values used; all other equation (9) covariance values are taken to be zero. The P.O.B. datum concept is defined as part of the GSDM in which the relative positions of a collection of points are defined with respect to an origin (P.O.B.) selected by the user. Admittedly, P.O.B. accuracy is not appropriate for sophisticated applications, but it is easy to use and may, in fact, be very appropriate for many routine local uses such as cadastral surveys, site development, or other construction surveys.

Another implementation issue is the underlying datum with which the GSDM is used. Although other user-defined environments could also be considered, the three following environments (initial datums) could be judged appropriate for using the GSDM in the United States:

- X/Y/Z values based upon existing published High Accuracy Reference Network (HARN) points;
- X/Y/Z values based upon a global polyhedron network such as the Continuously Operating Reference Stations (CORS); and
- The best possible (current epoch) International Terrestrial Reference Frame (ITRF) values referenced to the earth's center of mass.

Is it necessary or desirable for each of those datums to have a unique name associated with each of the possible accuracies? The answer might be "yes" if that is required to avoid a problem similar to the computer industry problem with the January 1, 2000, date. However, Malys et al. (1997) show that there is little statistical difference between WGS84 and the ITRF94 and conclude transformation between them is not warranted. That being the case, maybe the proposed accuracy names are sufficient in and of themselves. Otherwise, simple modifiers such as HARN, CORS, or ITRF could be used as appropriate.

Conclusion

The GSDM has been defined and proven but has not yet been widely adopted. Given its rigor, simplicity, and universality, the adoption of the GSDM as a standard for storing, manipulating, and exchanging geospatial data is viewed as a matter of time. Although the functional model equations and stochastic model procedures both enjoy mathematical specificity, it is conceded the proposed accuracy names are not consistent with the use of "network accuracy" and "local accuracy" as contained in the FGDC draft accuracy standards (FGDC 1997) published in 1997. Acknowledging other accuracy names might be more appropriate, it is hoped the GSDM and proposed accuracy names will be considered and discussed carefully before final geospatial accuracy standards are promulgated.

REFERENCES

Burkholder, E.F. 1991. Computation of level/horizontal distance." *ASCE Journal of Surveying Engineering* 117(3).

Burkholder, E.F. 1993. Using GPS results in true 3-D coordinate system. *ASCE Journal of Surveying Engineering* 119(1).

Burkholder, E.F. 1997a. Definition of a three-dimensional spatial data model for southeastern Wisconsin. The Southeastern Wisconsin Planning Commission, Waukesha, Wisconsin.

Burkholder, E.F. 1997b. The Global Spatial Data Model: A tool designed for surveyors. *Professional Surveyor Magazine* 17(7).

Burkholder, E.F. 1997c. Using the Global Spatial Data Model (GSDM) in Plane Surveying. *Professional Surveyor Magazine* 17(8).

Burkholder, E.F. 1997d. Three-dimensional azimuth of GPS vector. *ASCE Journal of Surveying Engineering* 123(4).

Burkholder, E.F. 1998a. Positional tolerance made easier with the GSDM. *Professional Surveyor Magazine* 18(1).

Burkholder, E.F. 1998b. A practical Global Spatial Data Model (GSDM) for the 21st Century. Presented at the Institute of Navigation National Technical Meeting, Long Beach, California, January 1998.

DMA (Defense Mapping Agency). 1991. Department of Defense World Geodetic System 1984: Its definition and relationships with local geodetic systems. *Technical Report* 8350.2, Defense Mapping Agency (now National Imagery and Mapping Agency), Fairfax, Virginia.

FGDC (Federal Geographic Data Committee) 1997. Draft Geospatial Positioning Accuracy Standards. FGDC, U.S. Geological Survey, Reston, VA.

Malys, S., J.A. Slater, R.W. Smith, L.E. Kunz, and S.C. Kenyon. 1997. Refinements to the World Geodetic System 1984. Presented at the Institute of Navigation, ION GPS-97, Kansas City, Missouri, September 16-19, 1997.

Mikhail, E. M. 1976. *Observations and least squares*. New York, N.Y.: Harper & Row.

Wolf, P.R., and C.D. Ghilani. 1997. Adjustment computations: Statistics and least squares in surveying and GIS. New York, N.Y.: John Wiley & Sons. ■