Cross-Validated Assessment of Geometric Accuracy

Kenneth C. McGwire

Abstract
The recent emphasis on metadata standards must be accompanied by vigilance in unbiased reporting of geometric accuracy. A cross-validation technique is shown to be capable of providing more accurate estimates of geometric error than the traditional method of using transformation residuals when modest numbers of ground control points are available. This method also provides a much more accurate indication of the effects of choosing different polynomial orders.

Introduction
A cross-validated method of estimating error in polynomial-based geometric transformations is presented. Polynomial rectifications are widely applied in both remote sensing and geographic information system (GIS) applications. For cases with moderate numbers of control points, the cross-validated method is shown to provide estimates of geometric error which are as good or better than traditional methods which use residuals directly from the transformation calculation. Polynomial rectification calculates a global mathematical transformation for converting from one image or map coordinate system (u, v) to another (x, y) (Moik, 1980; Castleman, 1979). The general form of the polynomial model is

$$x = \sum_{i=0}^{N}\sum_{j=0}^{N-i} a_{ij}u^i v^j \qquad y = \sum_{i=0}^{N}\sum_{j=0}^{N-i} b_{ij}u^i v^j \tag{1}$$

where u and v are the Cartesian coordinates of the original image, x and y are the Cartesian coordinates of the transformed coordinate system, N is the mathematical order of the polynomial equation, and a and b are empirically derived coefficients.

Thus, first-, second-, and third-order polynomials would have three, six, and ten terms, respectively, and would require at least a like number of control points for calculation. Additional points would result in a statistical fit for the transformation with increasing degrees of freedom. In order to choose the appropriate equation for this transformation, one must understand the trade-offs between the mathematical stability of lower order polynomials versus the ability of higher order polynomials to fit more complex patterns of geometric distortion.

Typically, the underlying distortion is not entirely understood when performing a geometric rectification, and an analyst must determine the appropriate polynomial expression from empirical evidence. The standard method for assessing the accuracy with which a transformation models the actual geometric distortion is the root-mean-square error (RMSE) of the residuals for points used in the transformation. RMSE is calculated as

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\right]^{\frac{1}{2}} \tag{2}$$

where n is the number of samples and σ is the estimated value minus the actual value.

RMSE is often calculated separately for the x and y components of each control point in order to provide information on relative error associated with that point. Total RMSE based on transformation residuals is often cited as an accuracy statistic for the entire image. However, because the control points are not independent from the transformation coefficients, RMSE will underpredict the actual error found elsewhere in the transformed image when the degrees of freedom are small. As the number of control points increases, the transformation coefficients will converge towards an improved estimate and RMSE will asymptotically approach actual error. This relationship is portrayed in Figure 1.

The most effective manner for testing the actual geometric accuracy of a transformation is to use an independent set of control points. These independent points are sometimes referred to as "pass" or "check" points. However, in many remote sensing applications, the availability of control points may be quite limited. This problem often arises in remote areas with few cultural features or when coarse spatial resolution limits the discrimination of ground features. In these cases, an analyst may not have the luxury of exempting potential control points from the calculation in order to have a pool of independent check points. The result will generally be an overly optimistic statement of geometric accuracy for the rectified image based on RMSE of transformation residuals.

Beyond the problems of reporting biased estimates of cartographic fidelity, a misunderstanding of the geometric accuracy for transformed products may have strong repercussions on subsequent analyses. For example, Townshend *et al.* (1992) document examples where misregistration by 0.2 pixels may cause 10 percent error in change detection using satellite imagery and 1.0 pixel misregistration may cause 50 percent error. Clearly, it is important in such studies to have an unbiased understanding of the true registration accuracy. Similar problems may occur when map products derived from the transformed imagery are combined with other datasets in a geographic information system (GIS) (Chrisman, 1989; Walsh *et al.*, 1987; Vitek *et al.*, 1984).

A biased error estimate may also create problems when the analyst uses the RMSE to choose between various orders of polynomials in the geometric rectification. This may be

Desert Research Institute, Biological Sciences Center, Reno, NV 89431.

Photogrammetric Engineering & Remote Sensing, Vol. 62, No. 10, October 1996, pp. 1179–1187.

0099-1112/96/6210-1179$3.00/0

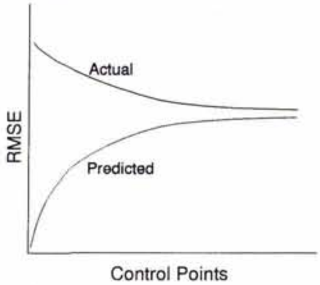

Figure 1. Idealized curves of estimated and actual RMSE versus number of control points.

ameliorated to some degree by comparing the RMSE of alternate transformations using similar degrees of freedom (i.e., more points for higher order polynomials). However, the lack of independent test points may obscure whether higher order polynomials are simply fitting localized displacements in the control points or whether they are modeling legitimate global trends. While higher order polynomials will fit complex distortions in the control points more effectively, they also may create large, undesirable deviations elsewhere in the image. Welch *et al.* (1985) demonstrate how higher order polynomials provide worse results for rectifications of Landsat Thematic Mapper (TM) data even though the RMSE suggests an improved transformation. Despite this finding, it is still not uncommon to find examples of third or higher order polynomials being applied to Landsat data.

Despite the great importance of the relationships between the number of control points, degrees of freedom, bias in the RMSE, and actual geometric error, I could not find any serious discussion of this in any of the seven college level remote sensing textbooks I have at my disposal. Of these, only a couple refer to the minimum number required to calculate a transformation or to the use of pass points. The *Manual of Remote Sensing* (Colwell, 1983) does cite the reduction in absolute geometric error with increasing numbers of points, however, I could find no mention of the bias in RMSE as calculated by most image processing software packages.

A possible solution to the problem of biased error estimates is offered by cross-validation methods. Instead of performing a single transformation with all the control points, the cross-validated method uses iterative sample substitution and calculation to create a pool of pseudo-check points which are independent from a corresponding pool of transformation coefficients. The mean error for predicting these pseudo-check points is expected to more effectively represent the actual error of a single geometric rectification calculated from all control points.

Approach

The cross-validation approach used here is based on the jack-knifing technique described by Mosteller and Tukey (1977). Rather than simply taking the residuals of a single transformation, the cross validated RMSE (RMSE*) is calculated by removing the first control point (u_i, v_i) from the pool and calculating a geometric transformation using the remaining points $(u_{!i}, v_{!i})$. This transformation is then applied to the first point (u_i, v_i) and its difference from the true value (x_i, y_i) is calculated. The first point is then replaced in the pool, and the process is applied iteratively to each point in the entire set. The differences in x and y are then substituted for σ in Equation 2. For example, the cross-validated RMSE for the x component of the rectification would be

$$\mathrm{RMSE}_x^* = \left\{\frac{1}{n}\sum_{i=1}^{n}[x_i - f(u_i)]^2\right\}^{\frac{1}{2}} \quad (3)$$

where $f(u_i)$ equals the transformation $u \rightarrow x$ calculated without sample i applied to sample i.

I first tested the cross-validated RMSE on coordinate pairs with a known geometric transformation. Thirty coordinate pairs (u, v) were generated to represent image control points. Simulated map coordinates (x, y) for these points were then created using a simple affine transformation with the addition of random local displacements. This was calculated by multiplying u and v coordinates by 30.0; rotating by 45° around the origin; and adding a random, normally distributed, circular error (mean = 0.0, standard deviation = 15.0) to each point.

After this, the cross-validated method was tested on Landsat TM and Daedelus NS001 scanner data with nominal ground resolutions of 28.5 and 18 metres, respectively. The study area for this test corresponded to four adjacent U.S. Geological Survey (USGS) 7.5-minute quadrangle maps covering an area of foothills in the Sierra Nevada of California (Millerton Lake West, Little Table Mtn, Knowles, and O'Neals). The area ranges in elevation from approximately 120 to 900 metres above sea level and contains moderate, rolling topography. Coordinates for 30 control points were identified in the TM imagery (Scene-ID LT5042034008821610) and digitized from the USGS map sheets. Thirty control points were also generated for the NS001 data (NASA flight # 88-112), with 24 of these points corresponding directly to those selected in the TM data.

Geometric rectifications of these three datasets were performed using different orders of polynomials and varying numbers of control points. The number of control points ranged from a minimum of one degree of freedom for the given order of polynomial to a maximum of 20 points. For each combination of polynomial order and number of control points, 25 subsets were randomly selected from the total pool of 30 control points. This replication was done to provide an indication of the range in values which might be encountered if different individuals were to have selected control points in the area. Ten independent check points were randomly selected to accompany each subset, and three different calculations of RMSE were performed on the transformation of each subset. The three methods for calculating RMSE were

- the standard method using residuals of the transformation calculation,
- testing transformation coefficients against independent check points, and
- the cross-validation method.

Results

Simple Affine Transformation

Results for the affine transformation with random local displacements provided a useful reference because the actual nature of the geometric distortion was known. Table 1 provides the mean value for each type of RMSE calculation and treatment (i.e., polynomial order/number of points). Boxplots in Figures 2 and 3 display the distribution of RMSE values from the 25 replications for each treatment. As expected, RMSE for the standard method based on transformation residuals started low and increased to a maximum of approximately 15 for both first- and second-order polynomials. RMSE based on transformation residuals stabilized at approximately seven control points (df = 3) for the first-order transformation. After this, the transformation residuals underestimated the actual error, as indicated by the check points, by approx-

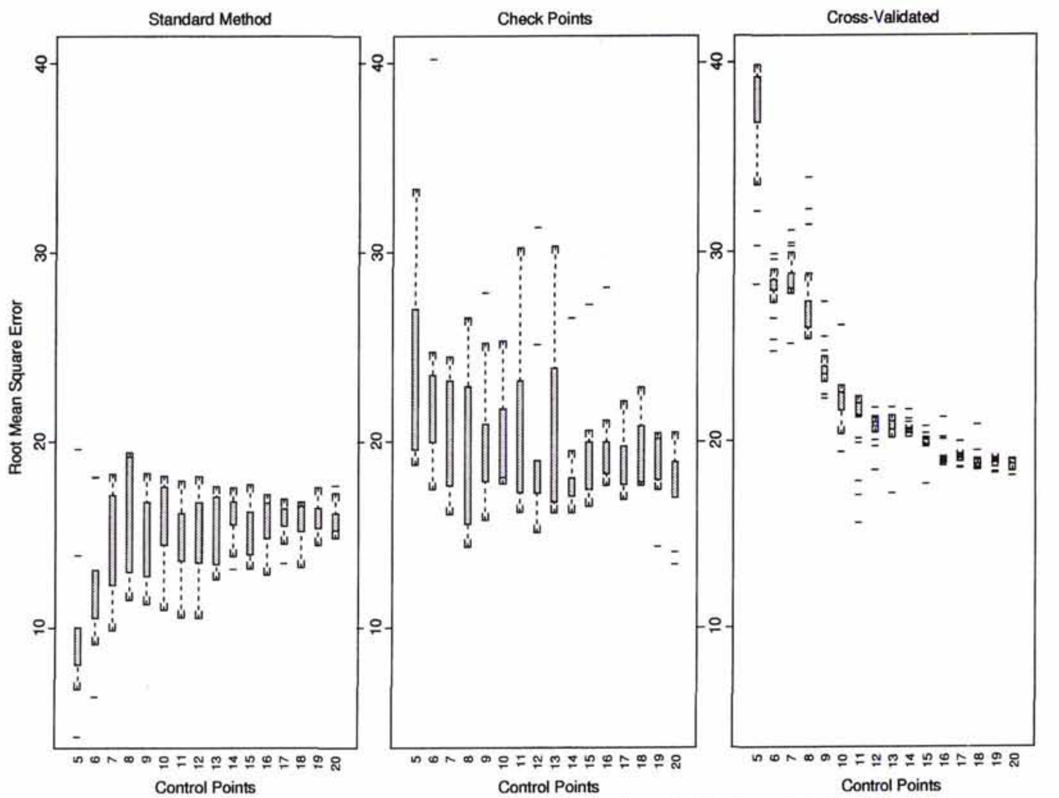

Figure 2. RMSE for first-order transformation of simulated data.

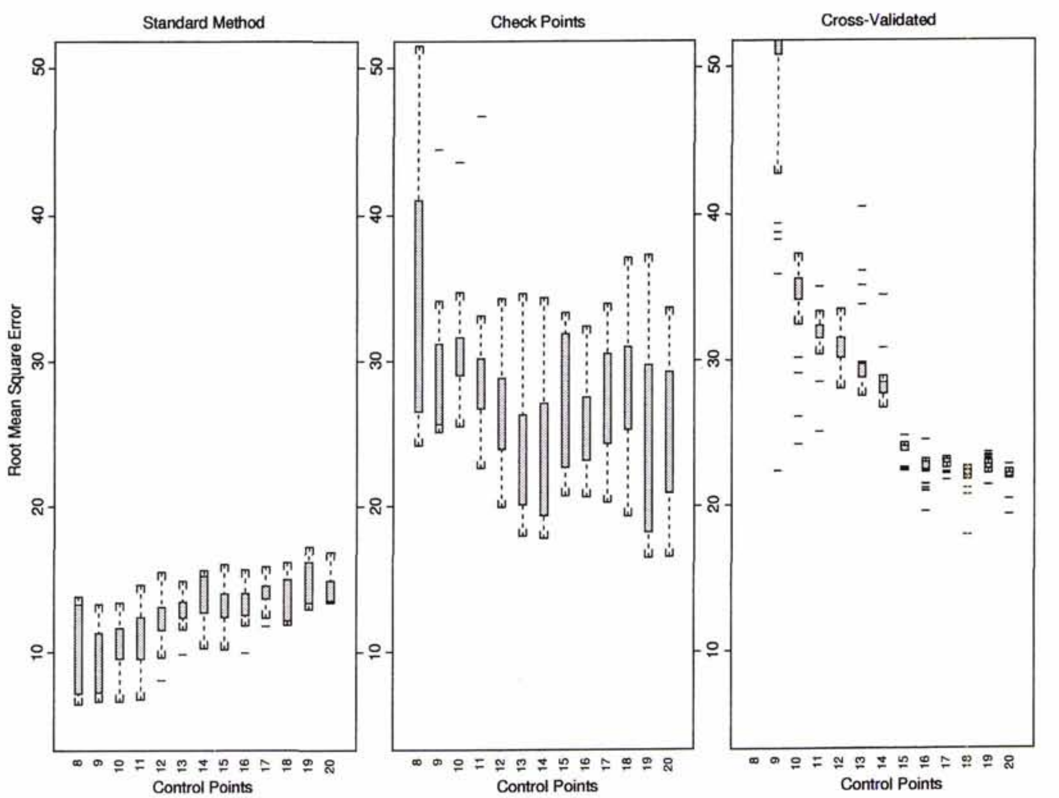

Figure 3. RMSE for second-order transformation of simulated data.

imately 15 to 25 percent. The cross-validated RMSE overpredicted greatly at low numbers of control points, but converged towards a more accurate estimate of actual error with nine or more control points.

As the number of control points increased, the RMSE based on transformation residuals for the second-order transformation increased more slowly than for the first-order transformation. Even when comparing across similar degrees of freedom, the traditional RMSE measure suggested better results than first order until 19 control points were used. However, the actual RMSE from the independent check points showed that the second-order transformation provided consistently worse results. Again, the cross-validated RMSE greatly overpredicted actual error of the second-order transformation when the number of control points was very small. However, it quickly converged towards much closer agreement with the independent check points. Even with 20 control points, the standard RMSE based on the second-order transformation residuals underpredicted actual error by 40 percent while the cross-validated measure showed a discrepancy of only 12 percent. Given the relatively small difference in the standard RMSE between first- and second-order polynomials, it is likely that an experienced analyst would correctly select the more conservative first-order polynomial. However, the cross-validated RMSE provided a much clearer indication that the first-order transformation is more appropriate.

I believe that the overprediction of error by the cross-validated method at very low numbers of control points arose from two sources. First, because one control point was left out of the transformation calculation during each iteration, the degrees of freedom were reduced and the calculation was more sensitive to local deviations from the global trend. Second, at small numbers of control points there was a much higher probability that the sample being omitted from a calculation fell outside of the convex hull of the remaining points. The resulting extrapolated estimates were likely to have particularly high error. The cross-validated RMSE for the first-order transformation was more accurate than the standard method when using nine or more control points and provided better results for second order with ten or more control points. Because the overprediction of RMSE was quite large when the minimum possible number of control points was used in the cross-validated method, the scaling of the *y* axis on many of the following figures was set to truncate large values. This decision was made to increase the readability of the diagrams as a whole.

One feature which is apparent in Figures 2 and 3 is the tight variance displayed by the cross-validated RMSE estimates. This occurs because each cross-validated estimate is itself the mean of multiple iterations. Variance for a single cross-validated RMSE (σ_*^2) is calculated as (Mosteller and Tukey, 1977)

$$\sigma^2 = \frac{\sum_{j=1}^{n} \mathrm{RMSE}_{*j}^2 - \frac{1}{n}\left(\sum_{j=1}^{n} \mathrm{RMSE}_{*j}\right)^2}{n-1}$$
$$\sigma_*^2 = \frac{\sigma^2}{n} \qquad (4)$$

Landsat TM Rectification

Landsat TM data are characterized as having very stable image geometry (Welch *et al.*, 1985). The range of elevations found in the test area, while not particularly extreme, were sufficient to create localized scale changes and displacements. It is not immediately clear from examination of topographic maps if a higher order polynomial would help to compensate for topographic trends. Mean values of the three different RMSE calculations are provided for treatments of the TM imagery in Table 2. Boxplots displaying the spread of values for first- and second-order polynomial models are displayed in Figures 4 and 5.

Once again, RMSE values based on transformation residuals rose as the number of control points increased. Comparison with check points indicates that the standard RMSE underpredicted error in the first-order transformations by at least 10 percent. It is somewhat surprising to find that the first-order polynomial continued to provide biased estimates even with the use of 20 control points. Using the standard RMSE, the second-order transformation appeared to have 10 percent less error than first order after RMSE levels off. However, check points show that the standard RMSE generally underpredicted error in the second-order transformation by at least 20 percent. As before, the second-order transformation actually provided worse results than first order. This result is consistent with the previous findings of Welch *et al.* (1985) for TM data.

The cross-validated RMSE overpredicted greatly with small numbers of GCPs, but converged towards very close agreement with the actual error documented in the check points. The cross-validated RMSE provided comparable results to the standard method when the number of control points exceeded 11 for first order. However, cross-validation

TABLE 1. MEAN RMSE FOR TEST COORDINATES USING (A) TRANSFORMATION RESIDUALS, (B) INDEPENDENT CHECK POINTS, AND (C) THE CROSS-VALIDATED METHOD.

# Control	First Order			Second Order		
Points	A	B	C	A	B	C
5	10	25	39			
6	12	24	28			
7	14	21	29			
8	15	20	27	9	37	139
9	15	21	24	10	31	51
10	15	20	22	10	32	34
11	15	21	21	11	31	32
12	15	20	21	12	27	31
13	15	20	21	13	26	30
14	16	19	21	14	24	28
15	16	19	20	13	27	24
16	15	20	19	13	26	22
17	16	19	19	14	27	23
18	16	20	19	14	27	22
19	16	19	19	15	25	23
20	16	17	19	15	25	22

TABLE 2. MEAN RMSE (METRES) FOR TM DATA USING (A) TRANSFORMATION RESIDUALS, (B) INDEPENDENT CHECK POINTS, AND (C) THE CROSS-VALIDATED METHOD.

# Control	First Order			Second Order		
Points	A	B	C	A	B	C
5	14	32	57			
6	17	30	42			
7	18	29	38			
8	18	26	33	11	50	203
9	19	24	31	12	38	64
10	19	23	29	14	34	56
11	19	23	26	14	30	45
12	19	22	25	16	26	42
13	19	23	24	15	26	35
14	20	22	25	16	26	33
15	20	22	25	17	26	29
16	20	22	25	17	24	29
17	20	22	24	18	25	29
18	20	22	24	18	25	28
19	20	23	23	17	26	26
20	20	24	23	17	26	26

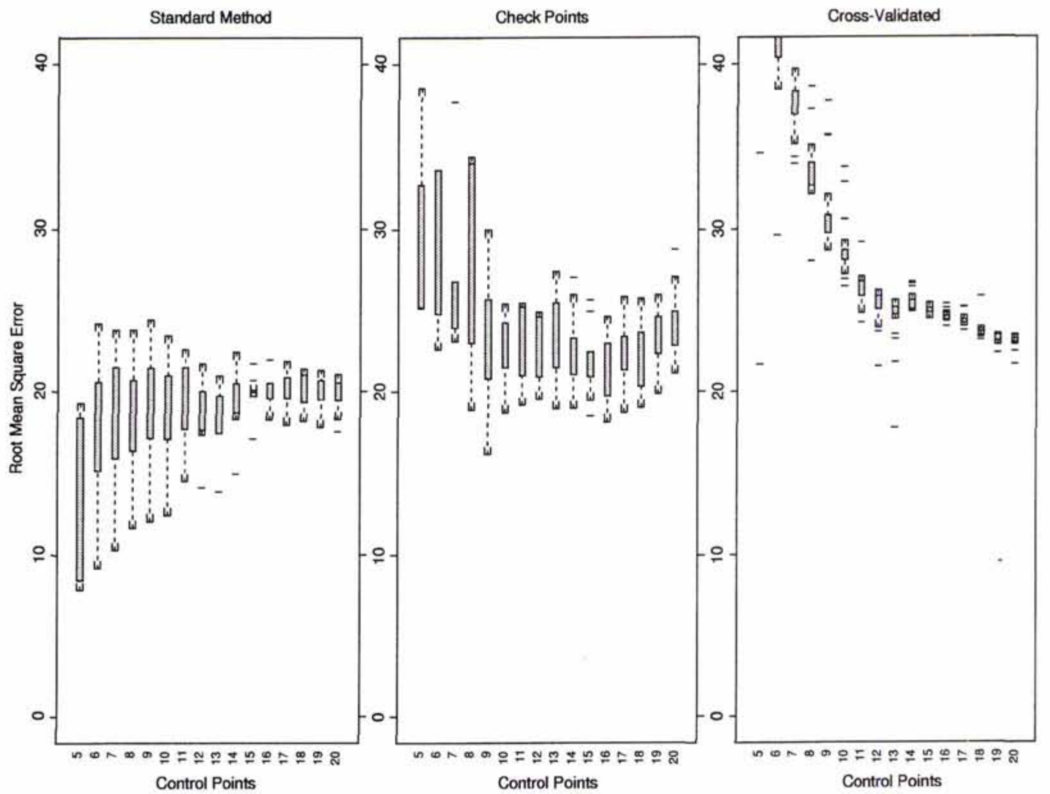

Figure 4. RMSE (metres) for first-order rectification of TM data.

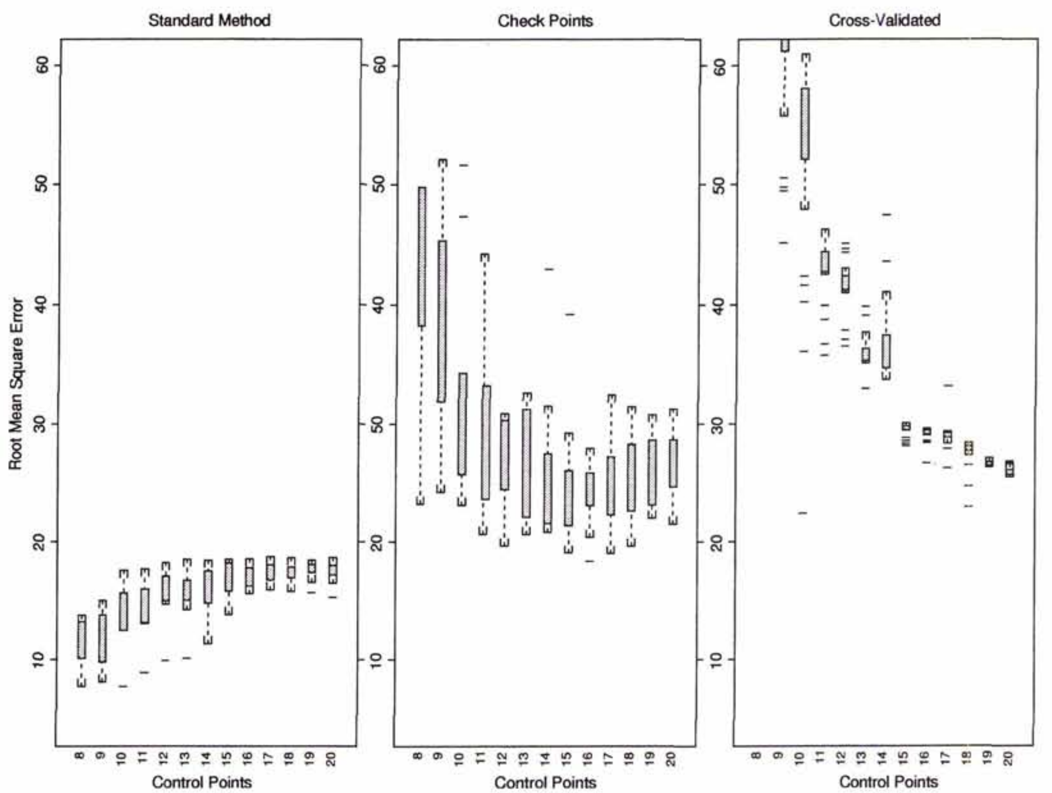

Figure 5. RMSE (metres) for second-order rectification of TM data.

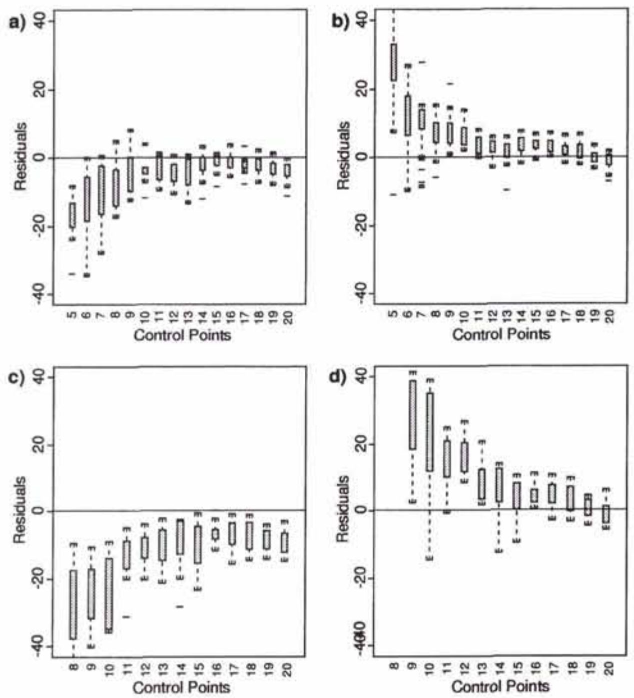

Figure 6. Bias in TM RMSE estimates: (a) standard/first order, (b) cross-validated/first order, (c) standard/second order, (d) cross-validated/second order.

provided consistently better results when more than 13 control points were used for second order. Despite the general overprediction with small numbers of control points, there were examples of cross-validation providing equivalent or better results than the traditional method with as few as seven points in the first-order transformations, or nine points in the second-order transformations. The cross-validated measure made it much clearer than did the standard RMSE that the second-order transformation was not appropriate for use in the rectification.

The bias in the standard RMSE and the cross-validated RMSE is displayed in Figure 6 by subtracting these estimates from the RMSE of the independent check points. The consistent underprediction by transformation residuals is contrasted with the overprediction of the cross-validated method. This might suggest the use of both statistics to establish upper and lower bounds on geometric accuracy when there are relatively small numbers of control points. However, cross-validation did underpredict the actual error of some second-order treatments in the test of the affine transformation, so this might not be a robust strategy. With both orders of polynomials, the cross-validated RMSE becomes practically unbiased with 19 or 20 control points.

Daedelus NS001 Rectification

The NS001 scanner is subject to much more complex geometric distortions than TM data due to the relative instability of aircraft platforms and the increased influence of topographic relief at aircraft operating altitudes. Further, the data used here had not been corrected for systematic changes in pixel size with off-nadir view angles. Thus, it was expected that a higher order polynomial would be required for rectification, though it was unclear what order polynomial would be appropriate.

In actuality, polynomial-based approaches are generally inadequate for this complex type of distortion. Analysts continue to apply polynomial rectifications to NS001 imagery, though much better results may be obtained using localized interpolation based on triangles or quadrilaterals (Chen and Rau, 1993; Devereaux *et al.*, 1990; Thormodsgard and Lillesand, 1987; Zobrist *et al.*, 1983), global positioning systems and digital elevation data (Fisher, 1991), or other new methods such as thin plate splines and the biharmonic, multiquadric method (Fogel and Tinney, 1994). This example is provided for the sake of exploring the robustness of the cross-validated RMSE, rather than endorsing the headache of performing polynomial rectifications on aircraft scanner data.

Mean RMSE estimates for rectifications of the NS001 imagery are presented in Table 3. Boxplots displaying the spread of values for first-, second-, and third-order polynomials are presented in Figures 7, 8, and 9. As usual, the RMSE estimated from transformation residuals increased consistently with the number of control points. Actual error indicated by check points generally decreased, though there was a small upturn in RMSE for the first-order transformation at high numbers of control points. This was likely due to a temporarily recurring pattern in the randomly selected subset of control points.

With 20 control points, the standard RMSE underpredicted the error of first order by 24 percent, second order by 31 percent, and third order by 54 percent relative to the independent check points. For first- and second-order polynomials, the typical overprediction observed in the cross-validated RMSE was less of a problem than the underprediction by transformation residuals regardless of the number of control points. The cross-validated RMSE performed better for third-order transformations when the number of control points exceeded 13. As with some treatments in the affine transformation test, the mean cross-validated RMSE for second-order transformations sometimes underestimated actual error by as much as 16 percent. However, this is still seen to be an improvement over the standard RMSE metric.

In examining the standard RMSE in Table 3, an analyst may be tempted to choose the third-order polynomial with 20 control points. However, even with the standard measure, it is clear that the third-order model is not necessarily an improvement when comparing with consistent degrees of freedom (i.e., versus 16 points with second order). The cross-validated RMSE correctly indicates that the third-order transformation provides worse results no matter how many of the 20 possible control points are used.

Table 3. Mean RMSE (Metres) for NS001 Data Using (A) Transformation Residuals, (B) Independent Check Points, and (C) the Cross-Validated Method.

# Control Points	First Order A	B	C	Second Order A	B	C	Third Order A	B	C
5	49	262	332						
6	73	216	228						
7	90	171	194						
8	91	188	159	48	308	409			
9	97	178	157	57	275	259			
10	104	173	159	73	205	244			
11	113	172	167	76	213	212			
12	115	169	163	79	182	188	37	496	1168
13	117	166	161	80	183	177	46	337	942
14	117	168	158	81	179	169	48	338	302
15	118	168	152	84	177	160	55	296	305
16	123	165	151	89	167	155	72	229	299
17	120	173	146	90	166	152	73	217	257
18	127	174	155	96	163	149	81	229	253
19	128	178	159	97	157	148	82	210	248
20	130	172	158	101	147	151	86	186	230

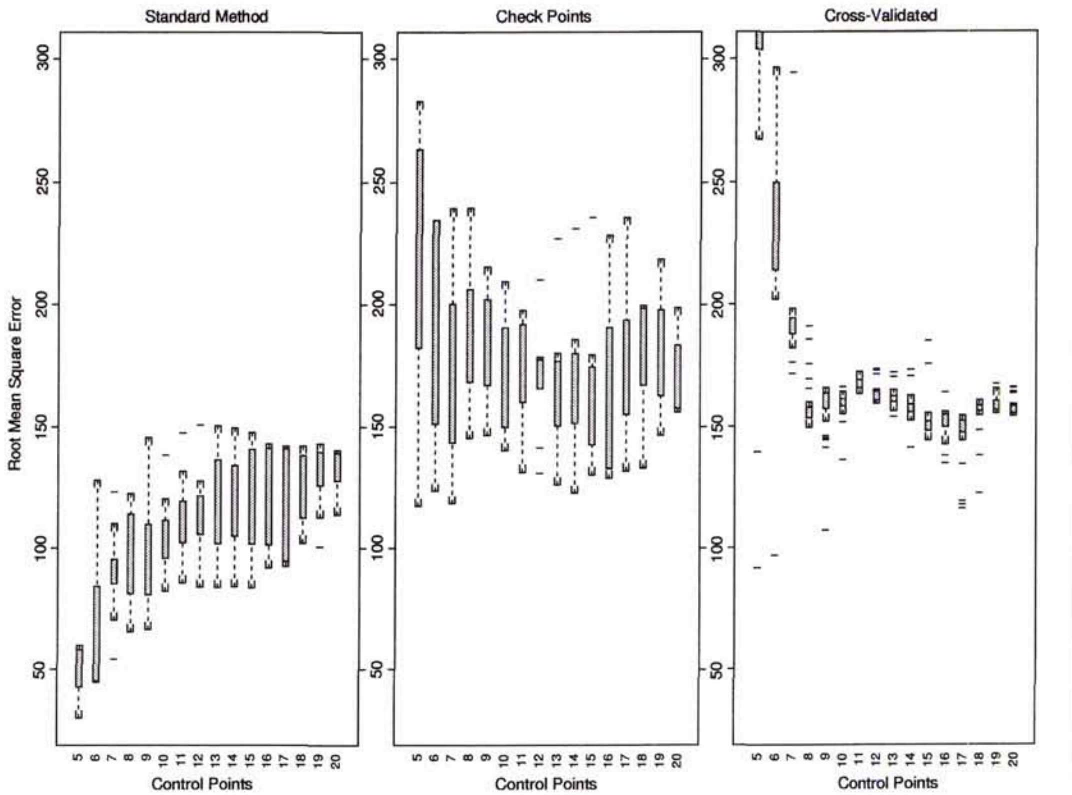

Figure 7. RMSE (metres) for first-order rectification of NS001 data.

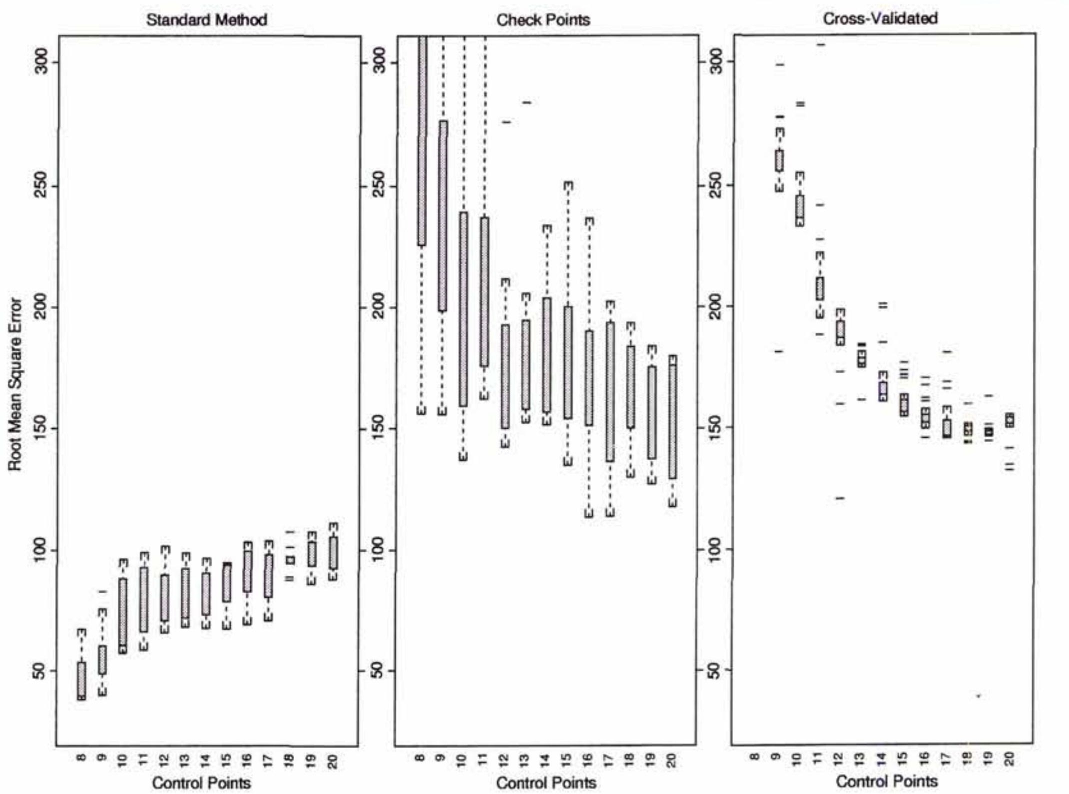

Figure 8. RMSE (metres) for second-order rectification of NS001 data.

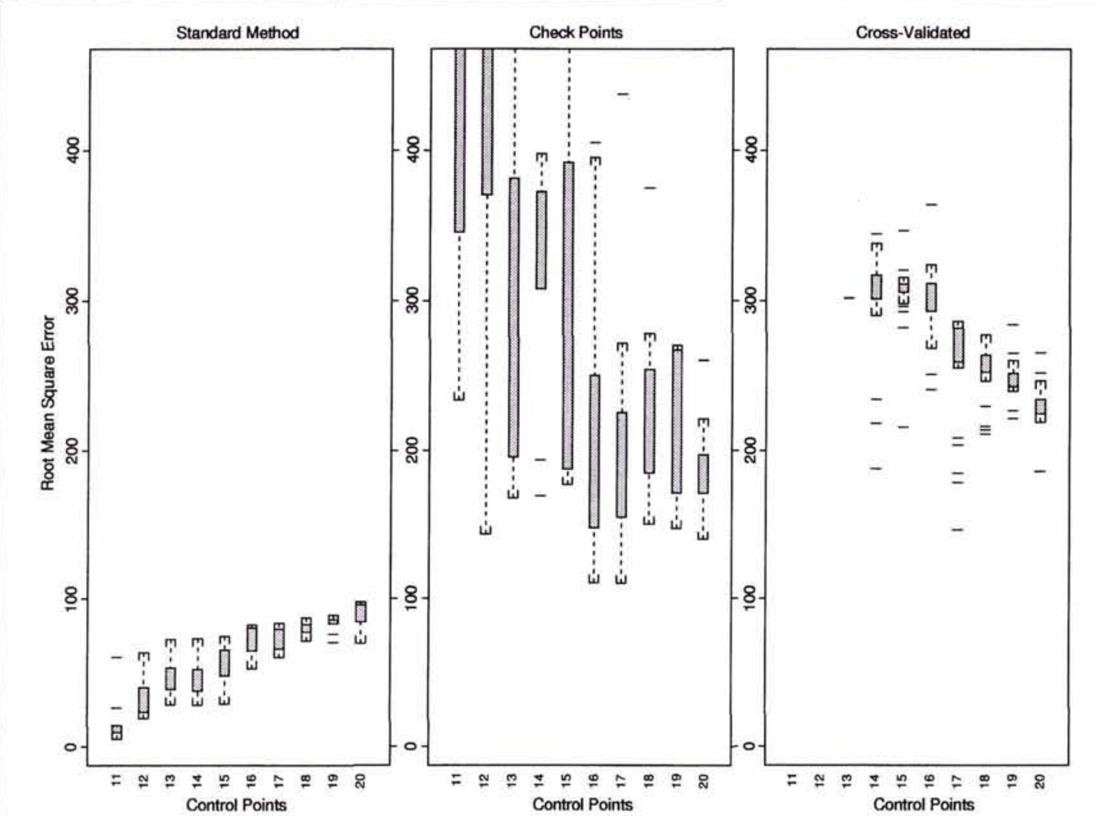

Figure 9. RMSE (metres) for third-order rectification of NS001 data.

Conclusions

The cross-validated RMSE was found to provide as good or better estimates of geometric error than those provided by transformation residuals when the number of control points exceeded a relatively small threshold value. It appears that the cross-validated RMSE should be suspect until the number of control points exceeds five degrees of freedom for a given order of polynomial. This requirement should not limit its effective use, however, because smaller numbers of control points are shown to make the standard RMSE calculation quite suspect as well. In addition to providing a more accurate depiction of geometric error, the cross-validated RMSE also greatly reduces the ambiguity in selecting the polynomial order of a rectification.

The cross-validated method is computationally demanding because it requires as many transformations to be calculated as there are control points. However, historical computational limits have practically vanished in the face of modern processing capabilities. The importance of accurate statements for the geometric characteristics of spatial data products has been widely acknowledged (Townshend *et al.*, 1992; Lunetta *et al.*, 1991; Chrisman, 1989; McGwire and Goodchild, 1996). The standard RMSE, as calculated from transformation residuals, is a diagnostic which has been confused as an accuracy statistic. It is actually a weak diagnostic at that, given that it generally cannot identify when a higher order polynomial provides significantly worse results. It is the author's opinion that the cross-validated RMSE, or a similar alternative, should be provided in addition to the standard RMSE measure when independent check points are not used. This requires both incorporation of the algorithm in image processing and GIS software, as well as its consideration as a metadata entry in data accuracy standards.

Acknowledgment

This work was supported in part by the Remote Sensing Research Unit at the University of California at Santa Barbara and by NASA Grant NAGW-1743. Tim Minor's review of the manuscript was greatly appreciated.

References

Castleman, K.R., 1979. *Digital Image Processing*, Prentice-Hall, 429 p.

Chen, L., and J. Rau, 1993. Geometric correction of airborne scanner imagery using orthophotos and triangulated feature point matching, *International Journal of Remote Sensing*, 14(16):3041-3059.

Chrisman, N., 1989. Modeling error in overlaid categorical maps, *Accuracy of Spatial Databases*, Taylor and Francis, New York.

Colwell, R. (editor), 1983. *The Manual of Remote Sensing, 2nd Edition*, American Society for Photogrammetry and Remote Sensing, Fall Church, Virginia.

Devereaux, B., R. Fuller, L. Carter, and R. Parsell, 1990. Geometric correction of airborne scanner imagery by matching delauney triangles, *International Journal of Remote Sensing*, 11(12):2237-2251.

Fisher, L., 1991. Geometric correction of multispectral scanner data using global positioning system and digital terrain models, *Proceedings of a special session of the ACSM-ASPRS Annual Convention: The Integration of Remote Sensing and Geographic Information Systems* (J. Star, editor), American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia.

Fogel, D., and L. Tinney, 1994. *Image Registration Using Multiquadric Functions*, Final Report for Dept. of Energy P.O. 7870M, Remote Sensing Research Unit, Dept. of Geography, University of California, Santa Barbara.

Lunetta, R., R. Congalton, L. Fenstermaker, J. Jensen, K. McGwire, and Larry Tinney, 1991. Remote sensing and geographic information system data integration: Error sources and research issues, *Photogrammetric Engineering & Remote Sensing*, 57: 677-687.

McGwire, K., and M. Goodchild, 1996. Accuracy, a chapter in *Integration of Remote Sensing and GIS*, Cambridge University Press, New York, in press.

Moik, J., 1980. *Digital Processing of Remotely Sensed Images*, NASA SP-431, National Aeronautics and Space Administration, Washington, D.C.

Mosteller, F., and J. Tukey, 1977. *Data Analysis and Regression: A Second Course in Statistics*, Addison-Wesley, Reading, Massachusetts.

Thormodsgard, J., and T. Lillesand, 1987. Comparison of the gridded finite element and the polynomial interpolations for geometric rectification and mosaicking of Landsat data, *Proceedings of the ASPRS-ACSM Annual Convention*, Baltimore, Maryland, pp. 139-151.

Townshend, J., C. Justice, C. Gurney, and J. McManus, 1992. The impact of misregistration on change detection, *Institute of Electrical and Electronic Engineers Transactions on Geosciences and Remote Sensing*, 30:1054-1060.

Vitec, J., S. Walsh, and M. Gregory, 1984. Accuracy in geographic information systems: An assessment of inherent and operational errors, *Proceedings, PECORA IX Symposium*, pp. 296-302.

Walsh, S., D. Lightfoot, and D. Butler, 1987. Recognition and assessment of error in geographic information systems, *Photogrammetric Engineering & Remote Sensing*, 53:1423-1430.

Welch, R., T. Jordan, and M. Ehlers, 1985. Comparative evaluations of the geodetic accuracy and cartographic potential of Landsat-4 and Landsat-5 Thematic Mapper image data, *Photogrammetric Engineering & Remote Sensing*, 51(9):1249-1262.

Zobrist, A., N. Bryant, and R. MacLeod, 1983. Technology for large digital mosaics of Landsat data, *Photogrammetric Engineering & Remote Sensing*, 49(9):1325-1335.

(Received 16 October 1995; accepted 17 January 1996; revised 13 February 1996)