BioOne DIGITAL LIBRARY

Uncertainty in Digital Elevation Models of Axel Heiberg Island, Arctic Canada

Authors: Cogley, J. Graham, and Jung-Rothenhäusler, F.

Uncertainty in Digital Elevation Models of Axel Heiberg Island, Arctic Canada

J. Graham Cogley* and
F. Jung-Rothenhäusler†

Department of Geography, Trent
University, Peterborough, Ontario, K9J
7B8, Canada
gcogley@trentu.ca
*To whom correspondence should be
addressed.
†Present address: RapidEye AG,
Wolfratshauser Str. 48, D-81379 Munchen,
Germany.

Abstract

We analyze errors in several digital elevation models (DEMs) of part of Axel Heiberg Island, Nunavut, Canada. We define total error as the sum of map-reading error, which measures the fidelity of the DEM to the map, and mapping error, which measures the fidelity of the map to the terrain. Three of the DEMs derive from estimates made by eye on maps prepared for research purposes and having scales of 1:10,000 ("10K"), 1:50,000 ("50K") and 1:100,000 ("100K"), while two are based on spatial interpolation from published topographic maps of scale 1:250,000 ("250K") and 1:1 000,000 ("1M"). Map-reading errors are reduced markedly by proofreading, and are unbiased but not normally distributed. They are also moderately correlated, with decorrelation distances of 1 to 3 DEM resolution elements. Replicate readings from maps 10K and 50K have rms differences of about 20 m, a figure which shrinks to 12 m when gross errors are identified and excluded. These differences represent the mapping errors. Total error in the 10K DEM may range from ~1 m up to, at worst, ~19 m, while for the 50K DEM the total error may reach 20 m. Excluding gross errors, the worst-case estimates of total error decrease to 11–12 m. Total error is estimated as 90 m for 250K and 158 m for 1M. We show that map-reading error is small in comparison with mapping error. However there are three obstacles to formal description of total DEM error. First, there is no objective basis for partitioning the mapping error between the two maps of a comparison pair. Second, gross errors cannot be accommodated satisfactorily. Third, because the usual statistical assumptions are violated the errors define confidence regions narrower than the usual 68% by some unknown amount. Maps of larger scale have steeper slopes. Differences in the frequency distribution of slopes are such that a simple additive correction would worsen, not improve, the gentle-slope bias of the smaller-scale DEMs. Elevation errors are roughly equal to one half of the contour interval of the parent map. It is not obvious why this should be so, but as a practical rule of thumb it should perform well.

Introduction

The question of digital elevation model (DEM) accuracy is typically difficult to answer, not least because studies which compare DEM elevations to ground truth are few. Statistical complications compound the difficulty of making quantitative statements about the uncertainty in a digital estimate of elevation or slope. The situation is worse in remote regions where the available information is often unsatisfactory because, for example, the best maps are of small scale. A problem peculiar to glaciological cartography (e.g., Aðalgeirsdóttir et al., 1998) is that uncertainty is particularly large in the accumulation zones of glaciers, where broad expanses of snow offer little or no contrast for the photogrammetrist to work from. We offer here an analysis of errors in several DEMs which cover the glacierized Expedition Fiord area of central Axel Heiberg Island, Nunavut, Canada (Fig. 1). Three of the DEMs are based on large-scale mapping with good or fair ground control, while two are standard digital products derived from the small-scale maps which are more typical of the coverage available for high-latitude regions.

The description of DEM errors has evolved rapidly in recent years (e.g., Robinson, 1994; Hunter and Goodchild, 1995, 1997; Ehlschlaeger, 2002). However there are still few detailed empirical studies of the accuracy of DEMs derived from maps, although DEMs derived directly from air photos have been well studied. For example, Davis et al. (2001) reported errors of 2 to 3 m, relative to GPS elevations with decimeter-level accuracy. Liu et al. (1999) compared DEMs from

1:250,000 and 1:50,000 scale maps of the Dry Valley region of Antarctica and found a mean difference of 75 ± 218 m and an rms (root mean square) difference of 132 m. Much better accuracy is available with modern methods. Bamber et al. (2001) showed that radar-altimetric elevations over the Greenland Ice Sheet had standard errors of about ±2 to 14 m for slopes less than 1°; these errors are relative to airborne laser altimetry, for which errors are at the submeter level. However in glaciology the mismatch between modern estimates and the less accurate elevations obtained from old maps is of considerable concern. It means that errors in estimates of volumetric change will be due largely to the old maps (Arendt et al., 2002).

This paper is organized as follows. In the next two sections we describe the DEMs to be studied and the methods used to generate them. Section four deals with the error model which we apply to the DEMs. The fifth and sixth sections address the two components of elevation and slope error, namely map-reading error (fidelity of the DEMs to their parent maps) and mapping error (fidelity of the maps to the terrain). The seventh section generalizes the results of analysis and offers a synthesis, and the final section summarizes the study.

Maps and Digital Elevation Models

10K, 50K, AND 100K DEMs

Early surveys of the Expedition Fiord area (Fig. 2) led to the publication of several large-scale maps, three of which are the focus of
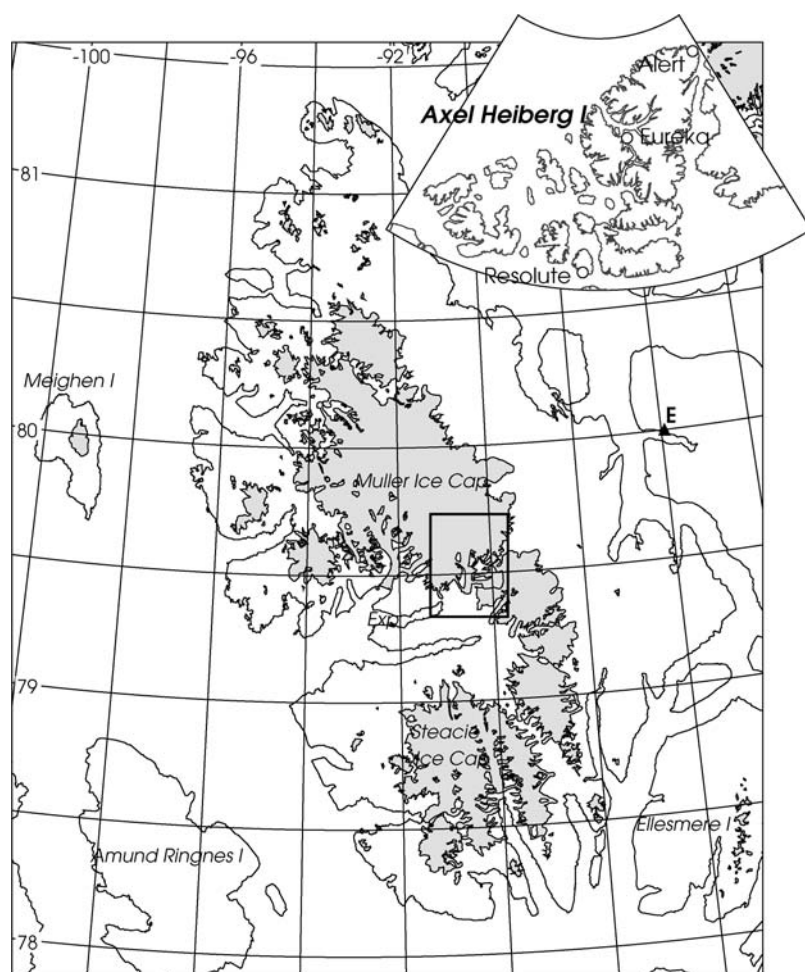
FIGURE 1. *Axel Heiberg Island in the context of the Queen Elizabeth Islands (inset). Glaciers are shaded. The rectangle in the center of the island encloses our study area. Exp: Expedition Fiord. E: Eureka. Universal Transverse Mercator projection, zone 15.*

the present work: the maps at scales of 1:10,000, covering White Glacier (WG; National Research Council, 1965); 1:50,000, covering the Thompson Glacier (TG) region (National Research Council 1962); and 1:100,000 (McGill University, 1963), covering the Expedition Fiord (EF) area more broadly. The 10K DEM is derived from the WG map, the 50K DEM from the TG map, and the 100K DEM from the EF map. In routine work we use a DEM which is a composite of these three basic DEMs, but the comparisons presented below are based entirely on areas where, by design, the basic DEMs overlap and provide separate estimates of elevation (see figure caption).

The ground survey and photographic program which were the basis for mapping are described in detail by Cogley and Jung-Rothenhäusler (2002), who also discuss positional errors in the maps.

The 10K DEM (contour interval c of the parent map equal to 10 m) and 50K DEM ($c = 25$ m) are independent estimates of the state of the terrain. The 10K DEM, based on 1960 photography, is one year later than the 50K DEM, but the mass balance of White Glacier measured for that year (Cogley et al., 1996) was the equivalent of −0.45 m of ice, which is small compared to the between-map differences discussed below. The 100K DEM ($c = 100$ m) is based on the TG map over the extent of the latter and on 1960 photography in its northern portion. The 10K DEM has horizontal resolution of 50 m. The horizontal resolution chosen for the 50K and 100K DEMs was 100 m.

250K DEM

The National Topographic Series 1:250,000 scale topographic map of the Expedition Fiord area was drawn from the same aerial photographs as those used in preparation of the TG and EF maps (Department of

Energy, Mines and Resources, 1967). Its contour interval is 152 m (500 feet). A digital vector dataset, including contours, is published as part of NTDB, the National Topographic Data Base (Centre for Topographic Information, 1999), and the raster topography derived from the contours is published as the digital elevation model CDED (Centre for Topographic Information, 2000). The CDED DEM, referred to here as "250K", has meridional resolution of 3 arcsec and zonal resolution of 6 arcsec, or 92 m by 34 m. Both NTDB and CDED are referenced to the NAD83 horizontal datum and to Mean Sea Level as defined in the Canadian Vertical Geodetic Datum.

The accuracy of CDED and NTDB is not specified in detail. NTDB "aims at attaining" positional accuracy of 125 m (Centre for Topographic Information 1999), while "in some NTDB data sets, horizontal inaccuracy goes up to ... 500 metres" (Centre for Topographic Information, 2000). No numerical estimates are provided for the accuracy of CDED elevations.

1M DEM

At a still smaller scale the Digital Chart of the World (DCW; Defense Mapping Agency, 1992) and the related GLOBE DEM (Hastings et al., 1999) cover Axel Heiberg Island at useful resolution (30 arcsec for GLOBE, or meridional and zonal resolutions of 930 m and 170 m). In the Canadian High Arctic, GLOBE derives from Operational Navigation Charts, which are 1:1,000,000 scale maps based on Corona satellite imagery of the early 1960s with typical scales of 1:300,000 to 1:400,000. The contour interval is 305 m (1000 feet) with a subsidiary contour at 152 m (500 feet).
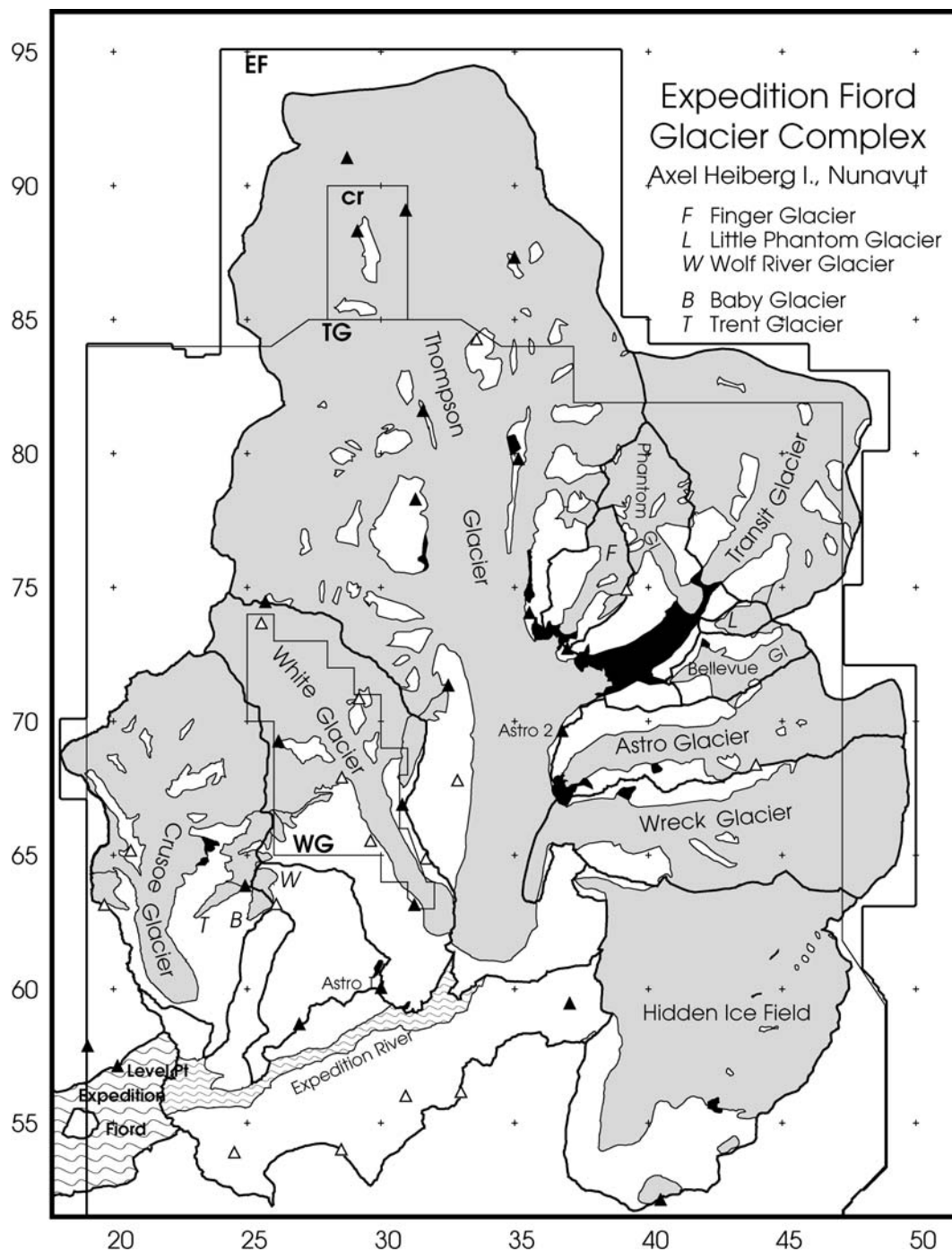
FIGURE 2. The Expedition Fiord area. Glacier ice, covering about two thirds of the terrain, is shaded; lakes are black. Solid, open triangles: primary and selected secondary control points of the triangulation net. The 10K DEM covers the glacierized portion of the White Glacier catchment. The 50K DEM covers the area labelled "TG". Within the area labelled "WG", estimates in the 50K DEM are samples every 100 m from the 10K DEM. These estimates are excluded in comparisons of 10K and 50K elevation, which are thus drawn from the periphery of the White Glacier catchment. The 100K DEM covers the area (labelled "EF") north and east of TG. Comparisons of 50K and 100K elevations are drawn from areas of planned overlap (not shown) between TG and EF. The 250K and 1M DEMs cover the entire Expedition Fiord area. cr: comparison region covered by Figure 6. Grid labels (in km) are arbitrary local coordinates, with control point Astro 1 assigned the easting and northing (30,000, 60,000) meters.

GLOBE is referenced to the WGS84 horizontal datum and to Mean Sea Level. "Absolute horizontal accuracy of the DCW hypsography is reported to be 2040 m rounded to the nearest 5 m at 90% circular error. Vertical accuracy is considered to be 610 m for contours, and 30 m for spot elevations" (Hastings et al., 1999). We refer to GLOBE as the "1M DEM" below.

Measurement Methods

MEASUREMENTS BY EYE

The 10K DEM is a digital version of the topography shown on the WG map, with accompanying landcodes (descriptors which distinguish between water bodies, land and ice). Relying on contours and on printed spot heights and lake surface elevations, we read elevations on the map to the nearest meter by eye, using a transparent gridded overlay. We adopted this old-fashioned method because we could not afford more modern resources, but we note that it enables us to evaluate uncertainty by the equally old-fashioned method of replicating our measurements.

We found that, no matter how carefully the work was done, systematic proofreading of the entire DEM by two persons was essential in generating a correct product. One person read from a printout while the second verified the readout against the map. Colored elevation and landcode maps, and a shaded-relief image, were also valuable proofreading aids. Proofreading is analogous to the "cleaning" process to which automatically generated DEMs must be submitted.

J. G. COGLEY AND F. JUNG-ROTHENHÄUSLER / 251

TABLE 1

TABLE 1

Replicate estimates of elevation (White Glacier)

Difference (m)	Raw measurements			After proofreading		
	$f_1 - f_0$	$g_1 - g_0$	Diff	$f_1 - f_0$	$g_1 - g_0$	Diff
Minimum	−10.00	−9.00	−16.00	−5.00	−4.00	−8.00
Mean	0.09 ± 1.00	-0.10 ± 0.63	0.28 ± 1.59	0.06 ± 0.58	0.05 ± 0.46	0.02 ± 0.92
Rms	2.44	1.30	3.21	0.95	0.86	1.29
Maximum	25.00	8.00	23.00	6.00	4.00	6.00

Subscripts 0 and 1 refer to the first and second elevation measurements of each of the two map readers f and g. The "Diff" columns show, in each case, the greater in magnitude of $f_0 - g_0$ and $f_1 - g_1$. For means, error ranges are $\pm$ twice the average standard error ($\Sigma_i \Delta_i / 2)/N$, Δ being the appropriate difference. The 50 m resolution of the 10K DEM yields a sample size $N = 400$. Rms: root mean square, defined as $[(\Sigma_i \Delta_i^2)/N]^{1/2}$.

Most of the 50K DEM was prepared in the same way as the 10K DEM. The nucleus of the 50K DEM was a subsample of the 10K DEM at 100-m resolution. As work progressed some of this subsample was replaced by estimates from the TG map, yielding replicate elevation estimates and allowing for between-map comparisons.

AUTOMATED INTERPOLATION

The EF map is more generalized and more poorly controlled than the TG map. For the 100K DEM, therefore, we developed a two-stage algorithm for interpolation of elevation estimates from digitized contours, formlines and spot heights. In the first or coarse stage, each point is assigned the average elevation of its upper and lower bounding contours. In the second, refinement stage, the band-midpoint elevation at each point is adjusted by adding a distance-weighted sum of nearby elevations at known distances from the grid point: the nearest points on the lower and upper bounding contours and any sufficiently close spot heights within the elevation band. A simple inverse-square-root rule was selected for the decay of weight with distance. Elevations more distant than 2205 m, representing a fractional weight of about 0.02, were excluded. Results presented below show that the 100K interpolation algorithm yields very satisfactory results.

PUBLISHED DEMS

The 250K DEM is interpolated using the ANUDEM software (Hutchinson, 1989, 1996). ANUDEM uses thin-plate spline interpolation, with an automatic constraint which maximizes the smoothness of the interpolated terrain surface. It is designed to incorporate information not only from spot heights and contours but also from stream lines and ridge lines. This tends to enhance the hydrological realism of the output DEM, as does a procedure for infilling of the spurious depressions created by the spline algorithm. For comparison of the 250K and 50K DEMs, 50K grid point positions were recast into WGS84 and their 250K elevations assigned from the nearest CDED elevation. It was not thought profitable to investigate possible differences between the vertical datum of 250K and the local vertical datum of the 10K, 50K, and 100K DEMs.

Like the 250K DEM, the 1M DEM is interpolated from vector information using ANUDEM, in this case the vector source being DCW. 1M elevations were resampled to the 50K DEM grid as for the 250K elevations.

Error Model

Suppose that at a point whose location is known exactly the true elevation is H, and that

$$H_a = H + B_a \pm \delta_a \tag{1a}$$

$$H_b = H + B_b \pm \delta_b \tag{1b}$$

are two estimates of H found on different maps, a and b. B_a and B_b are unknown biases and are assumed, of necessity, to be zero. δ_a and δ_b are random error terms of mean zero and variances equal to m_a^2 and m_b^2, respectively. m_a and m_b are the expected values of the mapping error for each map. (Note that they include errors of horizontal registration between the maps, which we cannot estimate explicitly.) Like H, H_a and H_b are unknown; we have to read the maps in order to obtain numbers h_a and h_b with which we can work. This gives rise to the map-reading errors, r_a and r_b, as follows:

$$h_a = H_a \pm \varepsilon_a \tag{2a}$$

$$h_b = H_b \pm \varepsilon_b, \tag{2b}$$

where ε_a and ε_b are random error terms of mean zero and variances equal to r_a^2 and r_b^2, respectively. From (1) and (2), a measure of the total error is

$$h_a - h_b = \pm \delta_a \pm \delta_b \pm \varepsilon_a \pm \varepsilon_b. \tag{3}$$

We can now identify the expected value Γ_{ab} of $h_a - h_b$ with a between-map rms elevation difference such as one of those in Table 5 below, and r_a, r_b with between-reader rms differences such as those in Tables 1 and 4. Γ_{ab}, which is measurable, is called the comparison error. Assuming that the various errors are independent,

$$\Gamma_{ab} = \sqrt{m_a^2 + m_b^2 + r_a^2 + r_b^2}. \tag{4}$$

Separating known terms from unknown, (4) yields the geometric sum of the two mapping errors.

$$m^2 = m_a^2 + m_b^2 = \Gamma_{ab}^2 - r_a^2 - r_b^2. \tag{5}$$

We must now make some assumption about the relationship between the two components of m. Two possibilities are

1. lumped error: one mapping error is zero and the other is equal to m;
2. equally distributed error: m_a and m_b are equal and each map has a mapping error of $m/\sqrt{2}$.

Given such an assumption, the total errors in the two maps are

$$\gamma_a = \sqrt{m_a^2 + r_a^2} \tag{6a}$$

$$\gamma_b = \sqrt{m_b^2 + r_b^2}. \tag{6b}$$

The mapping error characterizes errors in the placement of contours on the maps. We can only learn about it indirectly, by comparing one map with another. The map-reading error characterizes errors in extrapolating from given contour information. We can learn about it by

replicate sampling on one map at a time. The total errors can be estimated, but only if (5) can be solved.

Our error model, although it was developed for this study, simply implements ideas which are standard in the subject of error analysis (e.g., Bevington, 1969). Note that map-reading errors may be made either by human readers, who interpolate by eye, or by computers, which interpolate algorithmically.

Map-Reading Errors

For both mapping and map-reading errors, our estimation procedure is based on replication, that is, repeated measurement of the true elevation H. Table 1 illustrates replicate measurements of the WG map. Two readers each measured elevations twice on a 50-m grid covering a 1×1 km test area. One of these four sets of measurements eventually became part of the DEM. The other three are the replicates. The formal error range for the mean should be seen as a rough indicator of the likelihood that the difference reveals a bias. All of the means are small with respect to their error ranges, so the visual map-reading method yields estimates of elevation which are not affected by human bias. Table 1 also shows that there was no drift in the work of the two readers over time and no difference between their respective estimates. Mean differences were less in magnitude after proofreading than before. Likewise the minima and maxima grew smaller after proofreading, as a result of the elimination of gross errors.

In the absence of any reason to believe that one member of a difference is closer to the truth than the other, and of any evidence of bias, we can regard each pair of readings as a pair of random samples. If the N sample pairs are independent then the rms difference is an appropriate estimate of error, for it is proportional to the uncertainty in any single future sample of elevation. The uncertainty, however, is due solely to the reading of the map. Table 1 contains no information about the fidelity of the map to the terrain. Its value lies in showing that, for the 10K DEM, map-reading errors are of the order of 1 to 3 m before proofreading and 0.8 to 1.3 m after.

Figure 3 illustrates the frequency distribution of one set of between-reader elevation differences ($f_1 - g_1$ before proofreading; Table 1). The accompanying histogram for a normal distribution having the same mean and variance demonstrates that the observed differences are not normally distributed. In general, all of the difference distributions in this study fail the standard tests for normality, such as the Kolmogorov-Smirnov test. This means that one of the usual assumptions made when quantifying error estimates is violated. The rms difference between any two related estimates of elevation or slope is still proportional to the standard error, but we cannot make the usual inference that there is a 95% probability that the true quantity lies within 2 rms differences of the estimated quantity. The actual probability will be somewhat less.

The most likely explanation for the nonnormality of the observed differences is that they represent the sum of more than one population of errors. Several normally distributed sources of random error, each of different variance, will add up to a distribution which in general will not be normally distributed. In addition, some kinds of error are themselves unlikely to be normally distributed.

The other main assumption on which statistical estimates of uncertainty are usually based is that of independence of the samples. A spatial autocorrelation analysis was undertaken in the replicate cell studied above and in a second 1×1 km cell. Each two-dimensional array of elevation differences was correlated against copies of itself offset by varying distances (lags) in the easterly and southerly directions. All of the correlograms showed exponential distance decay. The decorrelation distances ranged from 44 to 248 m before and from 49 to 179 m (about 1–3 resolution elements) after proofreading. Adjacent elevation differences, and therefore presumably errors, are
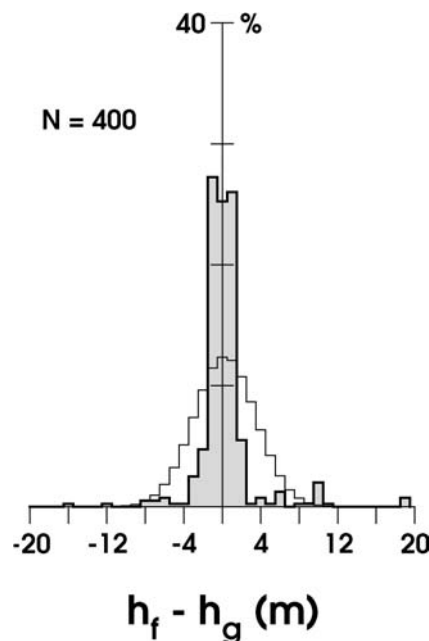

FIGURE 3. Frequency distributions of the difference in elevation estimates h made on the WG map by readers f and g in the 1×1–km cell with southwest corner at grid reference (29000,66000). Shaded histogram: observed differences before proofreading. Unshaded histogram: normal distribution having same mean and variance as shaded histogram.

moderately correlated rather than independent, and mistakes tend to occur in clusters. This is another reason why the rms difference is a somewhat generous estimate of uncertainty.

Table 2 compares the work of map readers who contributed to the 50K DEM. The frequency distributions of between-reader differences appear similar to that shown in Figure 3, but most of the error measures are larger in magnitude than those obtained for the 10K DEM. There is, however, no evidence that the map readers are biased one with respect to another. Minimum and maximum differences have magnitudes in the dekametre range, and rms differences are 4–8 m. Map reading for the 50K DEM was thus less accurate than for the 10K DEM, perhaps because proofreading was less thorough. It was not possible to devote the same proportion of effort to proofreading in the larger-scale 50K project as in the 10K project.

The finite-difference scheme used to estimate the slope at a point involves the elevations at its eight neighbors. It would be possible to calculate the error in the slope given the errors in these elevations, but knowing that the latter errors are correlated it is safer to rely on empirical estimates based on differences in observed slopes. In Table 2, mean slope differences are small and unbiased, there is no difference between the estimates from the different map readers, and the differences range up to 4° in magnitude. Rms differences are 0.4° to 0.9°. As for elevations, the slope statistics cannot be interpreted as confidence intervals. For slopes, the situation is aggravated by the dependence of adjacent estimates on each other. Each slope estimate shares three elevations out of eight with each of its neighbors.

We cannot estimate map-reading errors for the 250K and 1M DEMs because we have no replicate readings for either. This is a neglected advantage of the more time-consuming methods used for the 10K and 50K DEMs.

Mapping Errors

Where replicate estimates of elevation from two maps are available, we can compare the elevations and slopes and estimate the mapping error

J. G. COGLEY AND F. JUNG-ROTHENHÄUSLER / 253

TABLE 2

Between-reader differences (Thompson Glacier)

Difference	Elevation			Slope		
	$M - C$	$M - G$	$M - S$	$M - C$	$M - G$	$M - S$
Number	300	400	699	192	272	479
Minimum	−44.00	−62.00	−28.00	−1.33	−3.83	−2.40
Mean	0.26 ± 2.56	−1.89 ± 4.42	0.48 ± 4.32	−0.04 ± 0.31	−0.02 ± 0.55	−0.04 ± 0.44
Rms	4.65	7.81	6.44	0.45	0.86	0.64
Maximum	26.00	14.00	35.00	2.35	3.99	3.30

Elevation: Differences (m) between elevations measured by map readers M, C, G and S at coincident locations on the TG (50K) map. Slope: Differences (deg) in slopes computed from the work of the different readers.

according to the error model described above. Coincident elevations from two maps need not have equal standing. Given comparably careful work the larger-scale map should be more accurate. On the other hand, neither map can be regarded as exact. In general it would be a mistake to assign all of the random error to the smaller-scale map.

10K COMPARED TO 50K

Figure 4 and Table 3 summarize the distributions of differences in elevation and slope at locations common to the 10K and 50K DEMs. The magnitudes are all larger than in previous comparisons, but evidence for the unbiased character of the measurements remains as good as formerly. The rms differences, about 20 m in elevation and 2° in slope, point to substantially greater total uncertainty than in either DEM alone.

To apply the error model to the 10K–50K comparison, we take as estimates of the map-reading errors the medians of the rms differences from Table 1, $r_W = 0.95$ m, and Table 2, $r_T = 6.44$ m. From Table 3, column a gives a comparison error of $\Gamma_{TW} = 19.71$ m. Using (5), the total mapping error is $m = 18.60$ m, so it is the dominant contributor to total error. Map reading has introduced only a small extra uncertainty. If we make the lumped-error assumption, and for the sake of illustration assign all of the error to the 50K DEM, then from (6) total error is $\gamma_T = 19.68$ m in 50K and $\gamma_W = 0.95$ m in 10K. The equable-error assumption gives $\gamma_T = 14.65$ m in 50K and $\gamma_W = 13.19$ m in 10K.

The elevation distribution in Figure 4a has a striking positive tail. This tail comes almost entirely from a small area where the maps differ radically. The area is a featureless white expanse on the photographs used for the 50K DEM, while contrast is much better in adjacent 10K photographs. It is therefore likely that the 50K DEM is wrong. Of more general concern, inspection of elevation differences over the area shows that 129 of the differences can be regarded as gross errors, which is 2.4% of the 2×2695 elevations available, or one in every forty. The effect of these blunders can be seen numerically in Table 4.

In both rows of Table 4 the mapping error m is the largest contributor to total error γ. Thus map-reading, while it has added some uncertainty to the information from the map, is only moderately culpable.

50K COMPARED TO 100K

Figure 5 and Table 5 compare estimates read visually from the TG (50K) and EF (100K) maps. The histogram of elevation differences (Fig. 5a) is noticeably skew. Consistent with the relatively large mean difference, there is an excess of negative differences (100K higher than 50K) and a negative tail which is reminiscent of the positive tail in Figure 4a. It was not possible to localize the negative tail, however. The slope error grows more slowly between the 50K:10K and 50K:100K comparisons than does the elevation error: an increase by a quarter for slope against a doubling for elevation. This may be because of local autocorrelation of elevation errors.

To estimate the map-reading error r_E of the 100K DEM, visual measurements were made over a comparison region. The three resulting contour sets, showing close but not perfect agreement, are illustrated in Figure 6. The distribution of elevation differences is graphed in Figure 7 for the two stages of interpolation. Visual estimates and coarse interpolates differ roughly uniformly over a range comparable to the contour interval. The difference between visual
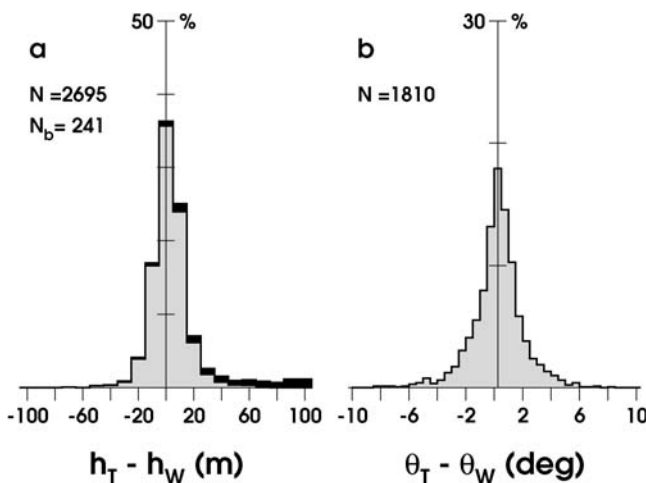

FIGURE 4. Frequency distribution of differences in coincident elevation (a) and slope (b) estimates from the 50K and 10K DEMs. In panel a the upper black portions of the bars represent the N_b elevation differences from a locality where the two maps differ radically.

TABLE 3

Between-map Differences (Thompson Glacier—White Glacier)

Difference	Elevation[a]	Elevation[b]	Elevation[c]	Slope
Number	2695	241	2454	1810
Minimum	−72.00	−43.00	−72.00	−8.50
Mean	5.56 ± 11.66	40.44 ± 43.34	2.13 ± 8.54	−0.02 ± 1.26
Rms	19.71	52.76	12.39	1.80
Maximum	99.00	99.00	66.00	8.01

Differences between measured elevations (m) and computed slopes (deg) at coincident locations on the TG (50K) and WG (10K) maps. a: all available elevations. b: elevations from a small area where the two maps differ radically. c: elevations excluding those in b.

<div style="display:flex">

TABLE 4

Errors in Thompson Glacier and White Glacier elevations

Elevations	Γ_{TW}	m	Lumped γ_T	Lumped γ_W	Equable γ_T	Equable γ_W
All	19.71	18.60	19.68	18.62	14.65	13.19
Gross errors excluded	12.39	10.54	12.35	10.58	9.85	7.51

In both rows $r_T = 6.44$ m and $r_W = 0.95$ m. The two rows correspond to elevation columns a and c of Table 3, respectively. Tabulated quantities are in meters.

</div>

TABLE 5

Between-map differences (Thompson Glacier—Expedition Fiord)

Difference	Elevation	Slope
Number	1059	518
Minimum	−92.00	−7.62
Mean	−11.58 ± 18.34	0.04 ± 1.69
Rms	25.05	2.46
Maximum	72.00	13.46

Differences between measured elevations (m) and computed slopes (deg) at coincident locations on the TG (50K) and EF (100K) maps. The work of all map readers was pooled for this comparison.

estimates and refined interpolates is less dispersed, the rms difference being much reduced (Table 6). When normalized by contour interval $c = 100$ m, this between-method difference of 0.13 is comparable to the between-reader differences of 10K and smaller than those of 50K. It is not strictly comparable to those estimates, but it is the only guidance available and so we adopt it as an estimate of the map-reading error for the 100K DEM, $r_E = 13.05$ m.

We can now estimate mapping errors and total errors of the 50K and 100K DEMs. From Table 5, the 50K:100K comparison error was $\Gamma_{TE} = 25.05$ m, and from the last subsection $r_T = 6.44$ m. From (5), then, total mapping error $m = 20.42$ m. In turn this gives a total error for the 100K DEM of $\gamma_E = 24.21$ m if all of the mapping error is lumped onto the EF map, or $\gamma_E = 19.43$ m if the mapping error is equably distributed. Thus, as in the 10K:50K comparison, total error is due largely to mapping error.

250K COMPARED TO 50K

Figure 8a shows the hypsometric frequency distributions of the 250K and 50K DEMs. The 250K DEM has more terrain at the highest elevations, above 1600 m, and less at 1400 to 1600 m. It has significantly more terrain at 50 to 200 m and much less at 0 to 50 m. This is because CDED is unable to resolve the flood plain of Expedition River, where it is too high because it lacks a source of information below the lowest contour at 152 m.

The histogram of 250K is noticeably more jagged than that of 50K, exhibiting a repeating three-bar pattern which suggests a connection with the contour interval, ~150 m.

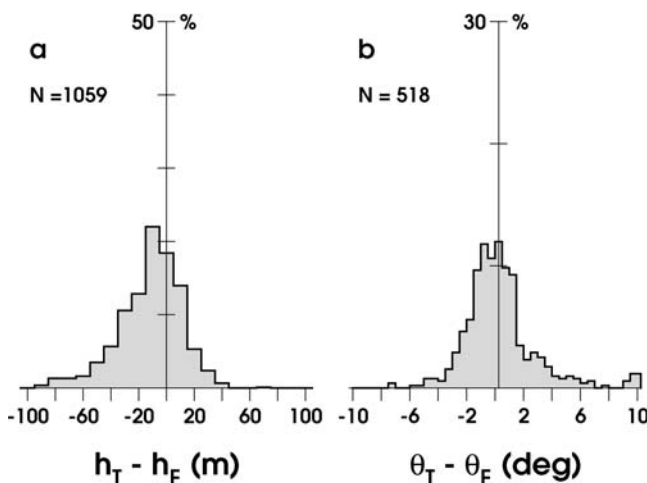Figure 8b, for the frequency distribution of slopes, shows that 250K is smoother and less rugged than 50K. The two DEMs have comparable frequencies of slopes less than 1°. The modal interval is 1 to 2° for both, but 250K has an excess of slopes of 1–5° while 50K has an excess of steeper slopes.

The histogram of elevation differences in Figure 8c is extremely broad. The histogram of slope differences in Figure 8d is more regular. There is a slight but noticeable skewness to the right (50K steeper than 250K), which is consistent with Figure 8b. Neither in elevation differences nor in slope differences, however, is there a significant bias.

Table 7 summarizes the difference distributions numerically. Elevation differences reach several hundred meters, and although the two DEMs are mutually unbiased their rms difference approaches 100 m. The rms difference in slope is not much less than 5°.

FIGURE 5. *Frequency distribution of differences in coincident elevation (a) and slope (b) estimates from the 50K and 100K DEMs.*
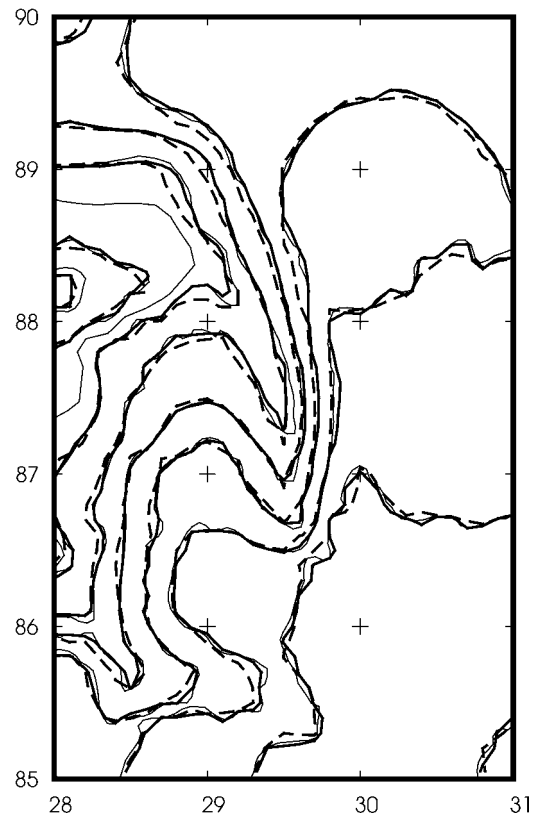
FIGURE 6. *Elevations in comparison region. Thin lines: contours digitized from EF map. Thick solid lines: contours of elevations estimated by interpolation algorithm. Thick dashed lines: contours of elevations measured by eye from map. Contour interval is 100 m, the lowest contour being at 900 m (southeast corner). Extra formline at 1550 m (west center) has no interpolated or measured counterpart.*
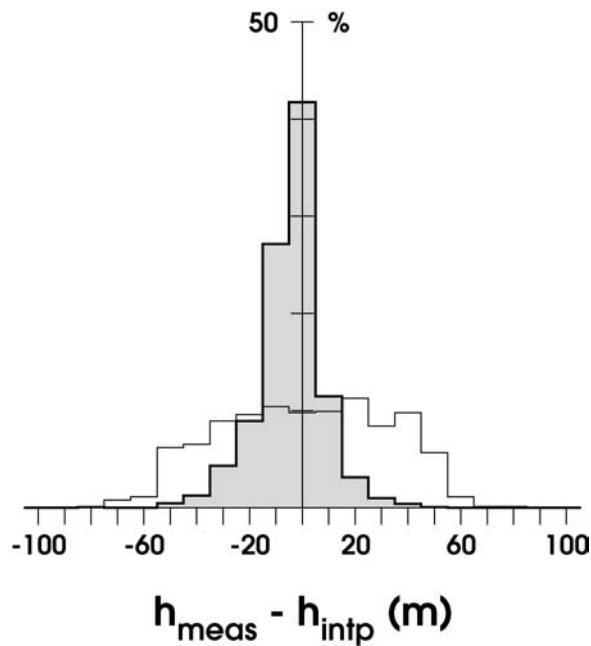
J. G. COGLEY AND F. JUNG-ROTHENHÄUSLER / 255

FIGURE 7. *Frequency distribution of differences in coincident measured and interpolated elevation estimates from the comparison region of the EF map. Unshaded histogram: differences after first, coarse stage of interpolation. Shaded histogram: differences after second, refinement stage.*

1M COMPARED TO 50K

Figure 9 summarizes the comparison of the 50K and 1M DEMs. As with 250K, there is an excess of the highest elevations in 1M, and a similar relative deficit of the lowest elevations (Fig. 9a). The jaggedness of its hypsometric distribution, with modes centred at multiples of its contour interval $c = 1000'$, is more marked than for 250K. 1M has a large relative excess of gentle slopes (Fig. 9b). The modal interval, 0 to 1°, contains 30.1% of all the slope samples, but in large part this is an artefact of oversampling. The low resolution of 1M means that each of its elevations appears roughly 2×9 times at the 100 m resolution of the 50K DEM. The 50K DEM has an excess of slopes in the range 2–26°, but beyond 26° 1M has an excess (all but invisible in Fig. 9b).

50K–1M elevation and slope differences (Fig. 9c, 9d) are still more dispersed than those of 50K–250K. Elevation differences (Table 7, second column) exceed 600 m in magnitude and the rms difference is well over 100 m. Slope differences (Table 7, fourth column) exceed 30°, with an rms difference of 7°. The difference histogram (Fig. 9d) is strongly skewed to the right, 50K being a markedly steeper DEM than 1M. The mode of Figure 9d is at 0.5 to 1.0° but the mean difference is 2°.

The only component of the error model which we can identify unambiguously for 250K or 1M is the comparison error Γ_{TC} or Γ_{TG}. For the present purpose, we adopt a total error $\gamma_T = 20$ m for the 50K DEM. Taking the comparison errors from the rms differences in the first and second columns of Table 7, and using (4) and (6), we obtain the following estimates of the total errors in the 250K and 1M DEMs:

$$\gamma_C = \sqrt{\Gamma_{TC}^2 - \gamma_T^2} = 90.26 \text{ m}; \qquad \gamma_G = \sqrt{\Gamma_{TG}^2 - \gamma_T^2} = 158.36 \text{ m}.$$

Thus, because γ_T is so much smaller than Γ_{TC} and Γ_{TG}, we might as well have assumed it to be zero. The 2-m difference between γ_C and Γ_{TC}, for instance, is unlikely to be of consequence in any working context where error estimates are needed for elevations.

TABLE 6

Coarse and Refined Elevation Differences (Expedition Fiord)

Difference	Coarse	Refined
Number	1500	1500
Minimum	−80.00	−78.86
Mean	0.28 ± 61.80	−4.27 ± 24.66
Rms	30.89	13.05
Maximum	81.00	45.90

Differences (m) are between coincident measured and interpolated elevation estimates from the comparison region of the EF map. Their distribution is shown in Figure 5.

SPURIOUS LOW-DIGIT NOISE IN ELEVATION ESTIMATES

Each DEM has a different contour interval c, and none of the c represents any physiographic length scale. The quantity $\eta = h \bmod c$ should thus be a random variable uniformly distributed between 0 and c. That is, the frequency function $f(\eta) = 1/c$ m^{-1} = const. Observed departures from uniformity (Fig. 10) arise from three contributions: real, physiographic features; random sampling variability; and shortcomings in methodology. Physiographic departures from uniformity of η are possible, and perhaps likely, within any one elevation band, but they should cancel rapidly as the number of bands $(h_{max} - h_{min})/c$ (Table 8) increases. We therefore assume that they can be neglected. We further assume that sampling variability appears as the high-frequency variation in the graphs (the "wiggles"). These variations are small. Their standard deviation is about 10^{-2} m for the 10K DEM and considerably less for the others. Except possibly for 10K, then, sampling variability is also negligible.

Each observed distribution is distinctive and peculiar. In 10K, the units digit 0 occurs more often than expected. For 50K, the four map readers who contributed to panel b have quite different distributions of η (not shown), and each reader has evidently adopted a different psychovisual interpolation algorithm; in particular, the sharp changes near $c/3$ and $2c/3$ in Figure 10b are accidental rather than due to a flaw common to all of the readers. The interpolation algorithm of 100K (panel c) tends to avoid elevations just above contours, and perhaps to favor elevations somewhat below the midpoint elevation at the expense of elevations somewhat above. Both 250K (panel d) and 1M (panel e) have too many elevations near to contours and too few in between. (The distributions plotted here are those of the resampled DEMs, but the distributions of the parent DEMs in geographical coordinates are equivalent.) The multiple spikes at both ends of 1M's range, and the more subdued pattern at the low end of 250K's range, are not understood.

We estimate a typical value for uncertainty due to algorithmic defects as follows. If η is in truth uniformly distributed, elevations must be in error by an amount sufficient to restore the observed distribution to uniformity. Exact uniformity is not realistic, but it is an objectively defined state which is reachable readily by reassigning the observed departures. We did this for each of the DEMs, minimizing the magnitude of the difference $\Delta\eta$ for each reassignment. Each negative departure from the expected $1/c$ was eliminated by drawing upon the nearest positive departure to its left (lower η). Thus

$$r_{algo} = \frac{\sum_\eta n\Delta\eta}{N} \qquad (7)$$

where n is the number of elevations reassigned by the amount $\Delta\eta$ and N is total sample size.

Table 8 shows that in general r_{algo} makes a moderate contribution to map-reading error r. For 10K it is equal to about 30% of r, and for 50K and 100K it is less than 5% of r. For 250K and 1M, r_{algo} is small
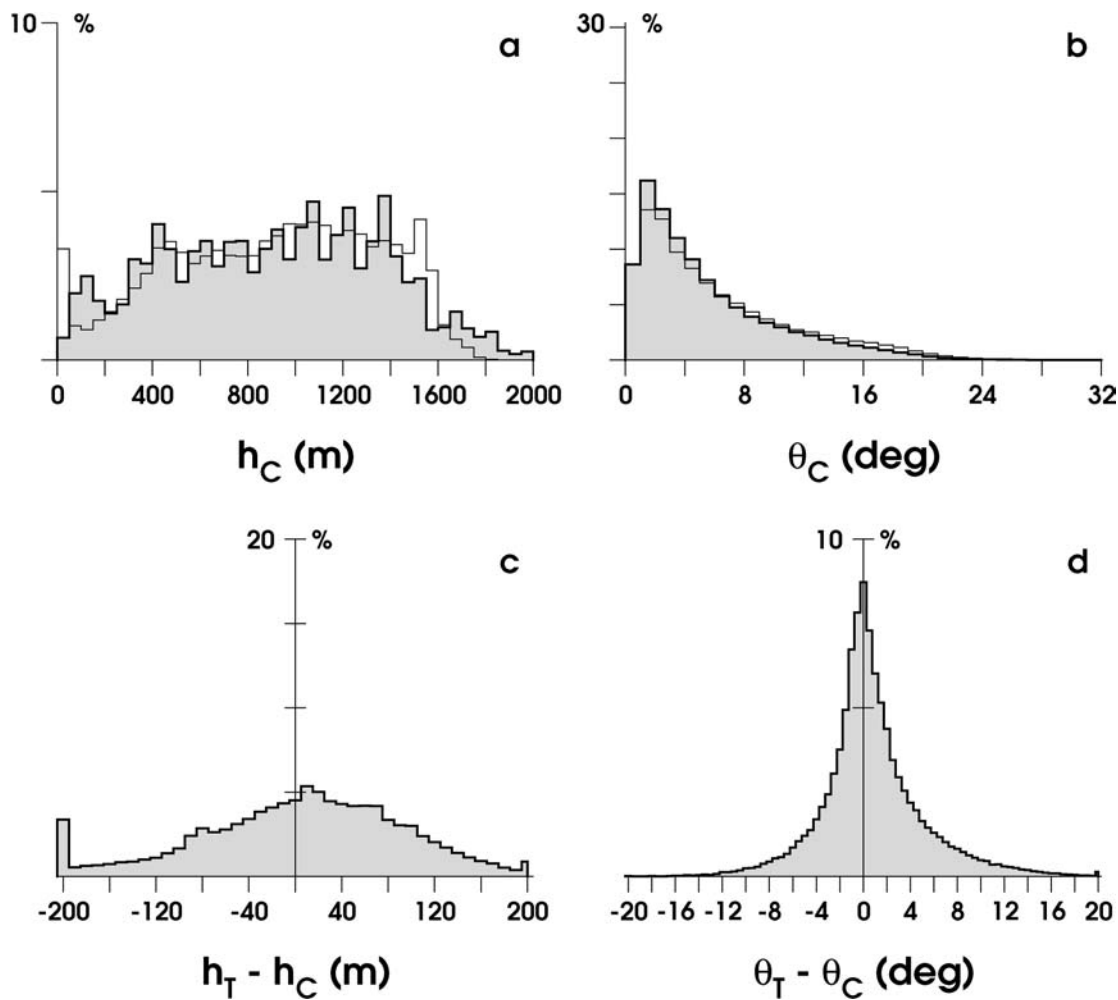
FIGURE 8. Distributions of elevation and slope in the 250K DEM. Elevations were transferred to the transverse Mercator grid of the 50K DEM by nearest-neighbor resampling. Sample size is of the order of 110,000 in each panel. a: Elevation (shaded histogram), with the hypsometry of 50K overlaid as the unshaded histogram. b: Slope (shaded histogram), with the slope distribution of 50K overlaid as the unshaded histogram. c: 50K elevation minus 250K elevation. d: 50K slope minus 250K slope.

by comparison with total error. Across the five DEMs it is equal to between 1 and 5% of the contour interval. Thus, in practice, the departures from randomness seen in Figure 10 are not the principal reason for being uncertain about DEM elevations.

Discussion

CONFIDENCE INTERVALS FOR ELEVATION ERRORS

The errors discussed here are not quantified in terms of confidence intervals. If certain conventional assumptions are satisfied, and if one of two compared DEMs represents the truth, the between-map rms difference represents the standard error of elevation, defining a ~68% confidence interval, in the untruthful DEM. The assumptions are (1) that each DEM consists of independent random samples which (2) are all drawn from the same probability distribution, and (3) that together the DEMs have a bivariate normal distribution. Each of these assumptions is violated in the DEMs we have studied, although more work would be needed to assess how serious the violations are.

The most confidence-sapping of the problems in our analysis, however, is the persistence of blunders (that is, gross errors). The only way to detect them is to sample the topography more than once, hoping that the same blunder will not be made in all of the replicate samples. But in general one must accept that undetected blunders can increase the uncertainty by 50% or more, and that they frustrate all conventional ways of representing the real error structure. In Table 4, for example, ignoring blunders makes the presentation incomplete, while including them inflates the error estimates for the great majority of blunder-free

elevations. We suggest that, where possible, attempts should be made to quantify the blunder rate in any investigation which relies substantially on DEM elevations. When an elevation estimate may be a blunder, conventional confidence intervals have little meaning. The map user needs to know separately the (blunder-free) total error and the probability of a blunder. Although our detected blunders happen to be peculiarly glaciological, we think it would be naive in any context to *assume* that the probability of a blunder is zero.

SYNTHESIS

Figure 11 summarizes the total errors γ estimated for the five DEMs in earlier sections. They can be explained well as a linear function of the contour interval *c*. Whether we exclude the 100K errors or not (see figure

TABLE 7

Between-DEM elevation and slope differences

Difference	Elevation		Slope	
	50K–250K	50K–1M	50K–250K	50K–1M
Minimum	−412.00	−618.00	−25.44	−34.85
Mean	3.49 ± 71.74	−30.45 ± 122.96	0.73 ± 3.05	2.01 ± 4.99
Rms	92.45	159.62	4.44	6.95
Maximum	377.00	682.00	28.28	32.80

Differences (m for elevations, deg for slopes) are between coincident estimates from the three DEMs. See also Figures 8 and 9.
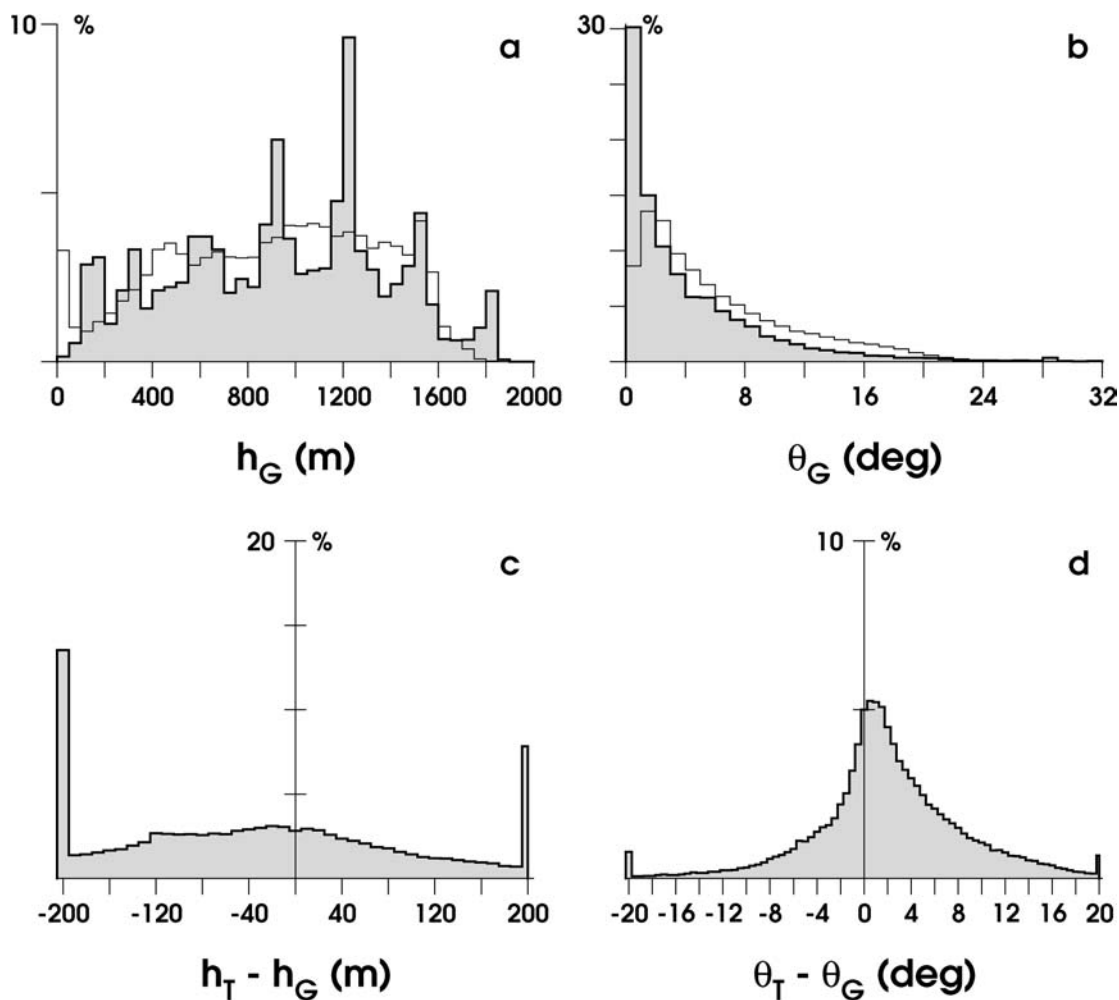
J. G. Cogley and F. Jung-Rothenhäusler / 257

FIGURE 9. Distributions of elevation and slope in the 1M DEM. Elevations were transferred to the transverse Mercator grid of the 50K DEM by nearest-neighbor resampling. Sample size is of the order of 110,000 in each panel. a: Elevation (shaded histogram), with the hypsometry of 50K overlaid as the unshaded histogram. b: Slope (shaded histogram), with the slope distribution of 50K overlaid as the unshaded histogram. c: 50K elevation minus 1M elevation. d: 50K slope minus 1M slope.

caption), we find that the total error is one half of the contour interval of the parent map. It would be interesting to know whether there is any deeper reason why this is so. It is an estimate which is often seen, but it is usually clear from the context that it is adopted as a convenience. Figure 11 represents moderately strong empirical justification, but there is no obvious starting point for a theoretical justification. The violation of

statistical assumptions and the probable presence of blunders are roadblocks to further reasoning at present.

This result may shed light on the conditions under which one map of a comparison pair may be regarded as exact. If two maps have contour intervals c_a, c_b, with $c_a < c_b$, map a is practically exact if $100(c_a/c_b)^2$ is a small percentage. For example it is 16% for 10K:50K
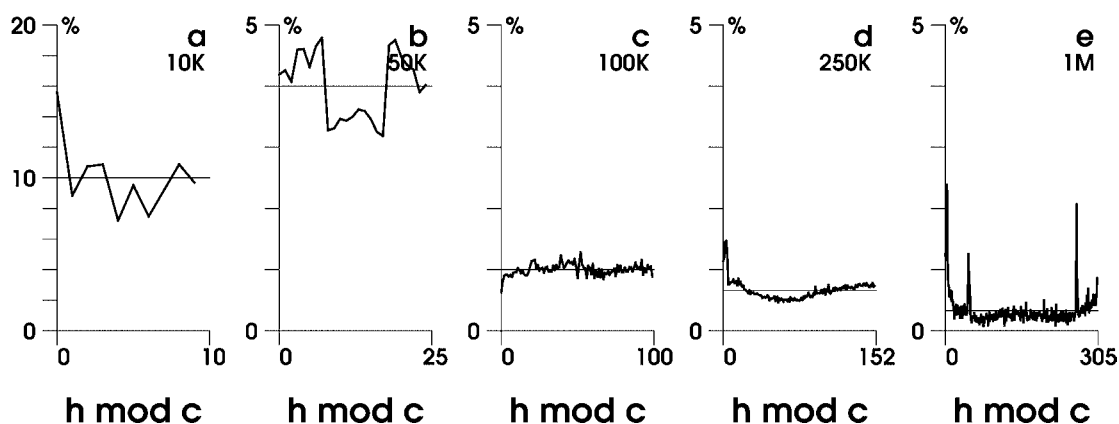

FIGURE 10. Distributions of $\eta = h \bmod c$ (mod signifies the remainder). The histogram expected on the hypothesis that η is uniformly distributed is shown as a horizontal line. a: 10K DEM; c = 10 m. Scale on frequency axis differs from that of other panels. b: 50K DEM; c = 25 m. c: 100K DEM; c = 100 m. d: 250K DEM; c = 500'. e: 1M DEM; c = 1000'.

TABLE 8

Attributes of the DEMs

Quantity	10K	50K	100K	250K	1M
c	10.0	25.0	100.0	152.4	304.8
N	21 710	85 596	21 894	111 526	111 526
$1/c$	0.10	0.04	0.01	0.007	0.003
$(h_{max} - h_{min})/c$	178	74	10	14	7
r_{algo}	0.28	0.29	0.61	3.76	16.01

c: contour interval (m). N: sample size. $1/c$ (m^{-1}): expected constant frequency $f(\eta)$ if η is uniformly distributed. $(h_{max} - h_{min})/c$: number of contour-bounded elevation bands contributing to estimates of $f(\eta)$. r_{algo} (m): map-reading error due to algorithmic defects, estimated using equation (7).

comparisons, which is not small, while it is 2.7% for 50K:250K comparisons, which probably is small.

Table 9 distills error estimates from earlier tables. For each DEM, map-reading error r is a best-case and γ(lumped) a worst-case estimate. Considering the uncertainty in the other estimates of uncertainty (and ignoring the question whether they underestimate the 100K errors), the rightmost column, $c/2$, offers what is probably the simplest choice among these alternatives for working purposes.

By definition, random errors are those which average to zero over many replications. This is a clear concept when applied to visual map-reading. In automated spatial interpolation, which is more usual, there is no such thing as replication because the algorithm yields the same answer every time. The meaning of "random" then becomes moot and, if geodetic surveys in the field are ruled out by their cost, it becomes difficult to quantify the errors in the DEM. One might use several different algorithms to generate replicates from the same source map, but this too may be impractical and it entails assuming that the algorithms are equally inaccurate. Such an exercise is of no value for estimating the accuracy of any single algorithm. Visual map-reading may offer a partial solution. Visual measurements of small test areas, if possible on maps of larger scale than that of the DEM source map, can yield the replicates which automated methods are unable to provide for themselves. Visual and automated algorithms are very different, but the evidence presented above suggests that they are indeed comparably inaccurate. We would not willingly repeat the labor of generating entire DEMs by eye, but nor would we wish to be without the understanding of map-reading errors which we have gained by it.

Summary

The map-reading errors of our visual measurements are unbiased, but they are not normally distributed and exhibit some mutual dependence. They are reduced markedly by careful quality control. The map-reading error is a small fraction of the contour interval of the parent map. Total error in DEM elevations is due mainly to mapping error and only secondarily to map-reading error.

Blunders, probably at the photogrammetric stage, have a major impact on uncertainty. This conclusion is not surprising, but replication allows us to quantify it weakly. We find a blunder rate of the order of one per forty elevation estimates, with the blunders strongly clustered. The rate cannot be generalized, but there can be no guarantee that it is zero on any given map.

In two small-scale DEMs, contour-interval artefacts are striking in hypsometric curves and in graphs of the quantity $h \bmod c$. These artefacts are however responsible for only modest proportions of total error.

Blunders aside, total error in DEM elevation can be explained well as a linear function of the contour interval of the parent map. The total error appears to be one half of the contour interval. This finding might repay further theoretical investigation. Meanwhile it may be regarded as a useful rule of thumb, offering some empirical guidance
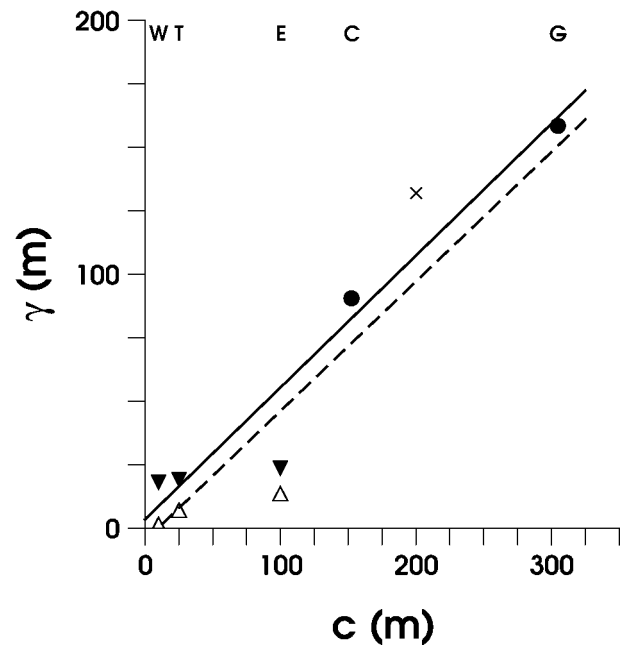
FIGURE 11. *Total DEM elevation errors as a function of the contour interval of the parent map. For 10K, 50K, and 100K, errors are shown for two assumptions: that total error is equal to the map-reading error (upward open triangles), and that total error is the map-reading error plus all of the mapping error identified in a between-map comparison (downward solid triangles). Dashed line: $\gamma = a + b\,c$; the intercept a is -4.9 ± 9.9 m and $b = 0.510 \pm 0.150$ is the slope. Solid line: a similar expression obtained when the 100K errors are excluded because they seem to be outliers; here $a = 3.4 \pm 4.8$ m and $b = 0.520 \pm 0.069$. Cross: a comparable estimate of DEM error from Liu et al. (1999).*

for DEM users who have nowhere else to turn. The information about errors which is supplied with published DEMs is often unsatisfactory. Sometimes, as for the 250K DEM, no information is supplied at all. The error estimates for 1M turn out (we presume) to be worst-case estimates, and ours are dramatically better.

Errors in calculated slopes grow more slowly with the contour interval than do elevation errors, but it is yet harder to quantify confidence intervals for slope errors. Maps of larger scale have consistently steeper slopes, a finding which can be explained by map generalization. On maps of smaller scale, more and more details of terrain structure are left out, and contours become increasingly inadequate representations of the real detail while no method of interpolation is capable of re-creating it. Further work is needed on how to correct this slope bias, for simple additive corrections will worsen the problem. The bias will be present in any map of a scale which does not resolve surface roughness elements.

TABLE 9

Candidate estimates of DEM elevation error

DEM	r	γ (equable)	γ (lumped)	$c/2$
10K	1	13	19	5
50K	6	15	20	12.5
100K	13	19	24	50
250K	—	(90)	(90)	76
1M	—	(158)	(158)	152.5

The various estimates are of total error in m. r: total error equal to map-reading error; the parent map is exact. Not available for 250K, 1M. γ (equable): mapping error shared equally with a comparison DEM. γ (lumped): mapping error of a comparison DEM deemed to be zero. $c/2$: one half of the contour interval of the parent map.

J. G. COGLEY AND F. JUNG-ROTHENHÄUSLER / 259

Acknowledgments

References Cited

Aðalgeirsdóttir, G., Echelmeyer, K. A., and Harrison, W. D., 1998: Elevation and volume changes on the Harding Icefield, Alaska. *Journal of Glaciology,* 44(148): 570–582.

Arendt, A. A., Echelmeyer, K. A., Harrison, W. D., Lingle, C. S., and Valentine, V. B., 2002: Rapid wastage of Alaskan glaciers and their contribution to rising sea level. *Science,* 297: 382–386.

Bamber, J. L., Ekholm, S., and Krabill, W. B., 2001: A new, high-resolution digital elevation model of Greenland fully validated with airborne laser altimeter data. *Journal of Geophysical Research,* 106(B4): 6733–6745.

Bevington, P. R., 1969: *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, New York. 336 pp.

Centre for Topographic Information, 1999: *National Topographic Data Base—Edition 3 Simplified User's Guide.* NTDB Customer Support Group, Centre for Topographic Information, Natural Resources Canada, Sherbrooke, Quebec. 13 pp. www.ccg.rncan.gc.ca/ext/html/english/products/ntdb/ntdb250k.html.

Centre for Topographic Information, 2000: *Canadian Digital Elevation Data Standards and Specifications.* Customer Support Group, Centre for Topographic Information, Natural Resources Canada, Sherbrooke, Quebec. 13 pp. www.ccg.rncan.gc.ca/ext/html/english/products/cded/cded.html.

Cogley, J. G., Adams, W. P., Ecclestone, M. A., Jung-Rothenhäusler, F., and Ommanney, C. S. L., 1996: Mass balance of White Glacier, Axel Heiberg Island, N.W.T., Canada, 1960–91. *Journal of Glaciology,* 42: 548–563.

Cogley, J. G., and Jung-Rothenhäusler, F., 2002: Digital Elevation Models of Axel Heiberg Island Glaciers. *Trent Technical Note 2002–1,* Department of Geography, Trent University, Peterborough, Ontario, Canada. 44 pp. www.trentu.ca/geography/glaciology.htm.

Davis, C. H., Jiang, H., and Wang, X. Y., 2001: Modeling and estimation of the spatial variation of elevation error in high resolution DEMs from stereo-image processing. *IEEE Transactions on Geoscience and Remote Sensing,* 39(11): 2483–2489.

Defense Mapping Agency, 1992: *Digital Chart of the World.* Defense Mapping Agency, Fairfax, Virginia. Four CD-ROMs. [Now the National Imagery and Mapping Agency.]

Department of Energy, Mines and Resources, 1967: *Strand Fiord District of Franklin Northwest Territories.* National Topographic Series map 59H at 1:250,000 scale. 2nd ed. 1988. Surveys and Mapping Branch, Department of Energy, Mines and Resources, Ottawa.

Ehlschlaeger, C. R., 2002: Representing multiple spatial statistics in generalized elevation uncertainty models: moving beyond the variogram. *International Journal of Geographical Information Science,* 16(3): 259–285.

Hastings, D. A., and 13 others, eds., 1999: *The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0.* National Oceanic and Atmospheric Administration, National Geophysical Data Center, 325 Broadway, Boulder, Colorado 80303, U.S.A. Digital data base on the World Wide Web. www.ngdc.noaa.gov/seg/topo/globe.shtml and CD-ROMs.

Hunter, G. J., and Goodchild, M. F., 1995: Dealing with uncertainty in spatial databases: a simple case study. *Photogrammetric Engineering and Remote Sensing,* 61(5): 529–537.

Hunter, G. J., and Goodchild, M. F., 1997: Modeling the uncertainty of slope and aspect estimates derived from spatial databases. *Geographical Analysis,* 29(1): 35–49.

Hutchinson, M. F., 1989: A new method for gridding elevation and stream line data with automatic removal of pits. *Journal of Hydrology,* 106: 211–232.

Hutchinson, M. F., 1996: A locally adaptive approach to the interpolation of digital elevation models, in Third International Conference/Workshop on Integrating GIS and Environmental Modeling, NCGIA, University of California, Santa Barbara. CD ROM available from www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/main.html.

Liu, H. X., Jezek, K. C., and Li, B. Y., 1999: Development of an Antarctic digital elevation model by integrating cartographic and remotely sensed data: a geographic information system based approach. *Journal of Geophysical Research,* 104(B10): 23199–23213.

McGill University, 1963: *Expedition Area, Axel Heiberg Island, Canadian Arctic Archipelago.* Map at 1:100,000 scale. Accompanies Müller, F., et al., 1963, *Preliminary Report, 1961–1962, Axel Heiberg Island Research Reports*, McGill University, Montreal.

National Research Council, 1962: *Thompson Glacier Region, Axel Heiberg Island, N.W.T., Canada.* Map at 1:50,000 scale. Photogrammetric Research Section, National Research Council of Canada, Ottawa, in conjunction with Axel Heiberg Island Expedition, McGill University, Montreal.

National Research Council, 1965: *White Glacier, Axel Heiberg Island, Canadian Arctic Archipelago.* Map at 1:10,000 scale in two sheets. Photogrammetric Research Section, National Research Council of Canada, Ottawa, in conjunction with Axel Heiberg Island Expedition, McGill University, Montreal.

Robinson, G. J., 1994: The accuracy of digital elevation models derived from digitised contour data. *Photogrammetric Record,* 14(83): 805–814.