Detecting Errors in Survey Data Processing

Raymond J. Hintz, PLS, PhD

University of Maine

Raymond.Hintz@umit.maine.edu

- Topic involves measurements not attribute information.
- No checks in survey data results in no way to identify measurement problems.
- Survey measurement checks involve repetition and redundancy.
- Survey checks are not limited to a given data collection system.
- Survey checks can only be created by intelligent field procedures.
- All concepts presented are in the Electronic Field Book Processor (EFBP)

- Survey measurements are tied to the concept of a point name/identifier.
- Every time you measure or set up on a point it should get the same point name.
- Do not give two different points the same point name.
- Backsight and occupied points are critical in identification by point name.

- Repetition can be of two types
- (1) Repeated measurements to the same point from the same setup.
- (2) Measurements of the same line from different setups. Note this could be in the opposite direction on the line compared to a previous setup.
- Note no coordinate generation is required to analyze repetitions.

- Repetition from the same setup allows

(1) comparison of current measurement to previous averaged information

Reported as “repetition error on multiple pointing to A4 is 12 seconds”

(2) Standard deviation, standard deviation in the mean, and maximum spread (largest difference from the average)

(3) Asterisks mark repetition error and maximum spreads above user defined tolerances, these are also sent to the .sum (summary) report

- Repetition from different setups allows

(1) Comparison of the horizontal distance average to the previous average value for that line

(2) Comparison of the elevation difference average to the previous average value for that line

Note these values will be averaged to create the current average values for that line.

- Concept of survey “connectivity”

Point names create a system of interconnected points.

The connections are defined by abstracted (average) values of horizontal distance, horizontal angle, and elevation difference.

- Are control coordinates needed to generate the relationship of points in a survey network?

If just a relative (assumed) relationship is all that is necessary, software can assume coordinates at the first setup, and assume a direction to the initial backsight.

This does require all points are connected, i.e. all setups must connect to one another.

EFBP would “best fit” the data by least squares and report residuals and misclosures.

Useful if you want internally check your data independent of control.

- Control adds a very important “external” check of our measurements.
- Control places our information in a reproducible datum.
- Control adds redundancy to our survey.
- EFBP will “best fit” the survey measurement data by least squares to the control and report residuals and misclosures.
- The point name is the link of what control points were occupied or measured to.

- Control normally needs a datum definition

(1) State plane zone

(2) Horizontal Datum (27, 83, HARN by year)

(3) Units (survey feet, international feet, or meters)

(4) Vertical Datum (29, 88)

Failure to correctly (1)-(3) produces an incorrect mathematical integration of survey measurements with the grid system control!!

- Least squares error estimation

(1) Horizontal distances grow in error slowly as they increase in magnitude.

(2) Trigonometric derived elevation differences grow in error quickly as the distance increase in magnitude (a large ppm than in (1).

(3) Horizontal angles contain larger errors if sights are shorter (setup error) and also contain a constant error related closely to the least count of the instrument.

- Two acceptable models of error estimation

(1) Repetition error plus user defined add-ons

Add-ons account for items like setup error not included in the repetition model.

(2) User defined error estimates (no use of repetition error). This model has to be used if no repetition error exists.

- Residuals in least squares are the difference between a measurement and the equivalent value derived from final coordinates.
- Snoop number = residual/ error estimate
- A snoop number greater than 3 indicates a suspect measurement at a 95% level of confidence.
- Snoop numbers larger than 3 are asterisked in the 1D and 2D least squares reports and output to the summary (.sum) report.

- Example
- Elevation difference on 200 ft. line has

Error est. of .02 ft. + 100 ppm or

.02 + (100/1,000,000)*200 = 0.04 ft.

A residual of 0.13 ft. yields a snoop number of 0.13/.04 = 3.25 so it will be asterisked and in .sum

The same residual on a 1000 ft. line will not be asterisked as its error est. is

(.02 + 100/1,000,000)*1000 = 0.12 ft

.13/.12 = 1.08 so no asterisk as reasonable snoop #

- Angle residuals

Due to setup error (or how well one can point) an angle with 20 ft. sights may have an error estimate of 60 sec! But 60 sec. In 20 ft. amounts to only .006 ft. This angle can have a residual of 45 seconds without detrimental effect.

While the same setup error will have virtually no effect in an error estimate on 1500 ft. sights so the error estimate will be represented by the angle constant error estimate. This angle better not have a residual of 45 sec!

- Detecting control errors

Assign small error estimates to the control coordinates in the .ctl file (0.03-0.1 ft.)

This will allow control to adjust towards where the survey measurements think it should be.

Residuals on control will appear in the least squares .1d and .2d files. Outliers can be identified.

Correct any found problems and re-run with control fixed again.

- Identifying systematic error

Finding a significant amount of larger residuals of the same sign is an indicator of systematic error.

Sources could be

(1) Incorrect instrument/prism offset

(2) Incorrect or missing zone/datum

(3) Incorrect elevation on .2d geodetic job

(4) Control problems

(5) Bad instrument calibration

- Quick checks on EFBP reports

(1) Go to .sum to see if any repetition error outside of user defined values or any residuals larger than 3 * error estimate

(2) Go to .gen and make sure a reasonable error estimate model was used.

(3) Go to .1d and make sure a suitable amount of benchmarks were measured to.

(4) Go to .2d and make sure a suitable amount of horizontal control were measured to.

(5) In .2d make sure the proper zone and datum were used.

- Live data examples
- Questions and answers