Procedures for Processing Borehole Strainmeter Data V1.1

Tuesday, May 23, 2006

hodgkinson@unavco.org

1. Introduction

PBO produces three levels of strainmeter data products. Level 0 are the raw data. Level 1 are the raw data converted to geophysical units. Level 2 products are processed data. Level 2 products are further divided into 2 sublevels, 2a and 2b. Level 2a products include cleaned gage data, derived time series of areal and shear strain, time series corrections for borehole effects and the earth tides. Level 2b products include, in addition to the strain measurements, a reassessment of the information used to produce areal and shear strain such as instrument calibration and borehole relaxation and curing parameters.

2. Level 0 and Level 1 Products

2.1 Data products

The Gladwin Tensor Strainmeter (GTSM) data logger collects strain data at rates of 20 samples per second (sps), 1-sps and 600-second intervals. Auxiliary data collected at the site such as temperature, pressure and field diagnostics will be sampled at rates of 1 to 1800-second intervals.

Level 0 products are the raw 20-sps, 1 sps, 1/300-sps strain and auxiliary data in the both the data logger's native format (bottle) and in miniSEED (Table 1). Currently only the 1 sps and lower frequency data are converted to miniSEED. Under normal circumstances the level 0 data will be begin to flow to the archives within 2 weeks of instrument installation. In routine operation, data will be downloaded from the GTSM datalogger on an hourly basis. The archives will receive 1-hour long data files every hour. The only processing involved in producing Level 0 products include unpacking and merging the data files into 1-hour long files and conversion to miniSEED format (steps 1 to 4, Table 2).

Table 1: Level 1 and Level 0 data sets

Data	Sample rate (sps)	Format	Generation frequency
strain gage 1	20, 1, 1/600	miniSEED/bottle	1 hour/24 hours
strain gage 2	20, 1, 1/600	miniSEED/bottle	1 hour/24 hours
strain gage 3	20, 1, 1/600	miniSEED/bottle	1 hour/24 hours
strain gage 4	20, 1, 1/600	miniSEED/bottle	1 hour/24 hours
barometric pressure	1	miniSEED/bottle	1 hour/24 hours
pore pressure	1	miniSEED/bottle	1 hour
Auxiliary data e.g., rainfall, temp	>=300	miniSEED/bottle	24 hours

Except for the conversion to natural units, level 1 data are identical to the level 0 data. Since the level 1 data are obtained by multiplying the level 0 data by the appropriate scale factor they are not archived. Rather, the user obtains them from the level 0 data set (step 5, Table 2). The scale factors will be those supplied by the manufacturer and retrievable from the level 0 data set.

Table 2. Strain data processing steps and the outputs at each processing step.

Step	Procedure	Output
Receiving and archiving the raw data (every hour).		
1.	Import tar files produced by the data logger.	
2.	Unpack and merge data logger files.	
3.	Convert raw data to miniSEED format.	
4.	Send miniSEED data and bottle data to archives.	**Level 0 Data Set**
5.	Raw data scaled to natural units s(performed by user).	**Level 1 Data Set**
Rapid data quality check (every 24 hours).		
6.	Linearize the gage data.	
7.	Remove tidal signal and borehole effects.	
8.	Check the detrended and de-tided data for outliers.	
Manual cleaning and verification of data (within 14 days of arriving at BSMAC).		
9.	Reduce the raw 1-sps data to 300-second data.	
10.	Linearize the 300 s gage data.	
11.	Remove tidal signal and borehole effects and inspect the residual gage data using field metadata, field system diagnostics and outlier criteria.	
12.	Apply edits to raw 1-sps data, linearize and reduce to 300-s intervals.	
13.	Generate and remove borehole trends.	
14.	Generate the tidal correction.	
15.	Compute areal and shear strain.	
16.	Convert to XML.	**Level 2a Data Set**
Post-processing and derivation of processing parameters (every 3 months).		
17.	Reprocess all 300 s data including additional field information and editing if required (Steps 11 to 15 repeated).	
18.	Recalculate the borehole relaxation and grout curing parameters using entire series.	
19.	Compute the tidal signal in the clean gage, areal and shear strain data.	
20.	Recalibrate.	
21.	Reassess the outlier criterion.	
22.	Recalculate areal and shear strain and corrections using output from steps 18, 19 20 and 21.	
23.	Convert to XML, store processing parameters	**Level 2b Data Set**

2.2 Rapid data quality check

On arrival at the Borehole Strainmeters Analysis Center (BSMAC) all strain data will under go an initial data quality assessment. The goal of this initial processing is to rapidly identify potential problems, such as spurious or missing data points, in the 1-sps data from each sensor. Strainmeter engineers will be alerted if necessary. The rapid data quality check processing steps are shown in Figure 1 and outlined in Table 2, steps six through eight. They are summarized as follows:

- **Linearize the data** (Step 6). The raw measurements from each gage are output from capacitance bridges. To convert the gage output, in counts, to strain along the axis of the gage the following equation is used,

$$U = \left(\left(\frac{\frac{R_{raw}}{1e+8}}{\left(1 - \frac{R_{raw}}{1e+8}\right)} \right) - \left(\frac{\frac{R_o}{1e+8}}{\left(1 - \frac{R_o}{1e+8}\right)} \right) \right) G / D \qquad (1)$$

  where U is the linearized data point, R_{raw} is the raw data point (1 count = 0.1 nanostrain), R_o is a data point early in the record, G is the reference gap within the strainmeter (200 microns) and D is the instrument diameter (87 mm).

  For PBO GTSM21 strainmeters extension is represented by a positive change in strain.

- **Remove the tidal and borehole relaxation and grout curing trends from linearized data** (Step 7). The tidal signal, borehole relaxation and grout curing trends are removed from the gage data to facilitate checking for spurious data. The tidal signal and borehole effect trends are generated using parameters derived as level 2b products.

- **Check the detrended and de-tided data for outliers and offsets** (Step 8). The data are checked for outliers and offsets by examining the difference between adjacent data points. If two adjacent differences exceed an Outlier Criterion (OC) the common data point is flagged as an outlier. The OC will be recalculated every three month for each instrument and will be first evaluated 6 months after installation (level 2b product). Email will be generated reporting any detected problems such as, offsets, outliers and missing data. These findings will be examined when producing the level 2 data set.

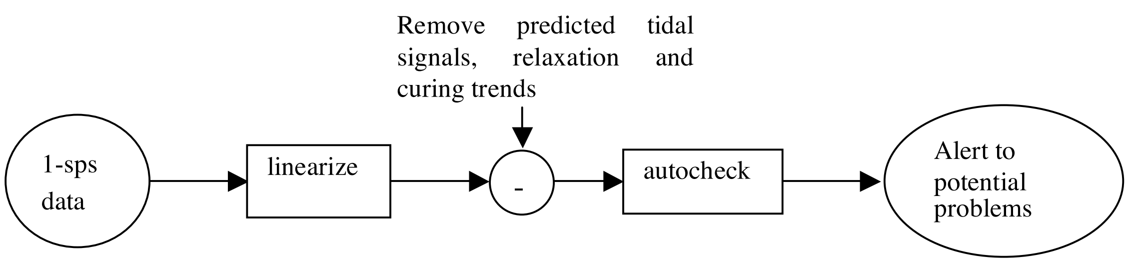

Figure 1. Rapid data quality check

The Gladwin Tensor Strainmeter records systems diagnostics data streams which will also be examined at this stage. The environmental data, temperature, atmospheric pressure and pore pressure are also automatically checked.

3. Level 2: Processed Data and Derived Products

Level 2 data products are summarized in Table 3. The level 2a data set will be no more than 14 days out of date and will be archived in eXtensible Markup Language (XML). Level 2b processing provides an opportunity to post-process the strain data. For example, if it is found that some part of the system has been malfunctioning for some time it is possible to flag the data as poor quality at this stage. At three-month intervals the entire 300-second interval data set will be reprocessed to form the level 2b data set. The clean gage and tensor strain time-series are regenerated by applying the edits found at level 2a and any additional editing found necessary.

Table 3. Level 2 Data products, (POD, PBO Operational Database).

Product	**Output Format**	**Product Level**
Areal and shear strain computed from gage data with and without borehole effects removed	XML	2a and 2b
Clean linearized gage data	XML	2a and 2b
Clean scaled environmental data	XML	2a and 2b
Borehole plus grout curing correction for each gage	XML	2a and 2b
Earth tide correction for areal and shear strain	XML	2a and 2b
Phase and amplitudes of main tides for each gage and tensor strain	XML, POD. Recalculated every 3 months.	2b
Borehole plus grout curing coefficients for each gage	XML, POD. Recalculated every 3 months.	2b
Calibration matrices	XML, POD. Recalculated every 3 months.	2b
Outlier Criterion	XML, POD. Recalculated every 3 months.	2b

3.1 Level 2a Processing

Level 2a processing consists of editing the data and forming the tensor strain and time series corrections for the earth tides and borehole effects. The level 2a processing steps are shown in Figure 2 and outlined in Table 2, steps nine through sixteen. They are summarized as follows:

- **The raw 1-sps gage data are down sampled to 300-second data** (Step 9). The 1-sps data from the borehole strainmeter will be low passed and decimated to produce 300-second samples using a series of filter-and-decimate stages. The filters used at each stage will be lowpass filters designed to provide a flat passband, minimal aliasing at zero frequency, and a minimum-phase finite impulse response (FIR filters). The filtering and decimation will be implemented causally. Causal minimum-phase FIR filters will be used because they minimally distort the onset arrivals from earthquakes (Scherbaum and Bouin 1997). The phase response can be computed from the filter weights given at each stage, which will be provided in the SEED and XML metadata.

- **Linearize the data** (Step 10). Linearize the 300-second gage data as for step 6.

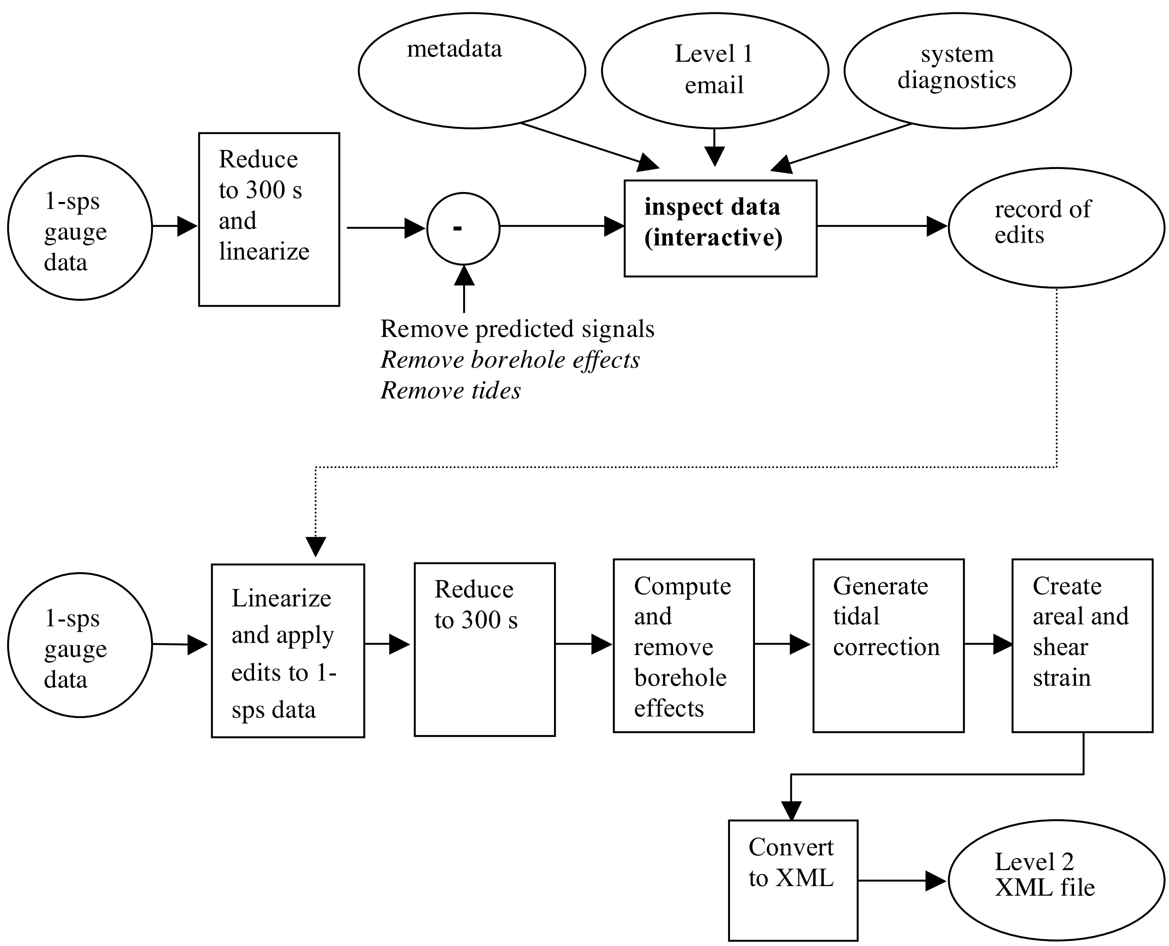

Figure 2. Level 2a processing steps.

- **Data editing** (Step 11). Level 2a data cleaning and editing will not be an automated process. In addition to using an outlier criterion to detect bad data, metadata such as information from the field engineers and the systems diagnostic data steams will be used as input to the cleaning process. Problems detected during the rapid data quality check indicating the presence of outliers, offsets or missing data will be investigated by inspecting the data. Auxiliary data such as instrument voltages and environmental data will also be checked. As for the rapid data quality check, the tidal signals and curing trends are removed from the gage data to facilitate cleaning. Any data that are considered outliers will be recorded and flagged as such in the level 2 XML file. The GTSMs provide a relative measurement of the instrument diameter and, unlike hydraulic strainmeters, do not have reset offsets. For this reason, only offsets that are known to be non-tectonic in origin, such as those caused during site maintenance, will be removed. The magnitude and time of any offset that is removed will be recorded in the level 2 XML file. Coseismic offsets will not be removed but their magnitude will be recorded in the XML file for those who wish to do so.

- **Apply edits to raw data** (Step 12). The edits recorded in step 11 are applied to the 1 sps data. The data are then linearized and reduced to 300-second sample intervals. The tidal and borehole effect signals are not removed from the data.

- **Generate and remove borehole trends** (Step 13). Time series of the borehole relaxation and grout curing effects are generated for the gage data. The borehole effects are subtracted from the edited gage data to form a set of gage residuals.

- **Generate the tidal correction** (Step 14). A time series of tidal corrections are generated for the areal and shear strain data. The parameters used to generate the corrections will be determined at level 2b and reevaluated every 3 months.

- **Compute areal and shear strain** (Step 15). The areal and shear strain measurements are computed from both the residual gage data and gage data that has not had the borehole effects removed. The tidal signal is not removed from this data set. (See Section 4 for equation used to calculate the tensor strain.)

- **Conversion to XML** (Step 16). The processed data are converted to XML. The XML file will contain all the level 2 cleaned strain data and edit information. It will also list the processing parameters used to detect outliers, estimate the tidal and borehole curing effects and matrices used to combine the gage data into areal and shear strain.

3.2 Level 2b processing

In addition to reprocessing the strain time series, level 2b processing shall include the following analyses of the cleaned data:

- **Reprocess entire data 300-second data set** (Step 17). Additional edits are made to the data that were not made at Level 2a.

- **Recomputation of the borehole relaxation and grout curing detrending parameters** (Step 18). The detrending parameters will be recalculated using the entire data record and applied to the level 2b data set. The parameters determined will be used to calculate the gage corrections for the next 3 months of level 2a data.

- **Computation of the observed tides in the gage, areal and shear strain data** (Step 19). The observed amplitude and phases of the earth tide signal will be calculated for the last three months of measurements. These parameters are used to calculate the tidal correction for the strain data.

- **Recalculation of calibration matrices** (Step 20). The instrument will be recalibrated using the last 3 months of observed tidal parameters.

- **Reassess the outlier criterion** (Step 21)**.** The criteria for detecting outliers is recalculated and updated if needed.

- **Recalculate areal and shear strain** (Step 22). Areal and shear strain are recalculated using the output from step 20. The borehole trends and tidal corrections are recalculated using the output from steps 18 and 19.

- **Convert to XML** (Step 23). Data and processing parameters, borehole and grout curing parameters, tidal amplitude and phases, calibration matrices will be converted to XML and archived. It is expected that the processing parameters will change with time and the entire series of parameters will be contained in the XML header.

3.3 Availability of Level 2 products

Level two products require the phase and amplitudes of the M2 and O1 tides to be known. These parameters are required to compute the matrices used to combine the gage measurements into areal and shear strain. The tidal amplitudes are also required to compute the tidal correction. It takes two to three months of strainmeter data to estimate these tidal parameters. The borehole relaxation and curing rates are largest in the months immediately following installation and can complicate tidal analysis. Level 2 products will therefore not be available for three to six months after instrument installation.

4. Calibration of Strainmeters

The strainmeters will be calibrated using the following equation

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{g_1} & 0 & 0 & 0 \\ 0 & \frac{1}{g_2} & 0 & 0 \\ 0 & 0 & \frac{1}{g_3} & 0 \\ 0 & 0 & 0 & \frac{1}{g_4} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\cos 2\theta_1 & \frac{1}{2}\sin 2\theta_1 \\ \frac{1}{2} & \frac{1}{2}\cos 2\theta_2 & \frac{1}{2}\sin 2\theta_3 \\ \frac{1}{2} & \frac{1}{2}\cos 2\theta_3 & \frac{1}{2}\sin 2\theta_3 \\ \frac{1}{2} & \frac{1}{2}\cos 2\theta_4 & \frac{1}{2}\sin 2\theta_4 \end{bmatrix} \begin{bmatrix} c & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & d \end{bmatrix} \begin{bmatrix} t_1 & t_1 & t_1 \\ t_2 & t_2 & t_2 \\ t_3 & t_3 & t_3 \end{bmatrix} \begin{bmatrix} e_a \\ gamma1 \\ gamma2 \end{bmatrix}^{regional} \tag{2}$$

where u_i are the gage readings, g_i the gage weightings, θ_i the gage orientations, c and d the areal and shear strain scale factors respectively, t_i the corrections for the topography and geology. The regional strain is described by e_a, *gamma1* and *gamma2*. Initially the topography and geology matrix will be set to the identity matrix (equation 2 is then equivalent to equation B4 of Hart et al., 1996).

The c and d values and the relative gage weighting g_2/g_1, g_3/g_1 and g_4/g_1 will be solved for by comparing the real and imaginary components of the M2 and O1 tides as observed on each gage with the regional M2 and O1 strain tides predicted using theoretical models. The parameters will be found via a least squares inversion of equation 2. If the least squares solution gives unrealistic areal and shear strain factors (following parameter ranges described in Gladwin and Hart, 1985) then equation 2 will be solved in an iterative fashion using starting values of 1 for the weightings and 1.5 and 3 for c and d respectively. Once the values are found the strain measured by the instrument is calculated using,

$$\begin{bmatrix} c & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & d \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{2} & \frac{1}{2}\cos 2\theta_1 & \frac{1}{2}\sin 2\theta_1 \\ \frac{1}{2} & \frac{1}{2}\cos 2\theta_2 & \frac{1}{2}\sin 2\theta_3 \\ \frac{1}{2} & \frac{1}{2}\cos 2\theta_3 & \frac{1}{2}\sin 2\theta_3 \\ \frac{1}{2} & \frac{1}{2}\cos 2\theta_4 & \frac{1}{2}\sin 2\theta_4 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{g_1} & 0 & 0 & 0 \\ 0 & \frac{1}{g_2} & 0 & 0 \\ 0 & 0 & \frac{1}{g_3} & 0 \\ 0 & 0 & 0 & \frac{1}{g_4} \end{bmatrix}^{-1} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} e_a \\ gamma1 \\ gamma2 \end{bmatrix}^{regional} \qquad (3)$$

The SPOTL tidal programs (Agnew, 1996) will be used to predict the amplitude and phases of the M2 and O1 strain tides. The Gutenberg-Bullen Model A Green functions shall be used along with the CSR 3.0 global ocean model that has been derived from TOPEX/POSEIDON altimeter data (Eanes and Bettadpur, 1995). Global ocean models often do not accurately calculate loading where there are large local tides, e.g., San Francisco Bay Area. Where they exist, local ocean models will be included in the calculations.

If any finite element modeling is performed to quantify perturbations in the strain field caused by local geology and topography this information will be included when calculating the regional strain.

For the first six months of operation it may be difficult to extract the tidal signal if the borehole relaxation and grout curing rates are high. Therefore for the first six months the gage data will be combined into areal and shear strain using gage weightings of one and values of c=1.5 and d = 3.0. A tidal calibration will be provided six months after the instrument has been installed.

5. Tidal analysis

5.1 Estimation of M2 and O1 phase and amplitudes from strain gage data

At level 2b the observed tides will be computed for the last 3-months of data analyzed from each gage. The tidal analysis will be performed using the program BAYTAP-G (Tamura, 1991). The program outputs a list of phases, amplitudes and associated error for a specified set of tidal frequencies. The data will have been detrended and had all offsets removed for input to the tidal analysis program.

5.2 Calculation of the tidal correction time series.

The tidal time series corrections are computed from the tidal amplitudes and phases as measured by the strainmeter. Once the amplitudes and phases of the M2 and O1 tides are known the hartid prediction program within the SPOTL software package will be used to generate the tidal correction tine series. The prediction program uses the amplitude of the observed M2 and O1 tide to scale and sum the deformation associated with 18 long-period, 51 diurnal and 50 semidiurnal constituents. The amplitudes of the additional constituents are found by spline interpolation of the ratio of the potential amplitude to that observed. The time series of the tidal deformation will be calculated, at any sample interval, for the areal and shear strain.

6. Borehole relaxation and grout curing trends

The dominant trend in strainmeter data is that of grout curing and borehole recovery. Corrections for these effects need to be generated for each gage as the trends will vary with orientation. Factors that contribute to the trends are; the compressive stress exerted by the grout expanding as it cures, dissipation of heat into the surrounding rock and the recovery of the borehole. The general equation to describe the combined trends is,

$$y = F + A_1 \exp^{(T_1 t)} + A_2 \exp^{(T_2 t)} + mt \qquad (4)$$

where F is a constant offset, m describes the linear trend that will dominate as time increases, and T_1 and T_2 are time constants. An approximate model for the borehole effects will be found through forward modeling of equation 4. When a good approximate model has been found the best fitting model will be determined using a least squares procedure.

7. References

Agnew, D.C., 1996: "SPOTL: Some Programs for Ocean-Tide Loading," SIO Reference Series 96-8, p.34.

Eanes, R.J. and S.V. Bettadpur, 1995, The CSR 3.0 global ocean tide model, University of Texas Center for Space research, Technical Report CSR-TM-95-06.

Gladwin, M. T. and Hart, R, 1985, Design Parameters for Borehole Strain Instrumentation. Pageoph., 123, 59-88, 1985.

Hart, R. H. G., Gladwin, M. T., Gwyther, R. L., Agnew, D. C., Wyatt, F. K., 1996, Tidal calibration of borehole strain meters; removing the effects of small-scale inhomogeneity, Journal of Geophysical Research, B, Solid Earth and Planets 101, no. 11 (19961110): 25,553-25,571.

F. Scherbaum and M. P. Bouin, 1997, FIR filter effects and nucleation phases, Geophys. J. Int. vol 130, 661-668.

Tamura, Y., T. Sato, M. Ooe, M. Ishiguro, 1991, A procedure for tidal analysis with a Bayesian information criterion, Geophys. J. Int. 104, 507-516.