units of length. These data have millimeter level precision and will be designated 1.1 and 1.2. The set of 1'1, 1'2, 1, and 1.2 data recorded by uninterrupted, continuous tracking of a particular GPS is referred to as a "pass" of data. Typically, the DR collects over 300 passes of data a day. Since the DR continuously collects phase data every second, simple arithmetic shows that over 1 million points per day must be processed. The differential span of both the pseudorange and phase data during a pass can be about 8000 km. The need to retain better than millimeter precision in the data thus requires high precision arithmetic.

STATISTICAL METHODS & SOFTWARE

All of the data are fit (per time series, per pass) with robust regression splines. The purpose of performing the splint fits is to produce residuals as one means to spot faulty data or other specific problems known to occur for the data. The GDPJ provides over 20 user specifiable parameters for controlling how the splint fits are done. Among these are splint parameters for setting (for each datatypethe constant knot spacing, the constant knot repetition to control the degree of derivative continuity, and the constant degree of the polynomial splinets, there are also parameters for controlling the iteratively reweighted least squares algorithm used to solve for the splines minimizing the Huber weighting criterion (1. Huber 1981). Bivariate discriminant analysis is performed with robust covariance matrices calculated with Median Absolute Deviation (MAD) based variances and covariances using the identity

\[ \text{COV}(X,Y) = \frac{\text{VAR}(X+Y) - \text{VAR}(X-Y)}{4}. \]

These algorithms are implemented in Fortran in the GDPL which is run in a batch mode.

Capability for interactive analysis of the automated editor's results is built into top of the S-111GS software (Statistical Sciences, Inc. 1992). This version of the S language (Becker, Chambers, and Wilks 1988) provides a Fortran interface for running subroutines of the data editor with convenient access to all arguments and results, graphical capabilities (X-windows and postscript), and a flexible programming environment used to perform special analyses and prototyping of algorithms for Fortran implementation.

PHASE DATA COMBINATIONS

Figure 2 shows plots of one pass of phase data in the linear combinations that are usually formed by GPS data analysts. The top plot shows the linear combination of phase data often designated LC(0 = 1.1 + 1.54(1.1,1-1,2)). This is the only combination of the phase data which is used by the orbit determination part of the Gill II as the dual frequencies are for ionospheric calibration purposes (to be made more clear soon). Since the differential span of the LC is typically thousands of kilometers while phase problems are often sub-meter in level, plots of LC seldom reveal any problems discernible by the unaided eye. The bottom plot of Figure 2 shows the ionospheric combination 1.1-1,2. This linear combination is dominated by ionospheric effects and is of a scale as to allow one to sometimes see directly the level of the noise in the data. There are some noticeable "discontinuities" in the 1.1-1,2 plot of Figure 2. They are manifestations of what are known as cycle slips - the main problem the GPS data editors must contend with. It has turned out that the Topex-GPS receiver rarely has cycle slips, so the data for Figure 2 is not chosen as a representative sample but rather as one of the few that does have the cycle slip problem.

THE CYCLE SLIP PROBLEM

As illustrated in Figure 2 the cycle slip problem results in discontinuities in the phase data. The sizes of the discontinuities are multiples of a half-integer when the data are in units of cycles in the original 1.1 and 1.2 (not in the LC and 1.1-1,2 combinations).

The data editing software tackles the cycle slip problem by forming the divided differences of the data in time thereby converting the discontinuities into outliers. Then robust regression splines are fit to the rate (LC) and rate (1.1-1,2) to produce residuals by which cycle slips are identified. See Figure 3 for time series plots of the rates of the phase. The bottom plot of Figure 3 overlays the splint fit on the data values.
It is highly desirable to fix the cycle slips if possible. The data editing software fixes cycle slips by using two-dimensional discriminant analysis to resolve how many half-integers were jumped. Figure 4 shows a plot of the residuals from spline fits to the data of Figure 3 in a scatter-plot along with contours of equal probability centered at half-integer locations. Note that the residuals in rate(1.1) and rate(1.1-1.2) have been transformed to units of cycles in 1.1 and 1.2. As usual in discriminant analysis the contours are based upon Gaussian assumptions but using robust covariance estimation as indicated earlier. The cycle slip identification algorithm proceeds by first determining the cycle slip center to which a point is closest. Then it checks that a point is within some user specified acceptance region to bound the probability of misclassification. If a point cannot be classified with very low probability of error (as specified by the user), then the cycle slip is simply marked as the beginning of a new pass. In this case all the cycle slips were properly identified and the fixed 1.1-1.2 combination is shown in Figure 5.

**STATISTICS FROM THE EDITOR**

The data editor calculates and saves over 30 statistics for every pass of data successfully processed. These statistics provide a means of monitoring the editor's performance and of monitoring the receiver's activities; unusual statistics could signal that processing is invalid or that some unanticipated problem slipped through the editor. An examination of these statistics for each of the 300 plus passes per day is the main way the editor's performance is validated before the edited data are passed onto additional processing. Of course, graphical representation of these statistics greatly aids in understanding. For each day of data processing the editor automatically produces over 10 postscript files of graphics (most with multiple plots) of these statistics. Figure 1 is a simplified version of one of these daily graphical summaries.

As another example, Figure 6 shows a scatterplot of the MADs of all the passes of residuals from the spline fits to 1.1 and 1.1-1.2 phase combinations for 31 May 1993. A point with extreme statistics appears in this plot and signals that the pass should be inspected for some type of failure in the editing. The 1.1-1.2 combination for that pass is plotted in Figure 7. For a GPS data analyst this plot reveals some obvious problems with the data that had not been anticipated when designing the editor. The appropriate action for this pass is to delete all of the data and determine whether it is thereceiver or the transmitter which is at fault. As it turned out, GPS 31 was experiencing hardware problems on that day.