Q: When Am I Getting My Data?

A: (by Bruce Chapman, JPL)

One of the most common questions from AIRSAR Investigators is how long processing will take. The answer is quite complicated, as it depends on many factors, such as data quality, data mode, processing backlog, funding issues, priorities, and sometimes even human error. During this talk I will discuss these factors in general ways, and describe the processing chain in detail, so that you may better understand the complexity of estimating how long it will take to process a particular job.

The short answer is: “It is going to take longer than you would probably prefer!”

Processing Throughput:

Processing throughput is a key determinant to answering this question. However, AIRSAR, while operational, is constantly evolving, and the processing throughput is quite variable. The processing throughput is dependent on the mix of data modes that are requested for processing, the computing resources that are available, and the manpower that is in-place to perform the manual interventions that are required to deliver a final product to a customer.

We measure our processing throughput by the number of “frames” that are produced, where a frame is 10km (along the flight track) of imagery. AIRSAR can process between 10 and 60km of imagery per single product. The longer the requested product, the more processing power that is required, the more disk space that is required, and the more likely it is that there will be a processing “problem”. On the other hand, the tie-pointing of a TOPSAR 10km product is almost as difficult as tie-pointing a 60km TOPSAR product (it may even be more difficult if it is more difficult to identify a suitable number of tie-points). If we are processing L-band in addition to C-band TOPSAR data, that likewise increases the processing time, disk space requirements, and manual interventions (such as tie-pointing) that are required.

We are currently ramping up production in several ways:

1) We now have three SGI computers (soon to be four) dedicated to processing. This is up from the two available just last year. In addition, we have a computer that is dedicated to survey processing and ATI processing (not discussed here).
2) We now employ a team of Caltech students to tie-point the data. Currently, we have 6 part-time students performing this function.
3) We are automating as many functions as is feasible. The need for human intervention can substantially slow down the throughput. However, at the same time, we must maintain our data quality standards, so these automated functions must be thoroughly validated when implemented.

For 2002, our objective is to double the processing throughput of 2001.

Calibration:

After a flight campaign, the highest priority is to calibrate the data from that campaign. This requires processing data from every mode over a corner reflector array, and determining the best calibration parameters for each mode. Often there have been some enhancements to the instrument; this sometimes may impact the processing, and the processing procedure may need to be modified. During a flight campaign, all the data must be processed to low resolution (surveys), and some updates are required to the database. There is usually a delay immediately following a flight campaign to perform these tasks.

The Processing Queue:

The processing Queue roughly manages when data gets processed. When you submit a job for processing through the AIRSAR website, the job is in an inactive state, until someone from the AIRSAR group activates the job for processing. When the job is activated, the job is given an initial processing priority. This initial processing priority is not well defined. It will depend on how many other jobs are in the Queue, how many jobs have been submitted by the PI in total, what the mode is, and an intangible estimate of how timely the processing must be. Currently, the processing backlog is extensive (over 1 year); therefore, we try to interleave processing requests between the different users, such that the highest priority jobs (as identified by the PI) are processed in a more timely fashion for each PI.

Sometimes, we are directed by NASA to give a higher priority to some processing requests.

The processing priority of the Queue currently does not differentiate between POLSAR processing, TOPSAR processing, and Duplications. We must balance our activities between these three functions within the capabilities of our limited resources. Clearly, it is not fair to put duplications at the lowest priority, since they require the least (though not insignificant) resources. Every week, we try to produce a fair mix between these three data products.

Funding issues also come into play within the Queue. International PI's pay for the AIRSAR processing, while processing for NASA PI's is currently subsidized by NASA's AIRSAR data library budget. These international requests must be balanced against NASA processing requests in a fair manner.

POLSAR Processing:

POLSAR processing is relatively straightforward and does not involve much human intervention. When the job comes up next in the Queue, the operator transfers the data to an available disk and performs some sanity checks (i.e. framecount is correct, no missing lines, etc), and checks the flight log for any problems noted. The standard AIRSAR processor then processes the data. It is the same processor that processes the TOPSAR data – except that no interferometry processing is performed. This processor calibrates the polarimetric data, and produces multilook products in the slant range projection (since we don't know the topography, we don't project the data to the ground range).

Problems that sometimes arise include: cross talk (can usually be corrected), channel miss-registration (can be corrected), phase errors due to instrument problems, P-band interference (can be corrected), ambiguities, and motion correction errors.

But, generally, POLSAR processing proceeds smoothly. Factors that may cause delays in processing include: transfer errors, and delays in staging the data to allow an equitable distribution of POLSAR/TOPSAR processing.

TOPSAR processing:

TOPSAR processing is generally more complex than POLSAR processing, due to the variety of data acquisition modes, the process of tie-pointing the imagery to maps, and the larger number of problems that may arise. TOPSAR processing also requires more processing resources, and more extensive QA procedures. In addition, there is more demand for TOPSAR processing. At the end of a flight campaign, the backlog of the Queue can be quite extensive.

There are several factors that may impact TOPSAR processing in particular:

1) Data Quality

   If the data quality is poor due to low signal to noise ratio (a problem for radar dark areas), steep topography, or turbulence experienced by the aircraft, this affects our ability to unwrap the interferometric phase, or may introduce artifacts into the imagery. Most TOPSAR imagery has "multi-path" banding, due to reflections off surfaces of the NASA DC-8. The magnitude of this affect tends to vary from scene to scene. If it is particularly bad, it can make accurate tie-pointing difficult.

2) Data Mode

   There are two principle modes: XTI1 (C-band TOPSAR) and XTI2 (C-band and L-band TOPSAR). XTI2 tends to be more difficult to process, as twice as many tie-points must be identified, and the L-band DEM data is noisier.

The AIRSAR TOPSAR processing chain from start to finish:

1) Data Request
   First, the user must go to the AIRSAR web-page, and select from the survey imagery what data is requested for processing. This is where users also select duplications of previously processed data, and POLSAR processing.

2) Data Processing Activation
   In order for the data to be processed, the order has to be activated. This is done via an internal software system. We will not generally activate TOPSAR jobs for processing unless we have access to maps to perform the processing.
   When the data is activated for processing, it is assigned a processing priority (a unique number).

3) Waiting for Processing

The data will be processed approximately in the order of this processing priority. Exceptions are that sometimes we have disk space for a small job that may get processed more quickly than its processing priority indicates. This processing priority is constantly updated as data gets processed, and as new processing requests are activated.

4) Check Flight log for problems
The operator checks the flight log for the data take, to make sure that there were not unusual conditions during the flight (bad weather, instrument failures, etc).

5) Transferring the data
The operator transfers the data from the high density digital tape media to disk. Depending on the length of the data take to be processed, the operator must examine the available processing resources to determine which computer and which computer disk to transfer the data to. We currently have 3 computers and about 20 data staging areas.

6) Initial quality assurance
The operator performs an initial assessment of data quality. We verify that the expected number of framecounts were transferred, that there were no bad lines, no byte shifts in the data, and that the Digital Chirp Generator was stable. Plots of the motion data, phase information, and the P-band spectrum are generated.

7) Standard processing
Currently, three SGI Origin computers perform the processing. The processor is the AIRSAR integrated processor (AIP). Currently, the version number is 6.33, where the version number if incremented whenever there is a change in the processing code. However, sometimes changes in the operations occur (for instance, verifying the geolocation) that don't result in changes in the processing version number.
We are currently in the process of adding processing capability to increase the processing throughput.

8) Miscellaneous checks
After the data is initially processed, we perform the following standard checks on image quality:
    a. Sub pixel misregistration between TOPSAR images from two antennas. If there is a large misregistration, the quality of the DEM will be degraded.
    b. Band to band misregistration. This is to insure that the C, L, and P-band data are co registered to each other.
    c. Geolocation errors. If possible, we verify that points on the map agree with the indicated latitude/longitude of the same points on the imagery.
    d. Phase unwrapping. In regions of steep topography or low signal to noise ratio, we sometimes have phase unwrapping errors (usually gaps where the software was not able to figure out how to unwrap the phase of the interferogram). We have several methods to reduce the impact of this

problem, such as parameter tweaking (i.e. Changing the correlation threshold), and splitting the job into two separate jobs.

9) Tie-pointing

An important stage of the processing is correcting for tilts and offsets to the DEM. These tilts and offsets are due to unknown uncompensated errors – theoretically, it should not be necessary to perform this step, but currently the corrections are quite large, and are required in order for the data to meet our data quality standards.

For this step, we employ topographic maps to correct these errors. We generally select between 12 and 16 tie-pints per image. These tie points are selected across the range direction of the scene to make sure we compensate for any cross-track tilts. We also try to sample in the along track direction, but generally the along track direction is pretty stable.

We now employ 3 to 6 Caltech part-time work-study students to perform this task. It generally takes from 4 to 8 hours for one student to tie-point one scene. They identify elevations on the topographic map that can be identified on the imagery (typically: mountain peaks, road intersections, river bends). This information is then entered into custom software that figures out how to tilt and offset the data to make the DEM agree best with the map values. This can take several iterations before the residual errors are acceptable.

It becomes more difficult if the maps are not good quality, if there is a scarcity of good tie points on the map where the imagery is, or if the topography is quite complex making identification time consuming. If the imagery contains ocean data, we will sometimes use those locations as tie-points. However, ocean imagery tends to be noisy, so care must be taken in those cases.

10) P-band interference

If there is interference in the imagery (generally, this only happens in the P-band imagery), then we make an attempt to Filter the P-band interference. Usually, the P-band interference is isolated in just a portion of the P-band spectrum. And, usually, the P-band interference occurs in geographically distinct areas of the imagery. The remainder of the spectrum is used to make the imagery. However, due to calibration issues, we always give users both the filtered and non-filtered imagery, so that the user can use whichever product makes most sense to them.

11) Error resolution

Sometimes, due to instrument failures, data anomalies, or processing errors, the data does not meet our standards and requirements. In that case, we discuss the problem and investigate the source of the error. We currently have over 30 problems that we track in our database. Sometimes, we are unable to fix the problem, and we must abandon the job. Other times, lengthy investigations are required to determine the nature of the problem. This can impact future data acquisitions as well, so great effort is made to understand the nature of these problems. The error resolution procedures and investigations can affect our processing throughput, as resources are sometimes required for lengthy periods before the job is either abandoned or corrected.

12) Radiometric corrections

After the DEM is verified as being accurate, the other data is projected to the "ground range" projection, and corrected for topographically induced radiometric calibration errors, based upon the C-band DEM. When both C-band and L-band DEM's are created (the XTI2 mode), the C-band data is used for correcting the P-band imagery.

In years past, when the P-band data was acquired with a bandwidth of 20Mhz (the maximum allowed in the USA), and the C-band and L-band data were acquired with a bandwidth of 40 Mhz (for greater spatial resolution), the P-band data could not be co registered to the C-band DEM. However, beginning this year, the 20Mhz P-band data will be co registered and calibrated to the 40Mhz C-band DEM.

13) Final QA

The operator now performs the final Quality Assurance assessment. We verify that the caltone signal (which is used to calibrate the phase) is stable, that the polarimetric phase makes sense, and that the backscattered power values seem reasonable. For the later two tests, the phase and sigma0 values are measured within selected uniform homogeneous areas and compared with general expectations of what those values typically tend to be.

14) Make color prints, print log file

Several color prints are generated combining the total power of the bands, the topography, and depending on the mode, combinations of the different polarizations. These color prints are archived in our library, along with log sheet

15) Process review

Once per week, we have a processing review of all processed data. The AIRSAR group supervisor (David Imel), the processing engineer (Anhua Chu), the science coordinator (Bruce Chapman), the software engineer (Wayne Tung), and the processing operator (Lori Carrico), and sometimes the AIRSAR system engineer (Yunling Lou) and the computer system administrator (Joe Kwan) attend this meeting and review the processing for each scene, and evaluate one last time the data quality, assess problems and their resolutions, and assign actions for improving the processing procedure.

16) Make cdroms and archive tapes

We then make cdroms for the customer, a cdrom for our archive, and an archive backup. Until recently our archive tape was an 8mm exabyte tar tape. But recently, we have begun using DVD for this backup. We may consider porting all our backup data to DVD.

17) Update database, and Load gif files to website

AIRSAR has an oracle data base that, in addition to managing the Queue, also records problems with individual datatakes, when the processing is completed, who it was sent to, and manages some of the functions for the interactive map interface for searching for data. A little known fact: at this point, the color photo gif images are put on the website. You can immediately see this image before it has been sent by postal mail.

18) Fulfillment of order

Now that the data is completely processed, we write up a letter for the user describing the processing, especially noting any problems that may have occurred, and any artifacts that might be present in the imagery. Then, we stuff an envelop with the letter together with a cdrom of the data. For international deliveries, we now ship by FedEx, as we have had some problems using regular mail.

Duplications

AIRSAR has an extensive collection of imagery that has already been processed, and it is not unusual for investigators to request this processed imagery. However, duplications do have an impact on the processing operations, and there can be delays even in providing this simple service to customers.

This past year, we have had some problems with archive media, which has meant that some duplications have actually required the data to be re-processed. In general, a duplication can be made available to users within a couple weeks of the request, assuming that the number of duplications do not impact our primary responsibility to process data for PI's.

Q: How do I search for AIRSAR data?

A: AIRSAR web-page description

This talk will graphically show the steps for navigating the AIRSAR web-page in order to search for AIRSAR data, and how to order it.

Q: What is the format for my AIRSAR data?

A: (By Bruce Chapman, JPL)

This talk may be found on the AIRSAR web-page following the workshop (airsar.jpl.nasa.gov) under frequently asked questions. It will describe the data formats of the files found on the cdroms that users receive when they order AIRSAR data.