

## DATA SYNTHESIS CHALLENGES FOR THE FUTURE

Bill P. Clark  
Computer Sciences Corporation  
System Sciences Division  
15245 Shady Grove Road  
Rockville, Maryland 20850

During the most recent ISPRS Congress Meeting (Kyoto 1988) there were less than ten papers that dealt routinely with the mix, match, and merger of remote sensing data using expert systems. Since then multiple organizations have evolved that use these concepts as drivers for their analysis of image data and its merger with other non image remote sensing data. The launch of multiple satellites in the future and the management of the data will be discussed here. In addition, there will be illustrations of the information gathered to date in this area and a projection will be made relative to the type of system the scientific user of the data will require in the next decade. Both sensors and ground systems will be discussed with emphasis on the merging of the two technologies and the diverging of different ideas as new and faster ground systems are designed.

KEY WORDS: Remote Sensing, Landsat, Space Segment, Ground Segment

### INTRODUCTION

This paper will discuss the problems that can be expected to occur in the future for the scientist that desires to use multimission and multisensor data in unique applications. Most of these problems can be solved quickly. However, others require a great deal more thought and must be placed in consideration prior to future missions. Many of these problems present opportunities for the remote sensing community that were previously unavailable. They are of sufficient magnitude that they must be documented from the perspective of the past use of multisensor data and a projection of how these past ideas might be extended into the future. The goal of this paper is to provide this type of orientation for the next generation of users of remote sensing data. Missions, data archives, multiple international satellites, new media for the users, special concerns about mathematical transformations for the future, data volume and rate of capture, data interchange and formatting, and the need for open discussion - and even debate - will be presented.

### MISSIONS

This paragraph will not be all-inclusive. Instead, six references will be given (NASA 87, NASA 91, Greenstone 91, Clark 88, Kostiuik and Clark 84, Clark 85). These articles discuss the report to NASA headquarters on Remote Sensing Applications, a report prepared for NASA Headquarters on the Earth Observation System, the proposed instrument suite planned by NASA for the Earth Observation System (EOS), NASA's projection of critical technologies for the future of remote sensing in the year 2000 A.D., the use of expert systems for the reduction of data for the nineties, and some work already published regarding data synthesis that can be used as a benchmark for future approaches to multisensor activities (Clark and Johnson 85, Clark, Sadowski and Johnson 87).

Multiple missions for the future have been planned. Many have been delayed due to technology and/or budgetary constraints. However, there are multiple missions that offer a logical data set that can be implemented early by investigators. These include the following:

- Landsats 4 and 5
- Landsat 6
- NOAA AVHRR (NOAA 9,10,11, etc.)
- ERS
- JERS
- SPOT
- Indiasat
- Radarsat
- Polar Platforms/polar platform free flyers

All of these with the exception of Landsat 6, Radarsat, and the Polar platforms have been launched and are currently providing data to the worldwide community. Radarsat is a Canadian Centre for Remote Sensing project and the polar platforms have three unique managers for the future. The best inputs to date (1992) indicate that both Europe and Japan will launch multi-instrument polar platforms during the next decade. The USA has had multiple task forces assigned to determine whether their methodology should be the same. Since the original study there have been conceptualized many free flyers to replace the platform. These instrument complexes are less expensive than a large polar platform to launch and provide the same data when all data sets are unioned together. It has been projected that in the EOS era there will be downlinked from the satellites (only the USA complex) an average of one Terrabyte every day. To put this in perspective we use the current Landsat (4/5) Thematic Mapper as an example. The TM data downlink rate is 84.9 Megabits per second. A full scene covers an area on the ground that is 185 kilometers wide. Each scene contains a nominal 300 megabytes of combined image and ancillary data. This means that the EOS instrument suite when fully operational will downlink the equivalent of 3400 Thematic Mapper scenes every twenty four hours!

The original system designed for TM data was built under a NASA specification requiring processing of 100 scenes radiometrically calibrated, 50 scenes geometrically corrected, 50 scenes taken to film, and 10 computer compatible tapes every day. The system was new and contained special purpose computer interfaces including a reworked parallel processing computer system needed for the data throughput rates. The technology at the time of build of this system was not adequate to do the job without using three shifts a day in operations and working for seven days a week. Eventually new methodologies were developed that enhanced the

system output, allowed these goals to be met normally in the "burst mode", i.e., allowed goals for any one or two days to be exceeded while maintaining a long term average that met the original plans. Hardware was slightly ahead of the state-of-the-art at the time of turnover to operations. It is now recognized that similar or even a new and different approach will be required in the ground segment for data capture and processing to work with the EOS data. Newer computer technologies pioneered by NASA and their contractors must be considered for the future.

Recognition of potential future problems with this volume of data has led to the proposed establishment of multiple archives and different production facilities. The archives will be separated geographically as will some of the production activities. This is necessary for purposes of cost and the utilization of existing facilities wherever possible. However, the user of multi-instrument data should be aware of the fact that there is currently no standardized method by which he can assure the capture of data from all spectral bands of interest, capture of multiple spatial resolution data sets, and capture of atmospheric data that are downlinked from various satellites at the same instant of time. Ground truth experiments performed at the time of overpass of the desired satellites are expensive. Users in the field want to know that they can acquire simultaneous data from multiple platforms. If this is not possible the thrust of future programs will be lost, the users will spend their monies unnecessarily and the remote sensing community will be left with many data sets yet be unable to perform a satisfactory correlation from one to another.

#### KEY TECHNOLOGIES FOR THE SPACE SEGMENT

In earlier studies (Kostiuk and Clark 84) it was determined that there were multiple technologies that would drive the development of new detectors in each spectral domain. Since then it has been necessary to modify some of the projections to the slow evolution of the state of the art of detector technology. Some areas have accelerated. Others have been either put in a hold posture or have not lived up to their expectations. In particular array technology for the long wave infrared has been particularly difficult to reproduce in a consistent fashion. However, some more mature technologies have resulted in visible and near infrared cameras with multiple bands that are currently being tested at the operational level. This technology is so well worked out that the aircraft data provides a very good spatial resolution (on the order of one meter dependent on altitude) This allows the capture of point target data which can be merged with MSS and/or TM to "sharpen" the imagery.

Many of the systems planned for the future require that the data be special processed, i.e., uniquely tailored to meet specific user requirements. This will become of increasing importance in the future as new instruments with pointable detector arrays are launched. Detector arrays have both good and bad characteristics. Good characteristics include the geometric fidelity of the image data due to the prelaunch measured positions of each detector in the array. This makes the geometric correction, currently done in software using millions of lines of code, much simpler. On the other hand the radiometric correction for the data is more difficult. To illustrate consider a 1000 x 1000 pixel array. In order to do a radiometric correction for this many detectors it is necessary that someone measure prelaunch the response of each

detector. In the ground data processing area it is necessary that there be established a file that carries one million gains and one million bias values. If this is done for a "small" system like Landsat with only seven bands then there must be placed in storage seven sets of two million values. This 14 megabytes of information will require periodic update as the detector response decays with time on orbit. This means there should be set aside multiple files - some of them open - that can be dated and can carry future calibration smoothing data. Obviously there is a tradeoff here. Furthermore the use of a million pixels in the array was only an example. If we were to use Landsat TM and propagate this type of analysis it would be obvious that there would be the following calculation:

Calculation: seven bands x 36 million pixels per band = 254 million gain and bias pairs required. This too is only a nominal number and presented here to illustrate the upcoming tradeoffs that must be made between the space and ground segment in the next decade.

Pointable systems and stereo systems have also been proposed in the literature. These systems can provide repeat coverage on a daily basis for some satellites that are polar orbiters. However, they provide a significant challenge both for the instrument maker and for the ground processing segment. The problems and opportunities in this area are presented in the next paragraph.

#### KEY TECHNOLOGIES FOR THE GROUND SEGMENT

Obviously the capture and dissemination of one terrabyte of data every day must make the builders of the ground segment take note of new technology. One of the key factors will be data base management. In order to do the "mission to planet earth" job correctly from the perspective of the multi instrument user there must be a methodology whereby there can be captured information from all centers about processed data over point targets at the same instant of time. For example, if the USA finalizes its approach and commissions six centers for either capture, processing, archiving, or all of these it is expected that each center will be unique in the type of data captured and/or processed. The user of the data needs a network, a catalog, and a method including perhaps an expert system available to him which will provide multisensor acquisition, processing, and archive data information.

This type of thought process has occurred at GSFC and at other NASA sites. However, it is presented here to illustrate the fact that it is significantly different than the current Landsat approach to archiving the data. Other countries with their responsibilities for the launch, operation, and monitoring of the earth using their Landsat equivalents will have the same problems. Since many countries do not deal with high volume production facilities yet produce some very high quality products they may have to rethink and perhaps separate their Landsat data from other instrument data archives.

During the past ten years we have seen some explosive growth in the development of hardware. Many computers that we use today were nothing more than someone's dream and projection ten to twenty years ago. Most scientists are no longer computer illiterate and use their systems every day. Many of them did not have the luxury the computer affords for complex calculations in the past. Today the USA

computer industry has broken into three unique industrial sectors. The microcomputer industry is booming, the mid sized computer industry is stable, and the large scale computers with parallel processing are available but not to everyone. There is some cross talk from one genre of computer hardware type to another and this overlap can be expected to be expanded as memory becomes cheaper and the logical progression of upgrading systems continues. This should make small, user friendly, extremely powerful systems available to the working scientist in the future. Furthermore, the new media for data storage i.e. diskettes, 8 mm tapes, 4 mm DAT tapes, CDROM, WORM, and Digital Paper (Clark 1992) will make the archiving process potentially manageable. Future ground segment builders should begin now with their plans to use special media for data distribution. They should also attempt to have at least one open system, i.e., one system in their laboratories where they can add to the computer hardware and software without a procurement of a totally new system. This recommendation allows the near term use of nine track tapes, the future use of cassette tapes, use of optical storage media, and the logical evolution from one to the other should the cost of the full system be too large. This approach also provides for newer technology as it evolves. Most international corporations have internal research and development programs that treat these elements of future computer hardware systems. It is imperative that the user of large volumes of data keep this in mind just to stay abreast of how to think of merging data sets.

In addition to hardware there is a need for the community of data users to start now in their development of specialized algorithms for the future. Most have not been appraised of the Landsat, NOAA, etc., plans for the future yet have algorithms in place for data classification. Users should beware!! No Landsat mission has produced data that is exactly like its predecessors. The same holds for AVHRR and other polar orbiters. There is no guarantee that these data will match future missions. In fact, there is a guarantee that none of them will match currently available data sets. This means the users must initiate studies on their systems relative to how they might upgrade their transformations for future missions. There is a "worst case" scenario where users that are not familiar with the data (old and new) will try to merge these data without cross mapping and therefore cross matching the data sets. If this is done the conclusions drawn from newer data sets will not match those from older. The result will be confusion.

In the context of transformations it should be noted that many vendors sell relatively cheap image processing systems. Thanks to the availability of these systems to the outside world many of us that have been users and continue in this area have found that there are multiple methods of displaying the data and multiple schema that are not generally used by others. In particular the users of future data should recognize that there are multiple nonlinearities in the systems that exist today. These should be considered and worked with using some care because the degree of nonlinearity in radiometry and geometry transformations for systems for the future is not as yet determined. However, older archives of data allow the nonlinear theorist to perform "proof of concept" experiments today. Using Landsat as an example the data for TM were subjected to ground truth testing by multiple principle investigations. In these study efforts the systems were assumed to be linear and there was never any contrast stretch, image enhancement

application, or information maximization from targets in the data. Such transformations were not attempted until small, user friendly image manipulation systems became available. Current nonlinear methods do have a physical and mathematical basis. These have been developed over the past decade. They have proven to be of significant value to unique applications.

Atmospheric corrections of the data will be extremely difficult regardless of the model used. To perform a correct atmospheric correction one must solve the complete set of coupled nonlinear partial differential equations that describe transmission through the atmosphere. Furthermore, this is a VERY LONG process that is extremely computer intensive when executed properly even on large systems today. This has led to multiple approximations that are in use today by some scientists. If we assume that there is a good approximation and it requires only one second per pixel to implement without a parallel processor then we quickly can find that we need three hundred million seconds to modify the data in the strictest scientific sense. (approximately 33333 hours) This calculation is not perfectly correct and it was given only to illustrate the problem. Even if co-terminus LIDAR data are present with a sounder, the two data types are so disparate that the merger of these data with image data will provide one of the most significant image analysis problems of the future.

Past ground truth experiments must be repeated for future systems. The author is unaware of any current plans to update these ground truth data. They cannot be performed with any overflight data until after launch of the instrument of concern. More than fifty instruments are described in the 1992 edition of the EOS instrument complex documents. This data, published as an upgrade to the 1991 data has only recently become available.

The reference (Greenstone 91) cited has minimal information regarding applications and no inputs regarding data fusion. This will become one of the most interesting areas given that there will occur a natural evolution of applications for each instrument both pre and post launch.

At the current time there exists some skepticism relative to the international community's capability to handle the volume of data that will be downlinked in the first decade of the next century. The following paragraph is quoted from the journal "SCIENCE NEWS", March 14, 1992, Volume 141, No.11, page 175. These data are brief and taken as overview information from the United States General Accounting Office Report published in February 1992.

"During its 15 year lifetime, EOS could collect roughly eleven gigabytes of data - more than 1,000 times the information stored by the Library of Congress. NASA is currently devising an EOS Data and Information System or EOSDIS. This system is devised to service an estimated 10,000 researchers." The article goes on to say that NASA is not currently funded to perform the work required to prototype the required systems for the future.

Given this type of input in the scientific press it is obvious that there must be some planning for the future that has not yet been performed. Individual users and data centers must start now if they expect to maximize their use of this vitally important data in the future. This gives the users a chance to think through their applications and provide inputs to NASA and other agencies that will be a part of this program.

There have been many very successful satellites in the past. The technical use of the data has been advertised by every country in the world that has launched a new instrument. Should history repeat itself we will find in the future that there will be a significant benefit and many more useful applications discovered than originally conceived by the instrument builders. The concept of data fusion is seldom mentioned, instead broad reference to "sharing" disparate data types is frequently mentioned. Benefits are left to the scientific community at large to determine through their research.

In order to provide the international community with data that crosses multiple spectral domains with multiple instruments it will be necessary that there be designed some data format compatibilities not present in the past. The community is moving in that direction today. However, for larger data sets it could be necessary to meet and discuss international implications of new formats using ISPRS, the UN, or other international forums.

Such standardization was initiated early for Landsat. The Landsat Technical Working Group started its work in this area in the mid 1970's. Since multiple countries have announced that they plan to have unique data sets in the future, the National Oceanic and Atmospheric Administration (NOAA) has established a Committee on Earth Observational Satellites or CEOS to standardize formats worldwide. However, this has not yet been done for future systems. This will be one of the more pressing needs for the international community as they pursue the open skies policy and exchange data in the future.

In addition the user must have some assurance in the future that he or she can purchase the data with no restrictions over specific parts of the world. This "open skies" policy has been followed by the USA for twenty years. This is an international political agreement that frequently must be renegotiated. The negotiations require a lot of time and effort. This is vital to our study of planet earth. If this is not done the investigators cannot be fully appraised of the status of their information in terms of acquisition, processing, and applications. These activities should begin now!!!

#### APPLICATIONS

Most applications for multispectral data at varying resolutions have evolved through the good work of scientists around the world. Using the differences in future mission instruments and some mathematical modeling of sensor response it will be possible to derive new algorithms for the future. If new algorithms are not developed some very incorrect conclusions could be made using current methodologies with new data. For example the radiance ranges and the modes of operation for Landsat 6 will not match what has been accepted for Landsat 4/5. This mismatch is in concurrence with the mismatch between the first series of Landsats (1,2, and 3). This means that there is a pressing need to cross compare data at the earliest possible time and even simulate what might happen if the sensor calibration does not provide "perfect" data for the future. The international community of scientists should begin this most important of their studies as soon as possible.

A multitude of challenges face those of us in the world of remote sensing, GIS, and other applications areas. As a community we must become more aware of the evolving nature of the technologies for sensor systems, their detectors, and the ground processing of multiple data sets. For those of us that either build, buy, or recommend imaging systems and remote sensing work stations there is a simple word of caution. Beware! Keep your systems open for new applications, make them as generic as possible, and follow the evolution of multiple computer technologies prior to utilizing future systems for data production. This recommendation is not limited to specialized systems. Do not be surprised to find that there has been no thought given by the community at large relative to how to mix, match, and merge future data production, evaluation, and analysis systems. I recommend that we start this work NOW!

#### REFERENCES

- Clark, B.P. 1992, "A Survey of Digital and Optical Media for Storage of Landsat Data", ISPRS, Commission II invited paper, to be published in the International Archives of Photogrammetry and Remote Sensing, Proceedings of the XVII Congress Meeting.
- Clark, B. P., 1988, "Expert Systems for Image Processing", Invited Paper, ISPRS KYOTO, International Archives of Photogrammetry and Remote Sensing, Volume 27, Part B2, Commission II, pages 61-71.
- Clark, B. P., Sadowski, F. J., and Johnson, A. J., 1987, ASPRS Technical Papers, ASPRS-ACSM Fall Convention, pp. 331, ISBN 0-937294-95-0.
- Clark, B. P., and Johnson, A. J., 1985, ACSM-ASPRS Technical Papers, pp. 821-832.
- Greenstone, R., 1991, EOS Instrument Summaries, GSFC, Greenbelt, Md. 20770
- Kostiuk, T., and Clark, B. P., 1984, "Spaceborne Sensors (1983-2000); A Forecast of Technology", NASA Technical Memorandum 86083, GSFC, Greenbelt, Md. 20771
- NASA, 1987, "Linking Remote Sensing Technology and Global Needs: A Strategic Vision", Paper prepared by the NASA Headquarters Working Group on Applications.
- NASA, 1991, Report of the Earth Observing System (EOS) Engineering Review Committee.