Purdue University
Purdue e-Pubs

LARS Technical Reports | Laboratory for Applications of Remote Sensing

1-1-1973

Analysis Research for Earth Resource Information Systems: Where Do We Stand?

David A. Landgrebe

Analysis Research for Earth Resource Information Systems: Where Do We Stand?

by

David A. Landgrebe
Department of Electrical Engineering
and
Laboratory for Applications of Remote Sensing
Purdue University, West Lafayette, Indiana

For a decade or more research has been conducted which is intended to lead to the design of operational earth resource information systems. Clearly this research has not yet been completed. It does seem to be far enough along however, that an assessment relevant to future operational activities might well be possible and beneficial. In this paper a model operational system will be discussed not so much as a proposal for an actual system, but merely as a viable objective upon which to focus the research. Possible system configurations and constraints will be noted, and the trend in the cost for data processing will be discussed relative to such a system.

INTRODUCTION

Before providing this discussion however, it may be helpful to list the major elements of such an information system and to establish certain aspects of the system.[1] Figure 1 shows the major elements necessary in an earth observational system. There must be a sensor system viewing the surface of the earth; there will no doubt be the need for some type of on-board processing of the data, perhaps including tagging the data at this point with location and calibration information. This is followed by return of the data to the surface of the earth, perhaps through a telemetry system. Once the data is at a suitable ground site preprocessing of some type will prove desirable, perhaps including the merging of ancillary data. This will be followed by the application of suitable analysis algorithms after which the information derived will be utilized by an information consumer.

It should be emphasized at this point that this is not intended as a finalized system concept; for example the blocks in this system are not necessarily even in the correct order. It may be desirable to do preprocessing or analysis operations on board the platform or in other ways to rearrange the steps through which the data passes. This is not important for our present purposes, however. It is only important at this point to recognize that this group of activities will need to be carried out in an operational implementation of such a system.

SYSTEM TYPES AND ANALYSIS PROCEDURES

A survey of current remote sensing systems would show that there exists today a duality of system types, in that the present capability to process data springs from two quite different stems of technology. The names given to these two types here are "image-oriented" and "numerically oriented". The differences between these two may at first appear to be quite subtle, but they are quite far reaching in their effect. Let us therefore attempt to make the subtle differences a bit more obvious.

Figure 2 shows the essential elements of a block diagram of the two types in such a way as to emphasize the differences. Notice that in the image-oriented system the "form image" operation is in line with the data stream and is therefore a key step in the processing of the data. Numerical systems on the other hand need not necessarily have a "form image" step at all. If they do it will be at the side of the data stream for purposes of allowing the monitoring of system performance and identification of special situations which require other actions.

To illustrate further, Figure 3 shows images of an agricultural area in three different parts of the spectrum. Four different types of agricultural surface cover are shown. If it were desired, for example, to identify corn as distinct from the other three at this stage of the growing season, this could be done in the .4- to .7-micrometer image on the left by noting the distinctive texture of the corn field relative to the other three materials. Thus, an image oriented approach has been devised, and a very typical image characteristic, texture, has been adequate to provide the desired analysis.

On the other hand, a sample numerical approach might be devised by noting that the particular set of gray values from spectral band to spectral band for the corn field is different from sets for the other three materials. That is, the corn field appears, in order, medium gray, dark gray, light gray in the three spectral bands, and this particular combination does not show up for any of the other materials. Thus, a quantitative form of this gray code as a function of spectral region could be used by a numerically oriented system to identify that surface cover.

Figure 4 shows the same concept in a more graphic form. At the top is a hypothetical, though generally accurate, sketch of the response as a function of wavelength for three classes of material: vegetation, soil and water. To quantify the above concept one could sample this spectrum at two or more wavelengths, for example λ_1 and λ_2 as indicated. If this data is replotted in two-dimensional space, that is, dimension 1 being the response in band λ_1 and dimension 2 in band λ_2, the result will be as shown in the lower portion of the figure. To identify any one of the three materials relative to the others, it is only necessary to determine into which portion of the space a given sample falls. This is a decision which is easily implemented by machine.

Let us consider one example in which this subtle difference between system types referred to above becomes quite obvious. This example occurred during a study of the effect of noise on analysis results. In this study an attempt was being made to determine the degree of sensitivity of the analysis results to the signal-to-noise ratio of multiband multispectral data analyzed by the above numerical technique. A typical data set from an agricultural experiment was used as a reference set and two new data sets were generated by adding a small magnitude noise and a larger magnitude noise to the original data set. Figure 5 shows a portion of one spectral band of each of these three data sets.

An identical crop species analysis was carried out on all three data sets and the results tested for accuracy. The measured accuracy turned out to be a 79.7 percent correct classification for the reference set, 63.6 percent correct for the low-noise case, but 78.1 percent correct for the high-noise case. At first these accuracy figures appeared inconsistent and a procedural error was suspected. After thorough checking it turned out, however, that no procedural error in the analysis had been committed. However, a difference in the generation of the noise sets (other than their magnitude) had occurred. The noise added in the low-noise case was uncorrelated from spectral band to spectral band whereas the noise added in the high-noise case was completely correlated between spectral bands.

There are several conclusions which can be drawn from this experiment. Obviously in this case correlated noise had a much less degrading effect on the numerically analyzed results than did uncorrelated noise. Parenthetically, the uncorrelated noise of the low-noise case simulates the situation which sensor detector noise injects into the system. Since a different detector is used for each different spectal band, the noise occurring in the detectors can be expected to be uncorrelated from spectral band to spectral band. On the other hand, the correlated noise of the high-noise case may simulate the type of noise which the atmosphere tends to add to the system, since with a multispectral scanner such as on ERTS, for example, the energy for all spectral bands passes through the same column of atmosphere.

But note especially from this illustration that while it is possible to judge data quality relative to an image-oriented system by viewing the imagery, it is not possible to do so for a numerically oriented system. Clearly in the case of this example, the data from the high-noise case would be ruled very useless while that from the low-noise case would be ruled very useful if judged on its image qualities alone. Though this example is perhaps an extreme one, it is not an isolated instance in which the importance of recognizing the difference between systems with image orientation and numerical orientation is important. This difference will be discussed again briefly.

ON THE STATE OF THE SCIENCE RELATIVE TO OPERATIONAL SYSTEMS

Let us now assess the state of the science of extracting information from remotely sensed data. Doing so will permit us to more realistically predict the future in terms of operational systems. Although a more up-to-date assessment of the technology is now in the planning stages, the 1971 Corn Blight Watch Experiment provides a good vehicle for a current assessment.[3] Though in the strict sense this experiment was indeed an experiment, it was both large enough and had adequately operational-like attributes to it to make it very useful for our purposes here.

This experiment, which was conceived at the time of a pending possible national emergency, involved an estimated thousand people from federal and state agencies. Figure 6 shows a plan view of the experiment. A multistage sampling scheme was to be used, and there were to be two different areas involved. The one was the whole corn belt including portions of seven states in which the stages of the sampling scheme were in order: the corn belt, flightlines as shown in yellow, segments within the flightlines as shown in red, and finally individual fields within the segments and ground samples within the field. There were 210 segments within this seven-state area, which was flown over by an RB-57F every two weeks throughout the growing season. The second area was in the western third of Indiana in which there were 30 of the segments which were also overflown by a multispectral scanner aircraft.

In carrying out the experiment ground observations were taken in eight to ten fields of each segment at the same time of the aircraft overflights. The data from both the ground observations and the aircraft overflights flowed to an analysis station, was analyzed, and the results reported to USDA each two weeks in time to begin the next cycle.

The photographic data from the 210 segments were analyzed by photo-interpretation on a field-by-field basis using the ground observation fields as bench marks. In the intensive area, in addition to the photo-interpretative analysis, the 30 segments were analyzed by a numerical (i.e., machine processing) method.

It should be noted that the identification problem itself, which forms the basis of this experiment, is at the more sophisticated or more difficult end of the scale in that it requires more than identifying vegetation as compared to bare earth or water and even more than identifying the species of that vegetation; it required the delineation of the state of stress of that particular species.

The results of the experiment showed that not only was it possible to identify the species to high accuracy throughout a major portion of the growing season, but it was possible to delineate the degree of blight infestation into two or three different levels of plant stress quite accurately.

It should also be noted that the system hardware which was used was not designed for the above purposes. The hardware was selected, of course, based more on the fact that it was already available or could become so very rapidly rather than that it was optimally suited for the purpose. In the case of the photo-interpretative equipment, in several respects the equipment was probably near to an optimal choice. Photo-interpretation is, of course, a very well-developed art and much operationally useful equipment is already available "off the shelf".

This is not the case for the machine processing part of the system, however. It is this latter point which is to be discussed further for a moment because of the tremendous potential for further development which exists in this case. In order to do so let us first review the steps necessary in such an analysis.

The data from half of the 30 segments of the intensive area was analyzed using a software system known as LARSYS. This system (Figure 7) consists of a number of different processors which are used more or less in order upon the data.

After preliminary data preparation and preprocessing a display processor is used for field location and selection. A clustering processor is typically used next to help develop so-called training statistics for the data; a statistics processor is then used to calculate the needed statistics. This is followed by a feature selection processor which selects the best set of spectral bands to be used, and then the classification itself follows. The final step is a display of the classification results and results evaluation. Certain other processing steps are possible in other circumstances. This system has been implemented on a variety of different types of general-purpose comuuters. The implementation at LARS/Purdue is on an IBM System 360 Model 67 time-shared system.

TRENDS IN PROCESSING COSTS

With this background let us now consider the situation relative to processing costs. Table 1 shows information which was accumulated during the corn blight watch on the direct cost involved in analyzing 15 of the 30 intensive area test segments by these machine processing methods. Note that the total cost was about $22,000 for the 15 segments per mission or about $1,500 per segment (a segment consisted of about 350,000 data points).

It should be immediately pointed out however that this $1,500 figure has very little relevance in an absolute sense in that on the one hand a number of costs of a more indirect nature such as the collection of ground observations etc. are not included, while on the other hand such factors as the over-qualification of staff carrying out the exercise and the software implementation of the system tend to enlarge the figures. The purpose then of discussing them here is to see the relative mix of people and machine costs for the various processing steps in order to predict what direction the technology is taking.

For example, current sensors, such as the ERTS satellite and the 24-channel multispectral scanner aboard the NASA C-130, carry out analog-to-digital conversion at the sensor, thus eliminating it as a variable cost. This same factor would eliminate much of the preprocessing which was needed in the corn blight watch to reformat the A/D data to a suitable form.

It was inherent in the design of the Corn Blight Watch Experiment that an unusually large amount of field location activity had to be carried out due to the experimental nature of the task; it was necessary to thoroughly test and verify the quality of results being obtained.

These are examples of factors which affect costs and which already exist in the technology. What can be said about the direction of future developments? There are two distinct elements to this technology flow of development which are relevant here. One is the relation of the size of the personnel costs to machine costs and the other is the machine cost itself. We will discuss these in order.

It is accurate to say, perhaps, that five years ago the distribution of cost would have shown a much higher portion in the people column as compared to the machine column for a typical analysis effort. An underlying objective in the development of new technology in this area has been to transfer the cost of analysis from personnel to machine. This trend will continue and for good reasons which will be discussed presently. But first consider the several ways in which this can be done.

The purpose of cluster processing has been to decrease the amount of manual effort involved in field location and selection for training purposes; research is continuing in this area which will no doubt lead to a further reduction in the personnel cost associated with field location and selection and cluster processing. Much has already been done to decrease the amount of manual activity involved in spectral band selection. Indeed for some sensors such as ERTS this step is not required at all.

Two further comments which are parenthetical to the people-to-machine cost distribution are in order at this point. First, in the future the proportion of total cost attributed to the classification step will undoubtedly be greater since it is quite possible in a single analysis now to analyze data from a very much larger number of scene elements; and second, it will shortly be argued that the cost of the machine portion of the processing can be significantly reduced. As this comes about it will prove desirable to do types of preprocessing which are not now normally conducted. I refer here to the registration of data from mission to mission thus making available temporal information about the scene and at the same time achieving a type of very precise scene correction of the data making it have very high quality cartographic properties. A greater component of cost in the future will also be in the Results Display area because of the larger diversity of formats in which the user community will require to have results.

Summarizing to this point then, it has been argued that the march of new technology is moving cost from manual activities to machine ones and eliminating some portions of the processing steps all together. The motivation for relying more proportionally on the machine portion of the system stems mainly from the second element of technology flow, i.e. the cost reduction which is possible in machine processing itself. Let us examine this point briefly.

There are at least three categories of machines relative to processing costs. These are, (1) general-purpose processors, (2) more specialized, but generally available computers, and (3) specially designed and built digital processing hardware. Significant decreases in processing costs are occurring in all three categories.

Much is being said in the computer industry of the manner in which the cost of general-purpose processors is decreasing. Although there may not be complete agreement on the rate at which this cost reduction is occurring few would dispute that it is of very significant size. Figure 8 shows one estimate of how it is occurring.[4] Note, for example, that in the large general-purpose class of machines a reduction in costs by an order of magnitude is predicted within this decade. The paper from which this figure was taken is more than a year old and new machine announcements since the time it has appeared generally tend to support this trend.

The second category of machines referred to above was the generally available, but more specialized computer. An example here would be the Illiac IV computer[5] which will go on line during the next year at the NASA-Ames Research Center. This machine represents a tremendous capacity for computation. Calculations indicate that it will do an amount of analysis computation per hour which is more than two orders of magnitude greater than an IBM 360/Model 67 computer while the cost figures associated with it are something less than one order of magnitude greater. Thus this would provide a potential cost gain if it can all be realized of perhaps more than a factor of ten.

The third category of machine, that of the special purpose specially designed hardware can probably be made even more cost effective. The price to be paid in each case is that of a higher initial investment and a decrease in available flexibility. In the case of the general-purpose machine, software is generally already available in usable form or can easily be modified. In the second category, one must devise and specialize software while in the third case both hardware and software must be committed to. The range of flexibility which is available decreases rapidly in moving from one category to the next. At the current state of technology and given the rapid advance which is now taking place it seems perhaps premature to commit to special purpose hardware due to the rapid obsolescence which is likely to take place.

Before leaving the matter of cost, one additional area should be treated, that of the cost of mass storage of data. Figure 9 shows Withington's projection on the cost of data storage over the next decade.[4] Notice that a reduction by 2 to 3 orders of magnitude seem possible to him. If this proves to be the case then the consideration of storing on line a volumn of data as large as, for example, one year's worth of ERTS data over the entire U.S. would probably be straightforward.

Summarizing, it is suggested that in the area of future costs relative to machine processing of data, further reduction in the manual portion of the processing activities are likely, and very significant reductions in the machine processing portions also seem in the offing. Proceding from these points then, let us turn to the manner in which the data stream portion of the system should be designed.

RECONCILING NUMERICAL TO IMAGE-ORIENTED SYSTEM DESIGN CONSTRAINTS

Recall the earlier discussion with regard to the duality of system types. It is suggested here that the data system hardware portion of the total system should be designed in terms of a numerical orientation since this tends to pose the most stringent requirements upon the system. At the same time it is argued that this choice is entirely compatible with image-oriented analysis tasks. Indeed, there may even be advantages for image-oriented systems in doing so. This latter statement will be illustrated using the ERTS system as an example. The MSS sensor system of ERTS with its onboard analog-to-digital conversion and digital telemetry system to the point of delivery of computer-compatible tape is very compatible with a numerically oriented system. We are all now familiar with the high-quality imagery which can be produced from this system. Figure 10 shows a standard product of such imagery. This particular image is of the south end of Lake Michigan showing Chicago, Gary and Hammond, Indiana and surrounding regions. Examining this photo using photographic enlargements would show the great amount of detail that is present in this imagery.

By remaining in a digital format up to the production of the final image, however, perhaps even greater detail is possible. Figure 11 shows an image from the same data but taken from the computer compatible tape. Though shown in black and white here, it was originally produced in color using bands 4, 5 and 7 of the ERTS data in a fashion similar to the production of simulated color infrared images. However in this case, rather than properly balancing the three colors relative to one another, a maximum in contrast was sought in each of the three spectral bands. The increasing amount of detail which can be displayed when continuing to use the data in digital form as the source from which the image is produced can be shown by referring to Figures 12, 13, 14 and 15.

SYSTEM DESIGN AND THE USER

Let us now return to some aspects of overall system design. Recall the major elements of the total system from Figure 1. The first point to be made here is that there is not a single user but a large number of them (Figure 16). Thus one has a single centralized sensor system at one end of the system and a large and diverse number of users at the other. A major question to be decided is the point at which the decentralized need should take precedence over the need for keeping the system centralized in order to achieve economies of scale. In the ERTS system, for example, the point of decentralization is after a portion of the ground preprocessing has been done, but before any analysis.

Taking a look at an additional system aspect, one must realize that in changing data to information, it may be necessary to pass the data through more than one (sequential) analysis step (Figure 17). For example, it may be necessary to first derive a map from the data showing a certain set of earth surface cover types, then analyze the map to derive the needed information. Different users may even require different maps of the same area (derived from the same data) from which to carry out subsequent stages of the analysis.

This would seem to impose constraints on the centralization of the system. But at the same time a general purpose system capability to derive maps and statistics of the occurrance of classes of earth surface types is quite fundamental to any system and is certainly possible at the present state of the technology. Thus one is led to conclude that in order to allow the user individualized opportunities for analysis, decentralization of the system data stream can begin not later than during the analysis step. All other things being equal, for reasons of economy, it should not begin any earlier.

ON THE AVAILABILITY OF THE TECHNOLOGY TO THE USER

It is now possible to discuss the final topic, that of how this rather complex and sophisticated technology can be made available cost-effectively to a large and diverse user community. There are at least three major elements to the availability of this technology at the user end of the system. These are hardware, which was previously discussed, software and training or understanding as to how to utilize the system. The following description of an experiment in making available this type of technology to a user community encompasses these three elements. The scheme being tested is the so-called multi-terminal system in which a relatively large central processor is utilized by means of terminals which are located at the user's location. This scheme allows not only for the centralization of processor hardware and software but also of the data base, relieving the necessity of shipping large quantities of data around the country. At the same time, it provides complete control of a set of standard analysis steps by the user. Only processing instructions and results need be shipped to remote sites. Note specifically that this arrangement incorporates not only the conditions of centralization but user flexibility and control just discussed above.

This experiment was first proposed to NASA in 1970. Its current status is that a suitable software system including LARSYS has been implemented on a relatively large machine at Purdue and that experimental remote sites have been established. These currently are at the Johnson Space Center in Texas and the Goddard Space Flight Center in Maryland. A third site at Wallops Island is expected to go on line shortly.

While it is early to draw definite conclusions about the results of this experiment, a significant amount of data is being analyzed on the system at this time. In addition, the system is proving very helpful in training individuals in using this type of technology regardless of how implemented. At NASA/Houston, for example, more than 60 individuals have now been trained on the use of the remote terminal system. Because of the centralization of the hardware and software, it is economically feasible to develop standardized training materials such as programmed texts, audio-tutorial tape/slide sets and the like which can be used at very low cost. These packages are now also under development and testing.

Should this experiment prove out the feasibility for using such a computer network system nation-wide on a cost-effective basis, operation of such a system could certainly begin quickly. Networks of computers and users for other applications are already in operation. The ARPA Net is one such system which now has services available for rent and is being used at a number of sites across the country for routinely carrying out processing tasks in one location using computers in another. Illiac IV will be connected to the ARPA Net. In addition to the ARPA Net at least two private organizations have requested permission from the government to engage in a network activity on a commerical basis.

SUMMARY

In summary, the author has discussed the state of the technology of earth resources information systems relative to future operational implementation. It was suggested here that though there has been a duality of system types in the past, these two system types are not incompatible with one another if the data system involved is properly designed. It was suggested that the cost of machine processing may be expected in the near future to decrease very rapidly. Some aspects of interfacing such an advanced technology with an operational user community were discussed so as to accomodate the user's need for flexibility and yet provide the services needed on a cost-effective basis.

ACKNOWLEDGEMENT

The work described in this paper was sponsored by the National Aeronautic and Space Administration (NASA) under Grant No. NGL 15-005-112, and has been presented at the 19th Annual Meeting of the American Astronautical Society, Dallas, Texas, June 19-21, 1973.

REFERENCES

1. Landgrebe, D. A., "A Systems Approach to the Use of Remote Sensing," Proceedings of the International Workshop on Earth Resources Survey Systems, Ann Arbor, Michigan, May 3-14, 1971, NASA Special Publication SP-283.

2. A Series of Papers Reporting on the 1971 Corn Blight Watch Experiment in the Proceedings of the 4th Annual Earth Resources Program Review, NASA/MSC, Houston, Texas, January 17-21, 1972:

   Johannsen, C. J., Bauer, M. E., and Staff*, "Corn Blight Experiment Results.

   Bauer, M. E., "The Corn Blight Problem - 1970 and 1971"

   Allen, Richard, "Corn Blight Review - Sampling Model and Ground Data Measurements Program

   Blilie, Ronald K., "Aircraft Data Acquisition"

   Phillips, Terry L. and Staff*, "1971 Corn Blight Watch Experiment Data Processing, Analysis and Interpretation"

   Swain, Philip, "Experiment Results Ground Measurements, Photo and Multispectral Machine Analysis"

   Nalepka, R. F., Morgenstern, J. P., and Brown, W. L., "Detailed Interpretation and Analysis of Selected Corn Blight Watch Data Sets"

   Clifton, J. W., "1971 Corn Blight Watch Experiment"

3. MacDonald, R. B., Allen, R., Clifton, J. W., Bauer, M. E., Landgrebe, D. A., and Erickson, J. D., "Results of the 1971 Corn Blight Watch Experiment," Proceedings of the Eighth International Symposium on Remote Sensing of Environment, Ann Arbor, Michigan, October 2-6, 1972.

4. Withington, Frederic G. "The Next (and Last?) Generation", Datamation, May 1972. pp. 71-74.

5. Bouknight, W. J., Denenberg, S. A., McIntyre, D. E., Randal, J. M., Sameh, J. H., and Slotnick, D. L., "The ILLIAC IV System," Proceedings of the IEEE, 60:369-388, April 1972.

*"Staff" refers to the staff of the Laboratory for Applications of Remote Sensing (LARS), Purdue University, W. Lafayette, Indiana.

Table 1. Corn Blight Watch Experiment Processing Costs Per Mission (8 Missions).

	People	Machine
A/D Conversion	407	2544
Preprocessing	1212	6050
Field Location/Selection	648	3225
Cluster Processing	345	1935
Statistics Calculation	—	1935
Spectral Band Selection	205	1290
Classification	—	1451
Results Display and Summarization	259	645
Total	$3,078	19,075

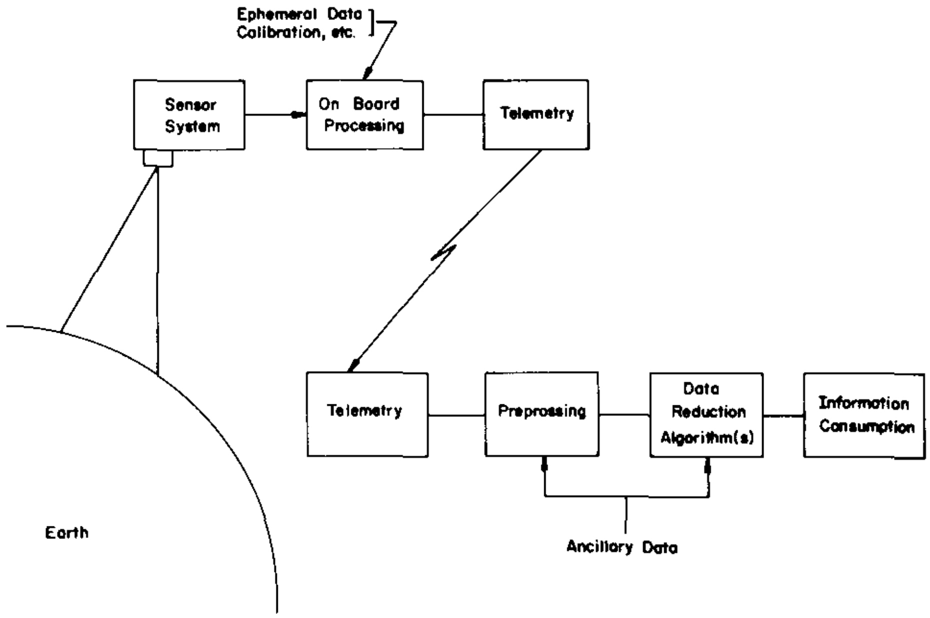


Figure 1. The Major Elements of an Earth Observational Information System.

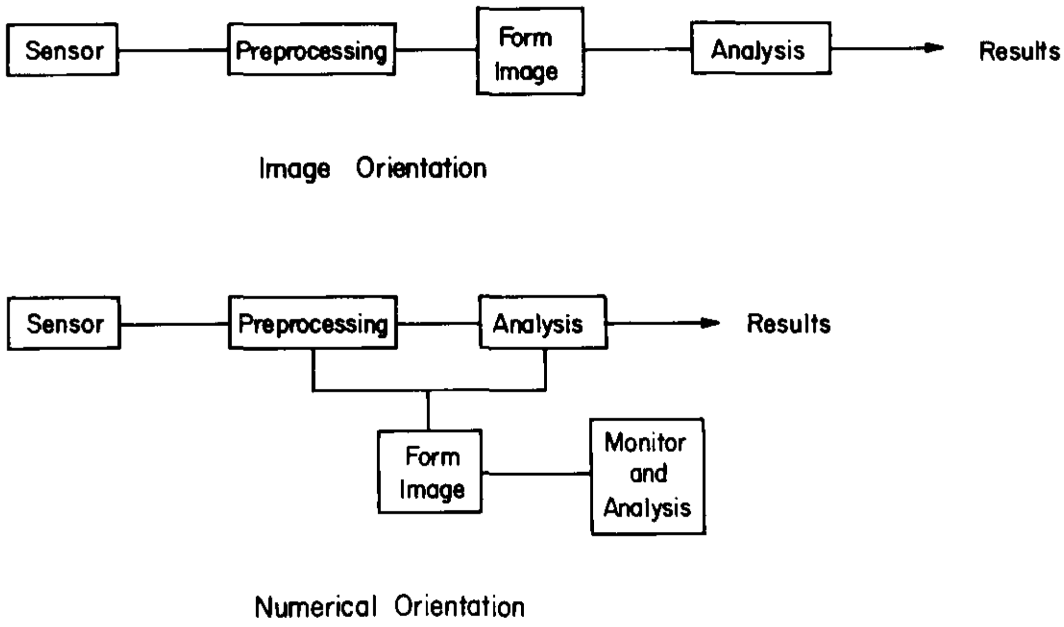
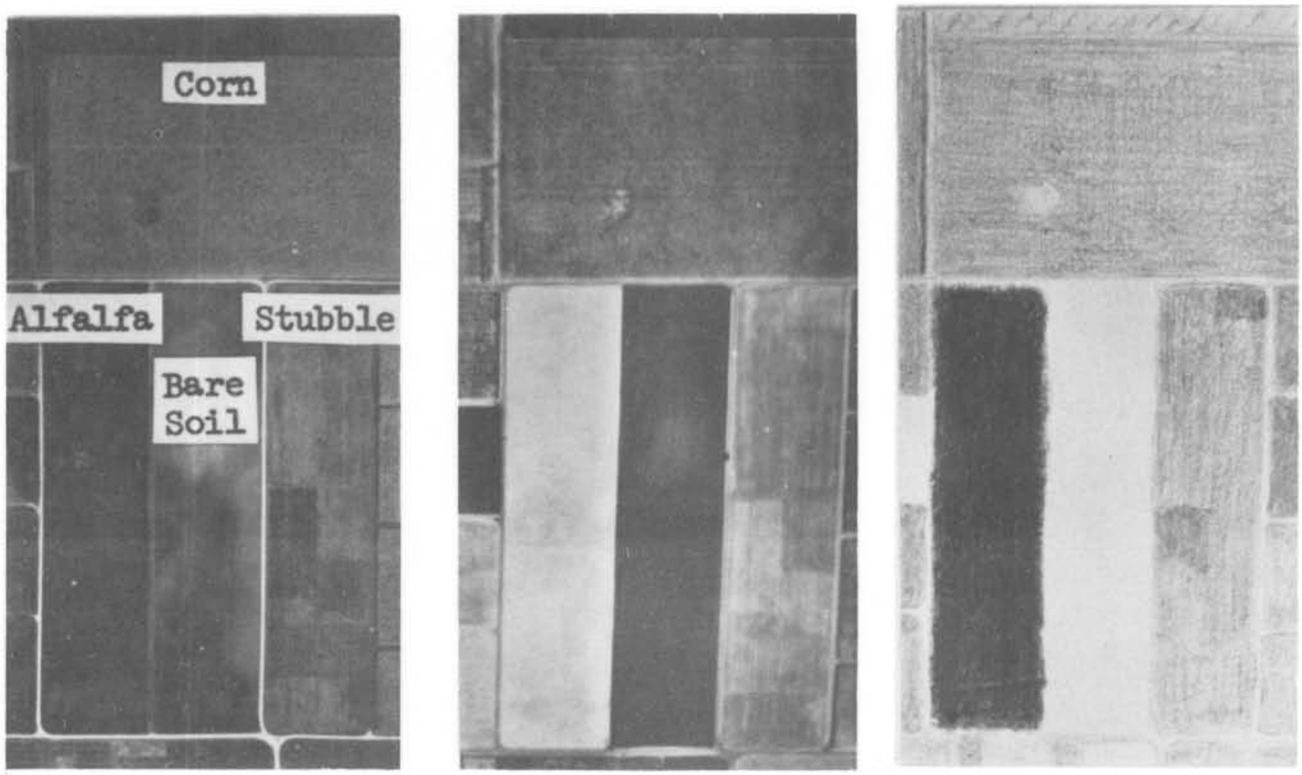


Figure 2. Organization of Image and Numerically Oriented Systems.



.4-.7 μm

.7-.9 μm

4.5-5.5 μm (Artist's Concept)

Figure 3. Multispectral Response of Corn, Alfalfa, Stubble, and Bare Soil.

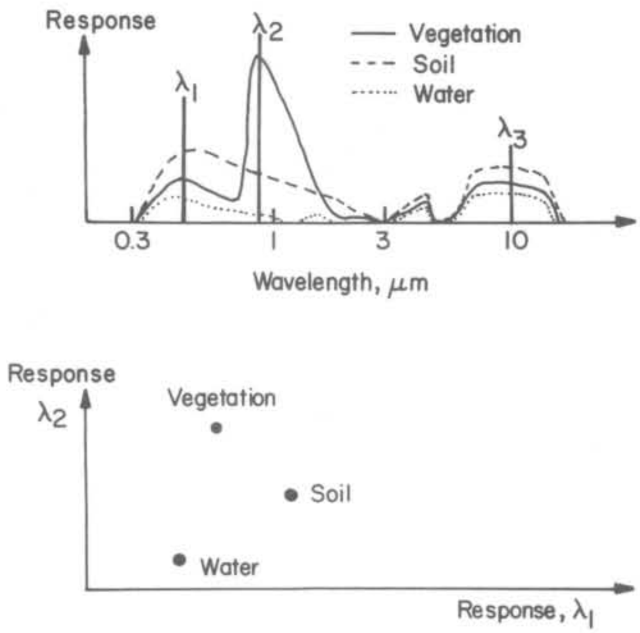
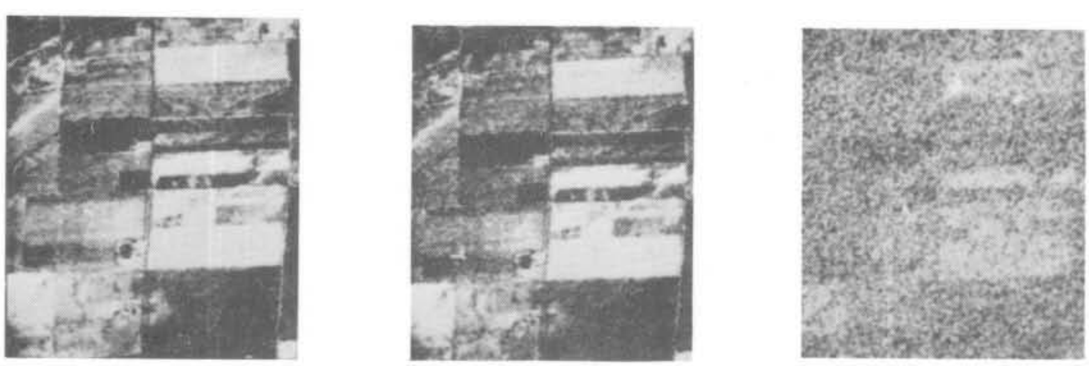


Figure 4. Spectral Data in Two-Dimensional Feature Space.



Reference (Original Data)	Low Noise	High Noise
79.7%	63.6%	78.1%

Figure 5. The Effect of Noise on Analysis.

Figure 6. A Plan View of the 1971 Corn Blight Watch Experiment Sampling Scheme.

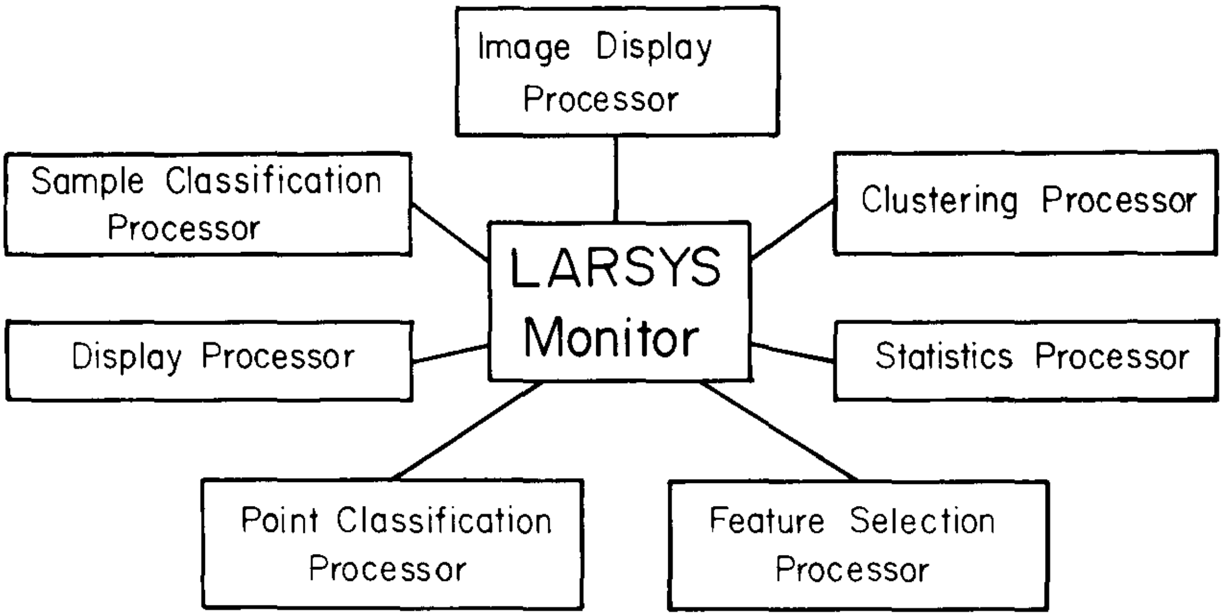


Figure 7. The LARSYS Multivariant Data Processing Software System.

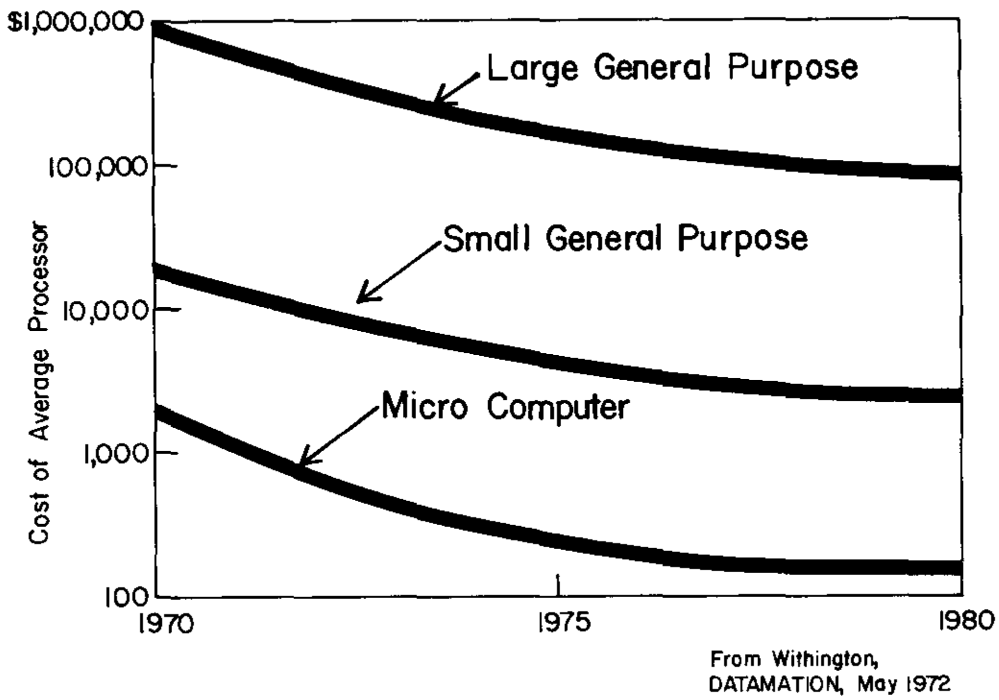


Figure 8. The Cost of Central Processors by Withington, Datamation, May 1972.

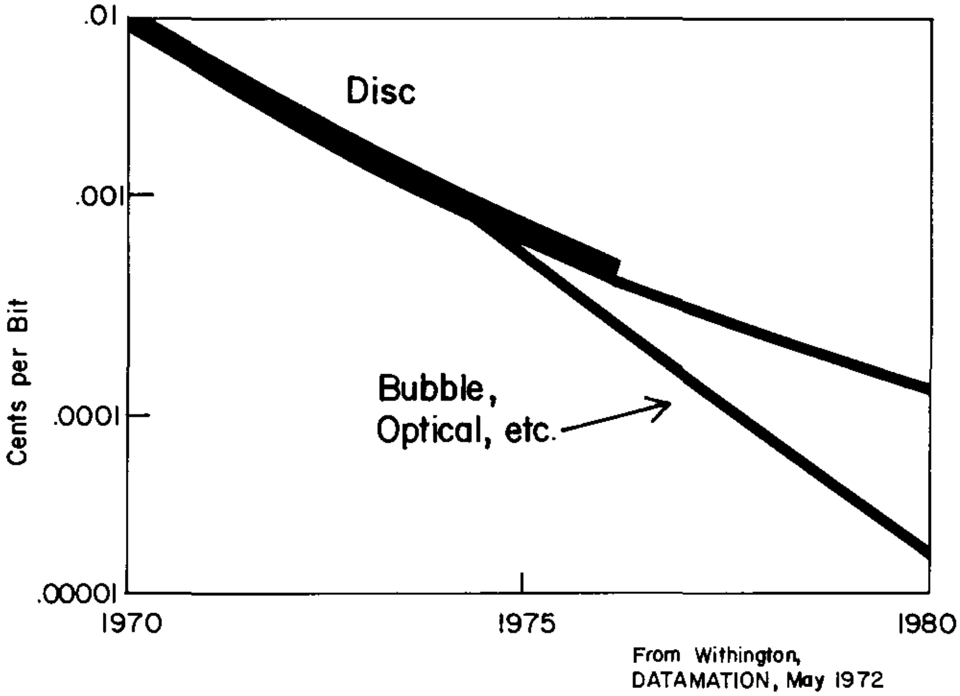


Figure 9. The Cost of Mass Storage by Withington, Datamation, May 1972.

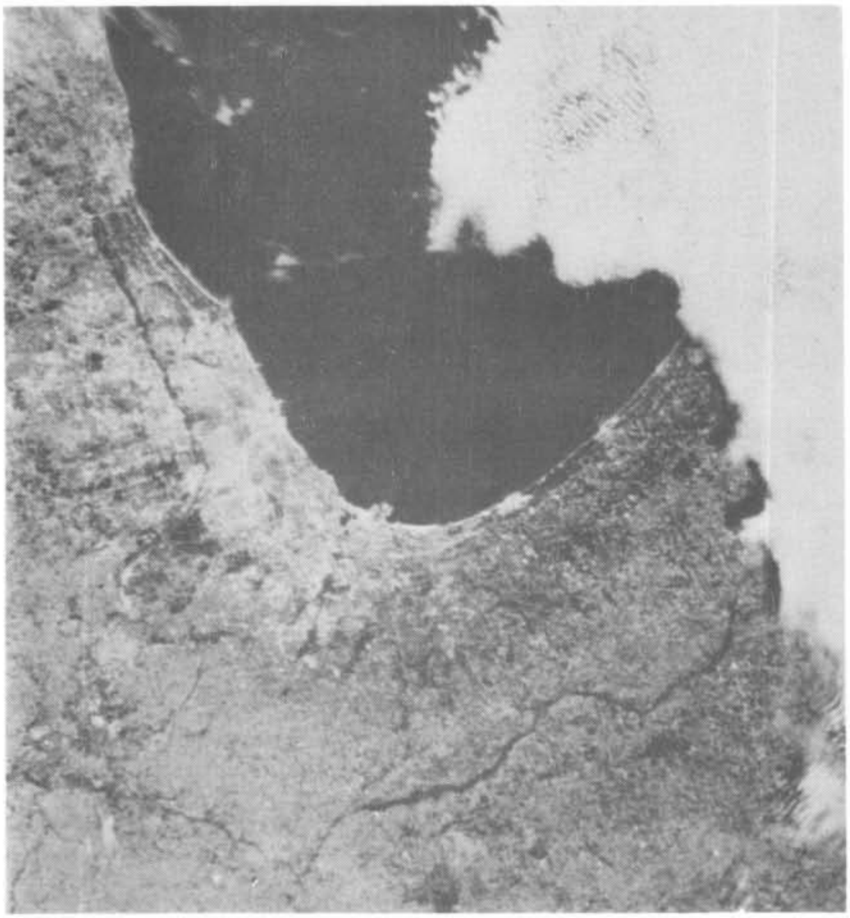


Figure 10. An Example of ERTS Data in Image Form: The Chicago Area on October 1972 in the 0.6-0.7 Micrometer Region.

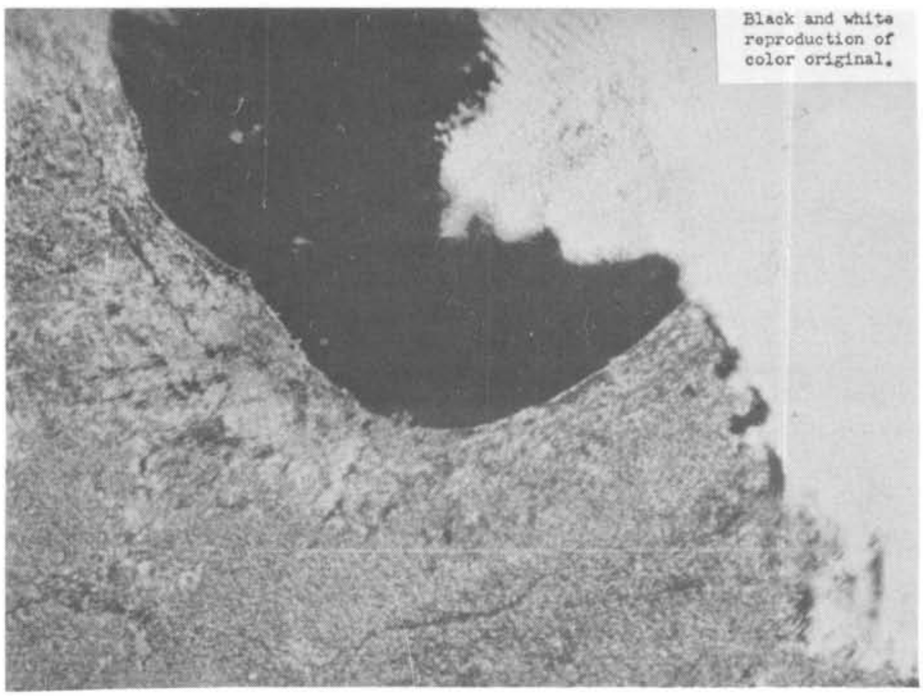


Figure 11. A False Color Version of ERTS Data Produced by Computer.

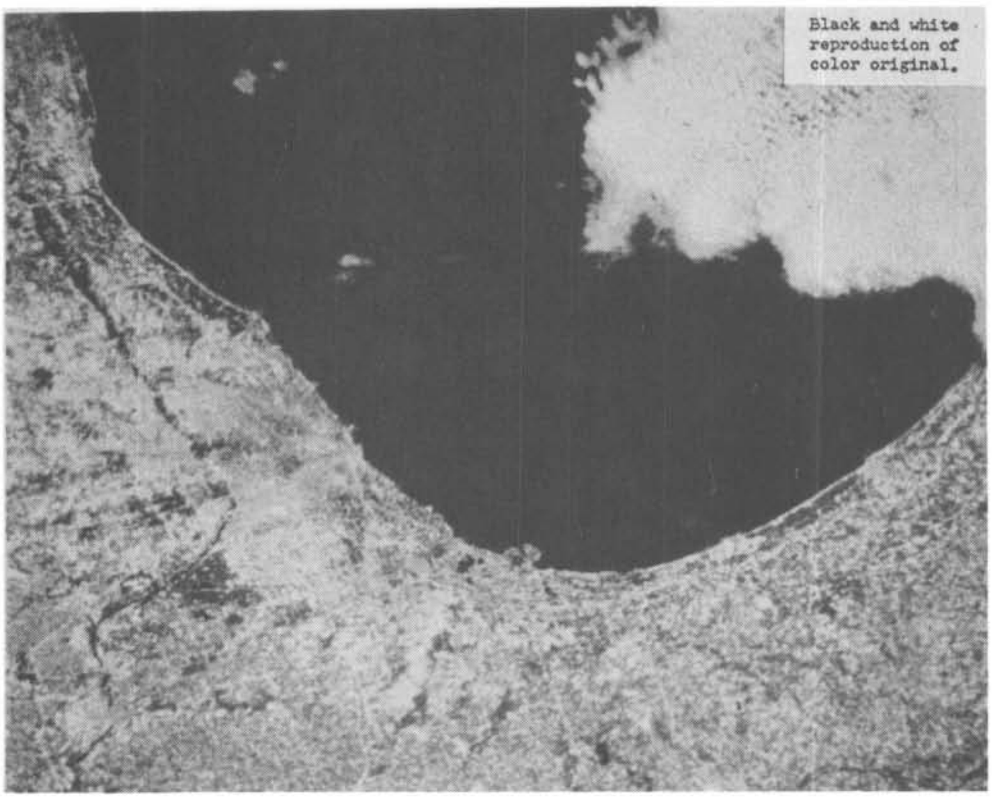


Figure 12. ERTS Image Produced by Computer with Increased Detail Displayed.

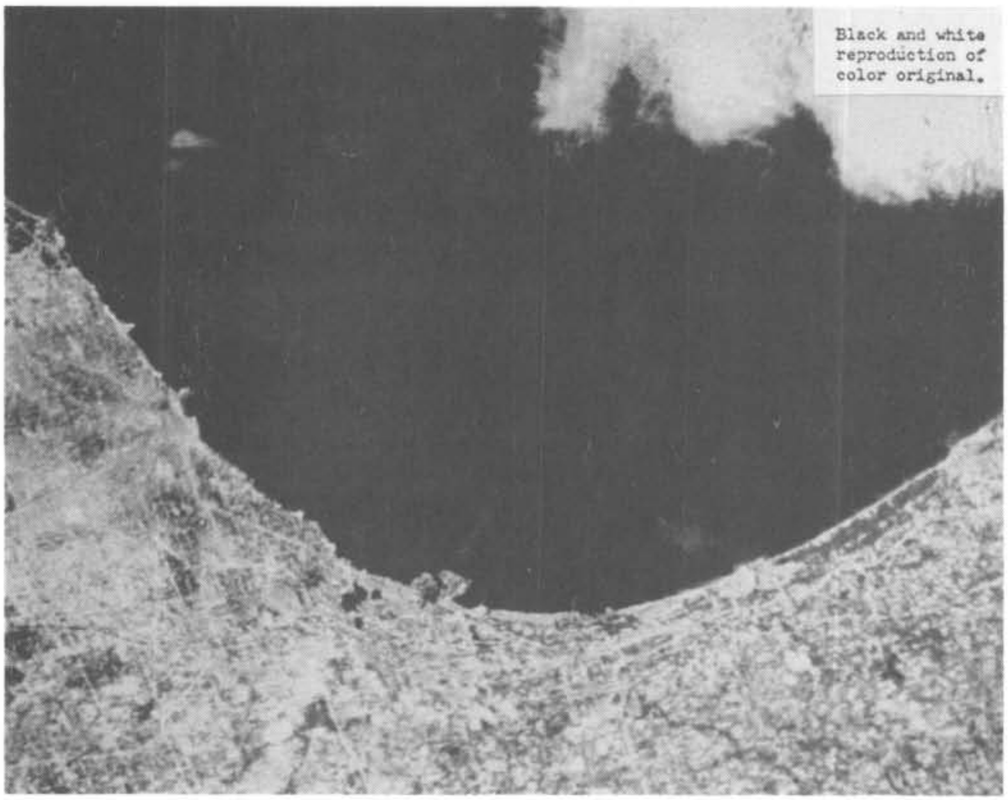


Figure 13. ERTS Image Produced by Computer with Increased Detail Displayed.

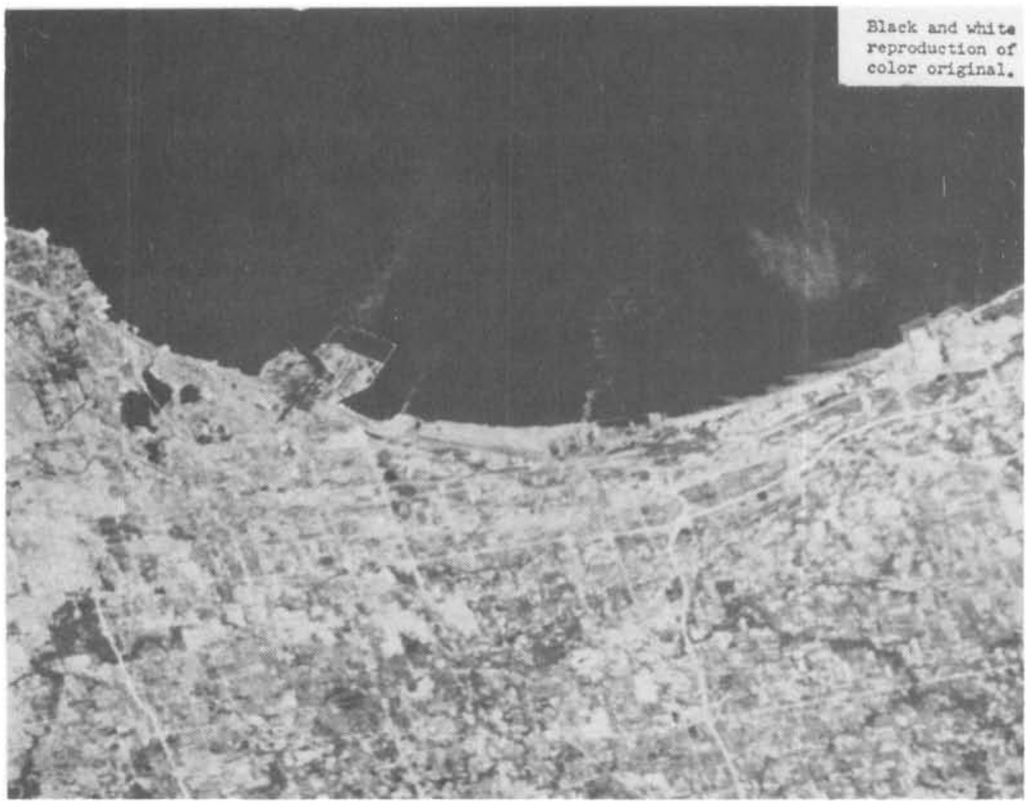


Figure 14. ERTS Image Produced by Computer with Increased Detail Displayed.

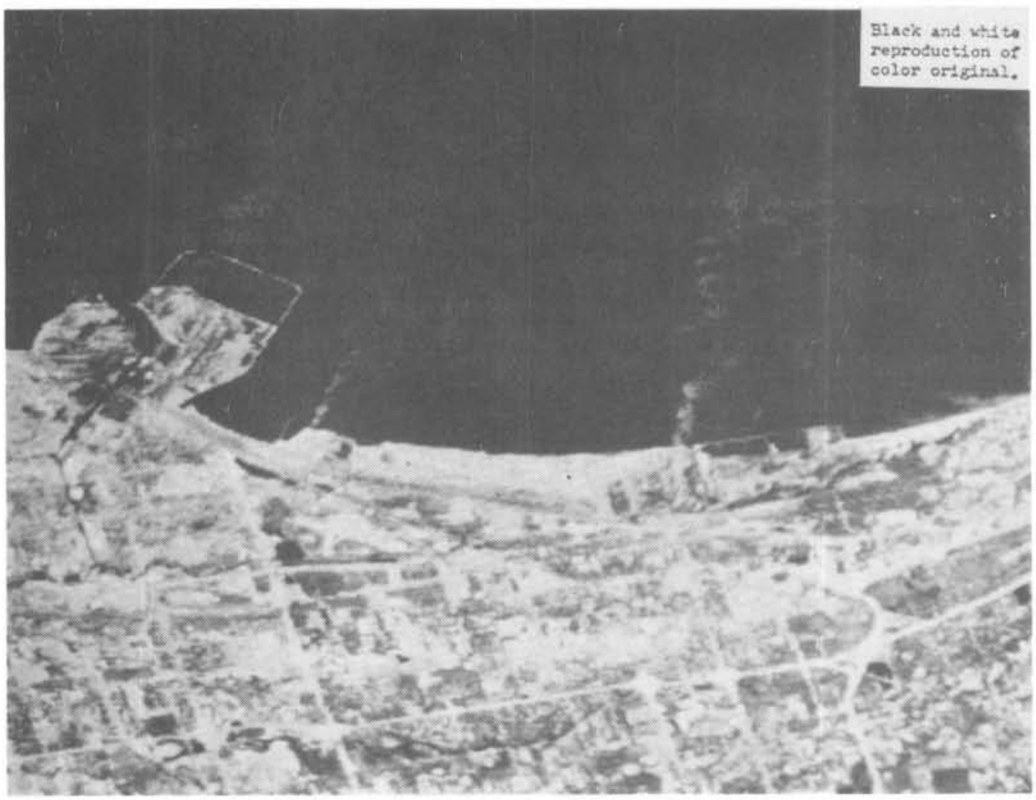


Figure 15. ERTS Image Produced by Computer with Increased Detail Displayed.

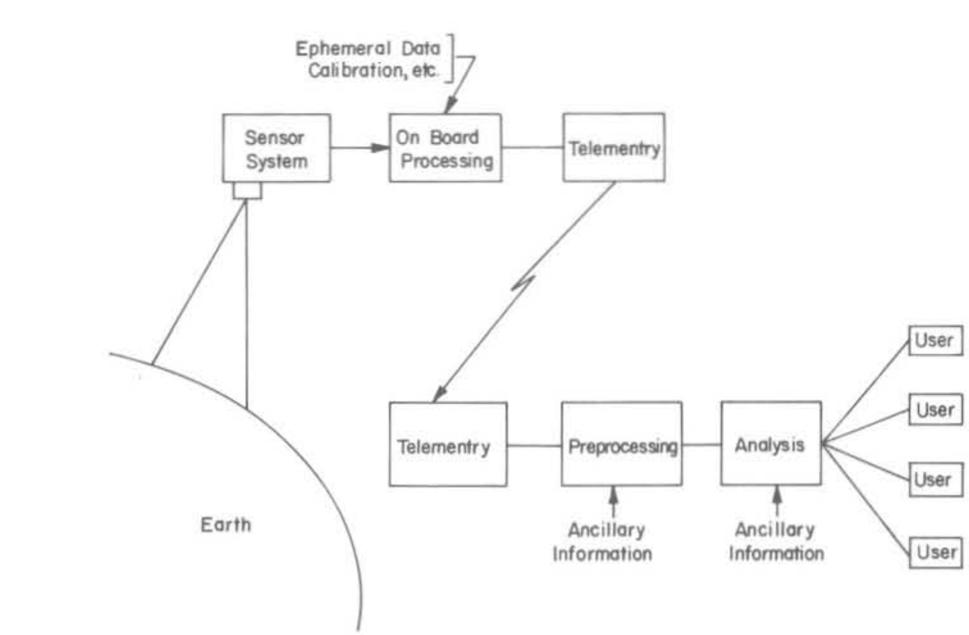


Figure 16. The Major Elements of an Earth Observational Information System Emphasizing the Fact that There are Many Users.

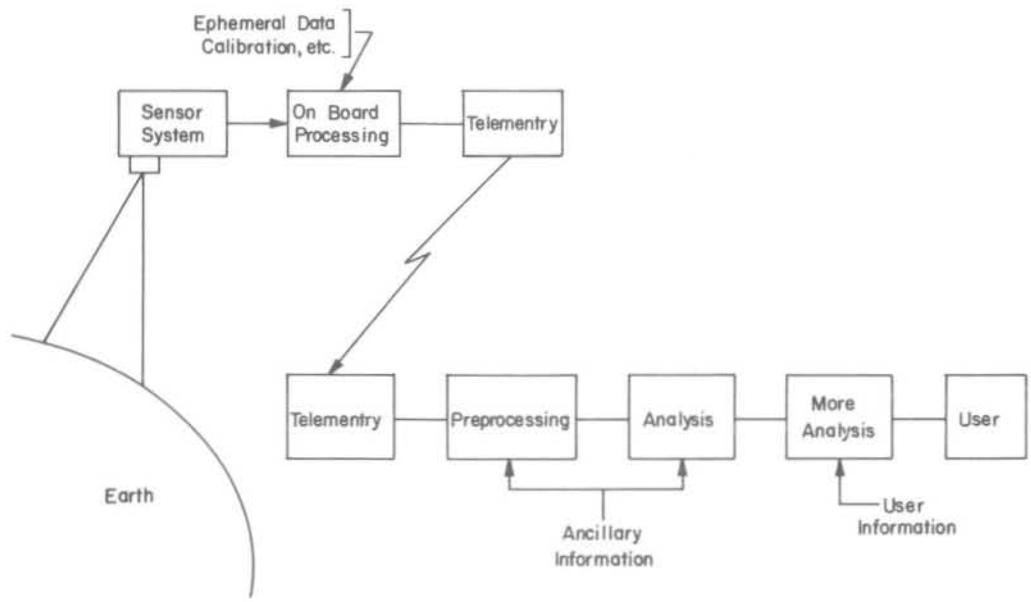


Figure 17. The Major Elements of an Earth Observational Information System Emphasizing the Users Part in the Analysis of Data.