

A Method for Multisensor–Multispectral Satellite Data Fusion

ANDREW S. JONES, KENNETH E. EIS, AND THOMAS H. VONDER HAAR

*Department of Atmospheric Science and Cooperative Institute for Research in the Atmosphere,
Colorado State University, Fort Collins, Colorado*

(Manuscript received 12 November 1993, in final form 8 December 1994)

ABSTRACT

Multisensor–multispectral scientific data applications require a tremendous investment regarding data preparation and analysis. A data fusion method is developed that is general enough for use with any scan-line-based datasets (satellite and ground based) and enables multisensor–multispectral datasets to be merged on a routine basis. In particular, advanced software processing techniques in the Polar Orbiter Remapping and Transformation Application Library (PORTAL) are used to show how polar (e.g., Special Sensor Microwave/Imager data from DMSP) and geostationary data (e.g., Visible–Infrared Spin Scan Radiometer data from GOES) can be spatially combined routinely in an efficient manner. The PORTAL system can also be used to combine data from sensors on the same satellite that have radically different earth scan patterns and ground resolutions. A self-describing generalized data format is used to modularize the data processing flow and obtain significant improvements in terms of flexibility, extensibility, and generality of application. A new data fusion method is introduced that consists of a comprehensive set of data fusion and merger software tools that generate merged datasets in an original satellite projection space. Computational efficiencies are compared between this data fusion method and that of conventional remapping methods. While comparable processing times are needed to physically merge the datasets, results show significant performance gains on any subsequent analysis of the merged datasets since scientific algorithms operate within the original satellite projection space. Effective microwave surface emittance is retrieved from multisensor–multispectral data as a demonstration of the technique.

1. Introduction

The ability to easily use multisensor–multispectral satellite data is becoming increasingly important to remote sensing research. Methods that combine multiple satellite datasets are, in general, complex and labor intensive. Much of the effort to process the data is application specific, and modifying code for use with another satellite sensor requires a considerable amount of redundant programming (Botts 1992). Many potential scientific users of satellite data in the geosciences are unwilling to pay the price in time and resources to use satellite datasets. Our own experience with the derivation of cloud liquid water over land surfaces using the polar Special Sensor Microwave/Imager (SSM/I) and the Visible–Infrared Spin Scan Radiometer (VISSR) on a geostationary satellite (Jones and Vonder Haar 1990) made us aware of the inefficiencies and difficulties involved with pre-1990 approaches in combining multispectral–multisensor satellite datasets.

The relevance of methods to easily and accurately combine multisensor–multispectral datasets is obvious when you consider the number of sensors that will be available on future satellites developed for the NASA

Earth Observing System (EOS) (NASA 1991). Data will likely be analyzed beyond the usual sense—that is, using one single dataset at a time—but in new ways that will routinely involve comparing datasets from multiple sources (Kahn et al. 1991). Another trend is toward more numerous and smaller objective-oriented satellite platforms within the EOS Program (King 1992). These two developments drive the need for more efficient, easy-to-use multisensor–multispectral and multiplatform data fusion methods. In recognition of the enormous task at hand, NASA's EOS Data and Information System (EOSDIS) is designed to support scientists so that they may use the massive amounts of data from EOS more efficiently (NASA 1992). Methods that improve the efficiency of multisensor–multispectral data processing and extend its use from special-case studies to general scientific use are therefore a necessity.

In addition to traditional remote sensing applications, the integration of remote sensing with Geographical Information Systems (GIS) poses similar problems that need to be addressed. Among them is the ability to handle multiple source datasets while encouraging the use of the data in a sense that allows the data quality to be known to the user (Star et al. 1991). Inherent to GIS (as well as to traditional remote sensing) is the problem of error propagation in the complex data integration procedures. Potential error sources

Corresponding author address: Andrew S. Jones, CSU/CIRA, Foothills Campus, Fort Collins, CO 80523.
E-mail: jones@phobos.cira.colostate.edu

must be identified to enhance the quality of the integrated remote sensing and GIS data (Lunetta et al. 1991). Analysis methods that combine multisensor-multispectral datasets are a source in this error propagation path. Attempts to simplify the procedure and provide a uniform efficient method to routinely combine multisensor-multispectral remote sensing datasets would encourage better error propagation analysis and improve the final integrated remote sensing GIS products. Software for such systems continues to be a critical limiting factor (Frank et al. 1991).

Previous examples of work using multisensor-multispectral data include remote sensing of sea ice (Comiso 1983; Steffen and Schweiger 1990; Comiso et al. 1991), characterization of vegetation (Paris and Kwong 1988), land classification (Welch and Ehlers 1987), and many others. The remote sensing of sea ice by Comiso et al. (1991) used the Synthetic Aperture Radar (SAR) (active microwave), Electrically Scanning Microwave Radiometer (ESMR) (passive microwave), and Airborne Multichannel Microwave Radiometer (AMMR) (passive microwave) data, in addition to numerous other data sources. The combination of the datasets showed that the active and passive microwave sensors generally complement each other as the sensors were sensitive to different physical properties of the sea ice. The work by Steffen and Schweiger (1990) required that Landsat Multispectral Scanner (MMS) (visible), Advanced Very High Resolution Radiometer (AVHRR) (visible and infrared), and SSM/I (passive microwave) data be merged for validation and analysis of sea-ice conditions. The technique used in this case required regridding of the 80-m resolution Landsat MMS data to match a SSM/I grid format at 25-km resolution (polar stereographic projection). Characterization of vegetation by merging Landsat MMS, Thematic Mapper (TM) (infrared), and Shuttle Imaging Radar (SIR-B) (active microwave) data was accomplished for a classification schema (Paris and Kwong 1988). Welch and Ehlers (1987) employed methods to enhance the resolution of land classification datasets based on the Landsat TM data, using SPOT high-resolution visible (HRV) data for more effective visual interpretation of the imagery. Work such as this exploits the complimentary nature of the remote sensing multisensor datasets to gain information that otherwise would be unavailable due to ambiguities and/or resolutions of the monosensor datasets.

a. Definitions and background

Data fusion has been defined as a generic term for the general field of applied methods used in integrating and displaying data from multiple sources and is one approach to the analysis of complementary remote sensing imagery (Franklin and Blodgett 1993). This term has been traditionally used for methods that add spatial detail to data by combining the sensors via color

functions or spatial filter techniques. The simplest technique of these is to register different datasets together and to create color displays by inputting a different sensor into each of the three color guns of the display device. A similar approach is to use a color transformation algorithm such as hue-saturation-lightness (HSL) (Ehlers 1991). Alternative data fusion definitions are related to intelligent data management issues, which are more abstract in nature (i.e., the focus is on how different datasets are coordinated and gathered with an emphasis on *how* to provide tools to facilitate their eventual combined use) (Campbell et al. 1989; Campbell and Crompt 1990). In this paper, our use of the term *data fusion* is similar to that of Franklin and Blodgett (1993) with the exception that the display process is not an integral part of the data fusion method as will be discussed later.

A primary need involved with any aspect of data fusion of multisensor datasets lies in the navigation and registration of the datasets. Two approaches are possible. The first takes the geographical overlay and distorts the overlay to the satellite image projection (*direct image referencing*) and the second method manipulates the image by resampling and fitting to a desired geographical map projection (*inverse image referencing*) (Emery et al. 1989; Ho and Asem 1986). The first method, direct image referencing, produces geometrically unique images that cannot be registered with other images (including subsequent passes of the same satellite). The second method produces images that are registered and is thus the most useful for merging multiple satellite datasets together and is also commonly known as *remapping*, *regridding*, or *image rectification* (Bernstein et al. 1983).

Two types of techniques are used to merge multisensor datasets using the inverse image referencing method:

- 1) *remapping* (or *regridding*) of a secondary dataset to a reference dataset projection space, which may itself have been gridded,

- 2) *enhancement* of a primary dataset via deepening of the data information content with the use of spatial and/or color enhancements that merge the secondary dataset information in a combination of color table manipulations (e.g., red-green-blue or HSL algorithms) and/or spatial filters that are then applied to the primary image (Ehlers 1991).

The enhancement method contrasts to the remapping method in that the final output is a single rectified image, whereas the remapping method may generate multiple rectified images, one for each spectral channel (Ehlers 1991). While both methods are quantitative, the enhancement method lends itself more readily to a display and visualization focused approach toward data utilization. Results from the remapping method are more abstract (i.e., having n -channel dimensions) and, thus, can be removed from the data visualization

process. It will be shown later that separation of the visualization process from the data analysis is considered an advantage.

Remapping can be abstractly thought of as two functions: the manipulation and coregistration of the data's spatial attributes, and the manipulation and interpolative spreading of the data values. The essence of remapping would of course be the manipulation of the data's spatial attributes; however, the data interpolation method used is also important. All data have associated attributes of time, space, and other sensor-dependent characteristics such as sampling and measurement precision and accuracy. Thus, when remapping is performed, which manipulates the data's spatial attributes, the *meaning* ascribed to the associated data values has changed. The goal of data interpolation in the context of remapping is to produce the best objective estimate of the data value that would have been measured had the data value been measured in its new spatial context (i.e., within the remapped coordinate system), including sensor-dependent characteristics. The data value that results from the interpolation procedure is only approximate. Therefore, some "error" is introduced in all remapping operations. It is not possible to perform errorless data fusion due to the mismatched spatial and temporal sampling of the original datasets. However, some recent work has shown that if instruments were designed to use continuous scan methods, such aliasing errors could possibly be minimized (Lee 1994). Data fusion thus inherently introduces such errors, but as previously mentioned, the scientific benefits are large when it is performed properly and includes an appropriate acknowledgement of the introduced errors. This is discussed in more detail in section 2.

b. Previous methods

Remapping satellite data for comparison with additional datasets is not new. Several studies have compared satellite and radar data but have relied on very specialized software such as MUDRAS (Kohn et al. 1978), RADPAK (Heymsfield et al. 1983), and CEDRIC (Mohr et al. 1986). Entire systems have been developed devoted to the analysis of satellite data in conjunction with other datasets (Suomi et al. 1983; Hibbard and Santek 1989). The challenge is how to design a system that provides for flexibility and extensibility, as well as efficiency and robustness in the data fusion method. Earlier work in this area can be categorized in terms of the data access method (and related data structures) and the data fusion method that are used.

1) DATA STRUCTURES

The data access method and its related data structures play an important role in the overall functionality of the data fusion method. For example, much of the

heritage of meteorological data fusion methods originated from the meteorological radar community. In part, this is because satellite datasets were collected at relatively few sites and tend to be very specific regarding their archival formats and navigation procedures. It is common for one satellite dataset to have several supported archival formats for each institution where the dataset was collected and resides. Meanwhile, radar sites are relatively numerous (plus their simpler navigation system makes the problem of a standardized data format more tractable). This forced the data transfer and access issues much sooner, resulting in the development of the universal format (UF) standard for data transfer to other sites (Barnes 1980). In relation to development of data fusion methods, the UF helped to not only achieve data uniformity and transportability between different radar sites but also allowed the construction of complex radar analysis packages (Oye and Mueller 1986). Satellite data processing systems have tended toward more data-specific implementations due primarily to the specific nature of their datasets. Thus, in today's situation of more widely available remote sensing data through Internet access, multisensor CD-ROM datasets, and the growing multiplicity of data formats, there is a stronger need to extend satellite data processing methods to non-data-specific methods.

The idea of a standard self-describing format was then later extended to abstract data structures and arrays with one of the first being the common data format (CDF) (Treinish and Gough 1987; NSSDC 1994). This allows for an efficient, practical system in which the management and access of data is decoupled from the actual analysis and visualization software. Thus, long-term software costs are reduced because new datasets do not require redevelopment of the software that provides the data access to the analysis and display software; in turn, new analysis and display techniques do not require new data access software (Treinish 1989). This concept can be generalized further to remove dependencies of the analysis software to the visualization software. Such a twofold decoupling (i.e., data access/

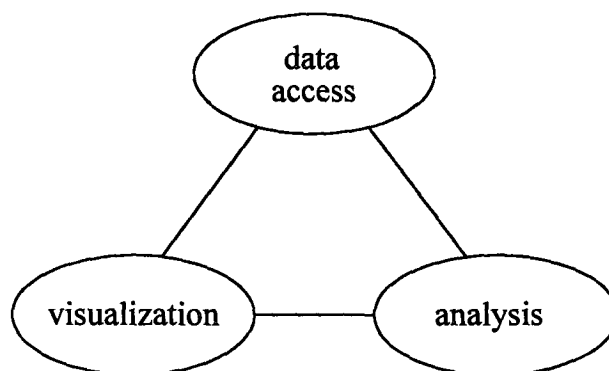


FIG. 1. Integration of visualization with data management and analysis (from Botts 1992).

analysis/visualization) as shown in Fig. 1, improves long-term flexibility and extensibility without undue complexity (Botts 1992). Further developments in the area of self-describing data structures for scientific datasets include named dimensions, machine-independent external data representation, C interfaces, and mechanisms for accessing larger aggregate data structures such as are in netCDF (Raymond 1988; Rew and Davis 1990; Unidata 1993). A hierarchical approach has also been used to provide several data access methods (interfaces) for data such as 8- and 24-bit raster images along with multidimensional scientific data (NCSA 1994). The hierarchical data format (HDF) developed at NCSA has been selected as the prototype standard data format for the EOSDIS project and includes a common library interface to both netCDF and HDF SDS file formats. Progress is relatively rapid in the design of self-describing data formats and many of the previously mentioned data formats continue to incorporate improvements (e.g., all of the mentioned self-describing data formats are now machine-independent).

2) REMAPPING METHODS

Data fusion methods employed to combine data of disparate types generally have relied most heavily on remapping methods. The basic procedure being first to select a working projection space into which all the datasets will be remapped (usually a raster image grid system), and then remapping successive images to this projection space for later analysis. The McIDAS, CEDRIC, and Candis systems are examples of such systems. In the case of McIDAS, data structures for grids, images, paths, and nonuniform data are defined, which allows an extensive set of remapping tools to manipulate the satellite datasets to the common grid space, but intimately ties the data access, analysis, and visualization processes together in a comprehensive but yet very complex system (Hibbard and Santek 1989). In such a system the basic remapping tools function in a similar fashion as the basic outline given above, in that the data fusion method rectifies images to a reference raster image working projection space. The CEDRIC system of Mohr et al. (1986) approaches the data fusion problem by interpolating data fields to a Cartesian grid. This simplifies the working projection space selection and is also a logical choice for radar-related work, since radar data are collected from a centralized stationary position in space. The Candis system (Raymond 1988) has a combination of capabilities. It can employ graphical merger techniques that function much like the enhancement method discussed earlier, in addition to being able to objectively analyze the datasets within a reference working projection space. In common with all these systems is the idea of remapping the satellite datasets into a reference working projection space and then having all subsequent analyses occur within this reference working projection space.

c. Organization

This paper introduces a new data fusion method that is implemented in a system that combines some of the previously mentioned concepts (i.e., it incorporates the advantages of a self-defining data structure with a generic comprehensive remapping and transformation software tool). This tool separates the data analysis from the data visualization in such a way as to obtain significant improvements in flexibility and extensibility in addition to an improvement in overall efficiency. The Polar Orbiter Remapping and Transformation Application Library (PORTAL) was originally developed to combine and remap satellite data from multiple sensors as well as other geophysical data (Jones and Vonder Haar 1992). Initially designed for Defense Meteorological Satellite Program (DMSP) SSM/I and Geostationary Operational Environmental Satellite (GOES) VISSR data, the method is general enough for use with other datasets without modification to the main PORTAL programs. The flexibility to use other datasets is obtained by defining a generalized data format that retains the full resolution and associated earth-location information of each dataset. The primary difference between PORTAL and previous software packages is a new data fusion paradigm in which the merged dataset products are in an original satellite projection space. This will be shown to have significant computational advantages for subsequent scientific analysis with the merged datasets. Section 2 introduces the concepts that are used within PORTAL and then section 3 contains a demonstration of the data fusion techniques applied to retrieval of effective microwave emittance from the SSM/I and VISSR datasets.

2. Concepts

There are two main concepts used in PORTAL that are simple but very powerful when used together in a cohesive fashion:

- 1) generalized data format structure,
- 2) a new data fusion method that generates merged datasets in an original satellite projection space.

It is the combination of both concepts that allows for flexible and routine multisensor data processing. Thus, both concepts are important, since each contributes to the effectiveness of the overall system design.

a. Generalized data format

It should also be noted that the following description of the generalized data format structure is of the particular implementation used in this data fusion software system. Many other variations are possible using other more widely accepted data formats (e.g., CDF, netCDF, HDF, and others) that are able to perform similar functions (NSSDC 1994; Unidata 1993; NCSA 1994). The discussion of the generalized data format is meant

to convey the structure of the overall system design and suggest data format characteristics that are desirable in a multisensor-multispectral data fusion software system.

Each sensor dataset (satellite or other rasterized datasets) is translated into a generalized data format (GDF) structure. All sensor datasets are therefore "identical" in format, which reduces the programming overhead required by special software design considerations for each unique satellite data format. The GDF acts as a buffer to the main software system, greatly reducing the complexity of the software system and is essential to the efficient processing of data from the numerous types of satellite-based instruments.

Aspects of the radar data UF (Barnes 1980) were used to develop and design the satellite data format structures that are used in the satellite GDF. The UF provides a complete description of each radar ray (data acquired for a given pointing direction) that is individually written to one or more physical records. The information includes all necessary information to reconstruct the radar ray in 4D space (i.e., site location, azimuth, elevation, range, and time) plus additional miscellaneous information. Each record contains several headers including a mandatory header that contains pointers to the start of subsequent headers. In an analogous manner, the GDF structures are based on satellite instrument scan lines (data acquired during individual sensor scan sweeps) and contains all the necessary information to reconstruct the instrument scan line in 4D space.

The selection of parameters to include in such a format specification is complicated by the fact that the satellite is in constant motion relative to the earth requiring some sort of navigation procedure for the satellite dataset. The usual practice with satellite data formats is to include the satellite ephemeris information to enable postarchival navigation of the satellite datasets by specialized software designed for that particular sensor. Another solution is to navigate the satellite information before archival and tag the navigated latitude and longitude information with the raw data values. Since it is unavoidable that the navigation of the satellite dataset will require very specialized software due the dependence on the very complex and unique combinations of sensor and satellite design configurations, the latter method was selected to simplify the data fusion software. By tagging each data element with earth-located latitude and longitude information, it is possible to exclude the sensor-specific navigation software from the data fusion software. However, the satellite ephemeris information is not discarded, but is stored in the GDF data structure for possible later use. In this way, the GDF data structure allows all possible navigation correction methods to be used [i.e., satellite ephemeris modification methods, digital image warping methods, etc. (Straka et al. 1993; Poe and Conway 1990; Wolberg 1990)]. Using this navigation method,

however, requires additional parameters, time, and satellite position, to sufficiently define the satellite scan line in 4D space given only the earth-located latitude and longitude information. Scan-line time information is usually very easy to obtain within most navigation software systems since timing information is critical to accurate earth location. The satellite position is defined by the earth-located satellite subpoint (position on the earth directly below the satellite at nadir) and satellite zenith angle. This satellite position representation was selected since it may prove useful in later scientific work and eliminates the need to recompute these values. Examples of when this may be beneficial include computation of atmospheric optical depth, which varies as a function of the satellite zenith angle (Liou 1980), and geometric cloud height determination techniques, which require solar azimuth and zenith angles in addition to the satellite subpoint and zenith angle (Purdum and Dills 1994). By specifying the scan line time, satellite subpoint, and the latitude, longitude, and satellite zenith angle for each data element within the scan line, the full 4D specification of the satellite scan line geometry is saved within the GDF data structure (see Fig. 2).

Aspects of the self-describing data formats mentioned previously were included in the GDF files. In accordance with the goals of a more simplified satellite data access method, the satellite sensor datasets were transformed from multiple files (one for each spectral channel) into one large multichannel satellite dataset contained in a GDF file. Random access calls can be used to retrieve GDF file elements of the satellite datasets, enabling easy updating of the GDF header information. Limitations are an a priori knowledge regarding the total number of spectral channels that are

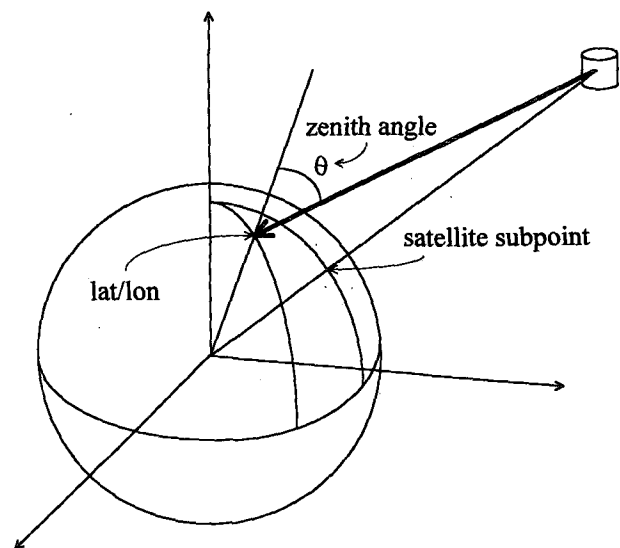


FIG. 2. Specification of 4D satellite scan line geometry.

to be contained in the GDF file, and scan line record length. Both limitations are minor, in that missing channel information is contained in the GDF header and satellite scan lines are usually uniform in length or are sectorized as such. The SSM/I sensor is an exception to this case and techniques used to handle this are discussed in the demonstration section (section 3).

The GDF makes use of a small subset of abstract data structures that contain the appropriate satellite data information required for multisensor-multispectral data fusion. They are also flexible and extensible so that numerous satellite sensor datasets may be used with the same data fusion programs without recompilation of the programs. The effect is a data-object-oriented approach to processing the satellite datasets in which redundant programming is significantly minimized, improving overall development-productivity. GDF datasets are thus transferred seamlessly from one stage of processing to another. Numerous satellite datasets have been stored in the GDF, including GOES VISSR, GOES VAS, DMSP SSM/I, DMSP OLS, RTNEPH data, NOAA AVHRR, HIRS, and MSU data. In addition, geophysical data and highly processed satellite data products have been stored in the same GDF data structures for easy use with the raw satellite datasets using the PORTAL software. Defining an extensible, flexible generalized data format also minimizes the amount of specialization and overhead required of application programs. Applications make use of a PORTAL library toolbox to handle many of the mundane chores of determining the appropriate channel number, scan line, and data-scale conversions, all in a consistent framework based on the GDF data structures.

Two classes of data structures are used to provide the necessary information.

1) HEADER STRUCTURES

The header structure contains information that is common to the entire data file, such as the following:

- data identifiers,
- sensor identifiers and operational information,
- data domain summary (time, latitude/longitude, and orbit ranges),
- internal information on how the data is stored,
- data flag information,
- sensor channel information,
- optional comments,
- history record containing information on how the file was generated.

The intention of the software was to provide the user with tools to *manage* the satellite dataset. The header structure allows for the capability of several useful features, such as, multiple data flags that can be used within the dataset to denote missing or unique conditions, the ability to access multiple sensor chan-

nels within the same data file, and additional text information relevant to the particular data file. Text comments may be inserted with the data for later reference. This may be useful, for example, if the data undergoes several stages of processing and requires more detailed identification. In addition, it is possible to include a history record that contains all of the input information from the generating program, thus eliminating ambiguities of how the file was generated.

2) SCAN-LINE STRUCTURES

The scan-line structure contains information that is only common to one particular scan line of the sensor, such as the following:

- scan-line number,
- scan-line time,
- scan-line status,
- satellite subpoint latitude-longitude information,
- scan-line data, (for each scan-line element),
 - (i) tagged latitude and longitude,
 - (ii) tagged satellite zenith angle,
 - (iii) multiple n -channels of data ($x_i, i = 1, n$).

Each scan-line element is tagged with its respective latitude and longitude earth-location values. This eliminates the burden of creating and maintaining a generalized satellite navigation package, and enables users to take advantage of navigation software provided with their particular dataset. Additional information stored in the scan-line structure, such as the scan-line time, satellite subpoint, and satellite zenith angle, allow for the capability of determining the position of the sensor beam within a 4D space. This would be useful when there are important temporal or directional effects. The scan-line time and satellite subpoint information is stored only once per scan line to save space since these values do not change significantly during an individual sensor scan-line sweep (time between scan-line sweeps is typically under 30 s, and often under 2 s). The earth-location and satellite zenith angle information, however, may vary considerably within one scan line and are stored for each scan-line element. If appropriate, the data are calibrated for each spectral channel and data values are stored as radiances, brightness temperatures, or other appropriate physical quantities. Additionally, the relevant calibration information is appended to the GDF header data structure for later reference. Many spectral channels (currently up to 512) can be stored and retrieved within one data file. For multichannel application algorithms, accessing one file (instead of several) simplifies the processing of the data.

b. Data fusion method

The PORTAL data fusion method introduces a new remapping paradigm that differs from the traditional remapping methods that we have discussed. To per-

form the high-resolution merger of disparate datasets, traditional remapping methods define a reference working projection space into which all data are "mapped" and in which all subsequent scientific analyses occur. For example, suppose we have two disparate datasets, A and B, which are from sensors on different satellite platforms, or sensors that may have radically different earth scan patterns and ground resolutions. In the traditional remapping method, both datasets are remapped to a reference working projection space (see Fig. 3a). This reference working projection space could be defined as any user-defined geographic coordinate system, or even a regular grid from a selected satellite projection space. Two remapping operations are performed using the traditional method, one on each original dataset. In comparison, the PORTAL remapping method also includes a reference working projection space, but only one dataset undergoes any remapping operations. As shown in Fig. 3b, dataset A is remapped to the reference working projection space, but in contrast to the traditional method, dataset B is *not* remapped to the reference working projection space, and the remapped data from dataset A is remapped to the original satellite projection of dataset B. Thus, in regard to dataset B, the remapping direction is reversed compared to the traditional method. When one compares the number of total remapping operations in both methods, Fig. 3 shows that both methods have a total of two remapping operations. Thus, no extra remapping operations are involved. Two alternate perspectives are possible in classifying the PORTAL data fusion remapping method. One is that dataset A undergoes a two-step remapping procedure and dataset B remains unchanged, or that dataset A underwent one remapping operation, and that the remapping operation of dataset B has been reversed in direction. Either view is appropriate.

The significance of the new remapping paradigm is that of subsequent computational efficiency. Since two remapping operations are performed in each method, there is no fundamental difference in the computational speed during the remapping process. However, there can be significant computational differences in subsequent operations on the merged datasets. This characteristic is related to the differing resolutions of the original datasets and in the resulting size of the final merged datasets.

To examine the computational efficiency, consider the two approaches that are possible using traditional remapping methods. The first can be characterized as an overbinning technique that subsamples the highest resolution data and reduces the resolution of the final merged datasets. The second is more of a "brute force" hyperfine grid technique in which the remapping is performed using a working projection space that is of higher resolution than that of the highest resolution dataset that is to be merged. For example, if datasets A and B are to be merged on a hyperfine grid (see Fig.

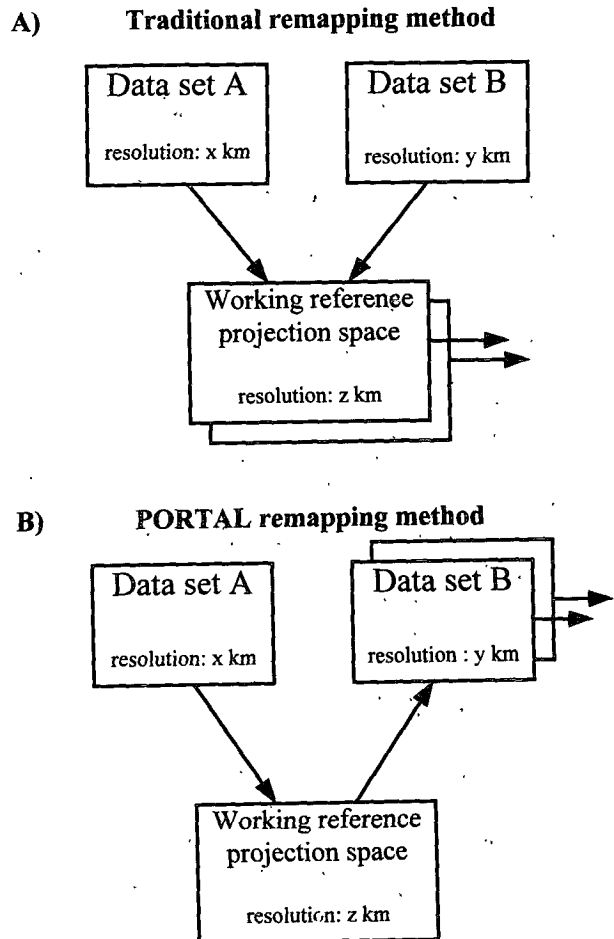


FIG. 3. A comparison between (a) traditional remapping methods and (b) the PORTAL remapping method for multisensor-multispectral data fusion.

4) the areal coverage of a data element in projection space B (the shaded area in Fig. 4) is actually a composite of several data elements within projection space A. Spatial sampling is performed on the hyperfine grid for each projection space, A and B. Thus, when the resolution of the hyperfine grid is increased, the spatial characteristics are more accurately resolved for the data elements within datasets A and B, and the remapping accuracy improves. However, as a resulting side effect, the hyperfine grid technique generates very large and often spatially redundant merged datasets [e.g., each hyperfine grid element within the areal coverage of a data element in projection space B (the shaded area of Fig. 4) contains the same information when the hyperfine grid elements are also from the same data element within projection space A (areas in Fig. 4 that have the same intensity-shaded areas)]. The major difference between the hyperfine grid and overbinning techniques is that of the size of the reference working projection space and how sampling (if any) is per-

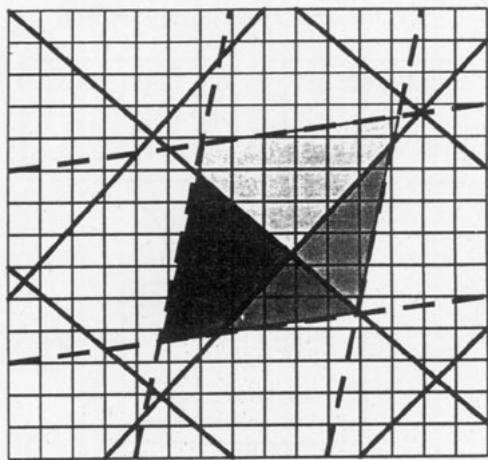


FIG. 4. Illustration of the hyperfine grid remapping technique. The heavy solid and dashed lines denote the projection space of datasets A and B, respectively. The light solid lines represent the hyperfine grid structure. The shaded area is the areal coverage of an individual data element in projection space B. The various shading intensities within this element represent the differing data element values within projection space A.

formed. The overbinning technique can be thought of as a particular implementation of the hyperfine grid technique in which the hyperfine grid is selected to be of lower resolution than one or both original datasets. The error characteristics of the overbinning technique in relation to the hyperfine grid technique is discussed further in the error propagation subsection and is shown to produce lower quality results when compared to the hyperfine grid technique. Of the two methods, the hyperfine grid technique is preferred for obtaining full resolution merged datasets, since the resolution of the original datasets is maintained. The major disadvantage of the hyperfine grid technique is that the very large hyperfine grid structure is then used in subsequent algorithms that operate on the merged dataset and thus propagates the computational penalties of maintaining the large grid into all subsequent processing.

The PORTAL remapping method also implements a large hyperfine grid within the remapping process to maintain data resolution; however, the results of the PORTAL remapping method are in an original satellite dataset projection (projection space of dataset B, in Fig. 3b), which can be *significantly smaller* than the hyperfine grid used as the reference working projection space. Thus, subsequent computational overkill is minimized using this method, since the resolution of the merged datasets are tailored to the original dataset resolution rather than the resolution of the large hyperfine grid (which is a computational artifact). Consider the example shown in Fig. 3. If the resolution of datasets A and B are x and y km, respectively, and for simplicity also assume that the area covered by a particular data element can be represented as x^2 and y^2 km², then the computational differences between

PORTAL and traditional remapping methods can be related to the ratio of the resolution of the low-resolution dataset (dataset B, in this example) to the resolution of the hyperfine grid, z , used in the remapping procedure. The computational expense per unit area for the traditional method can be written as

$$c_{\text{trad}} = \left(\frac{1}{x^2} + \frac{1}{y^2} \right) R + \frac{1}{z^2} S, \quad (1)$$

where R and S are the computational expenses per data element of the remapping method and of subsequent scientific algorithms, respectively. Similarly, the equivalent PORTAL computational expense per unit area is

$$c_{\text{PORTAL}} = \left(\frac{1}{x^2} + \frac{1}{y^2} \right) R + \frac{1}{y^2} S. \quad (2)$$

Both methods share the same computational expense for the remapping operations, since the remapping difference between the methods is limited to the reversing of the remapping direction on the low-resolution dataset (dataset B) (see Fig. 3). The overall-system improvement of the PORTAL method over the traditional remapping method is

$$\begin{aligned} c_{\text{ratio}} &= \frac{c_{\text{trad}}}{c_{\text{PORTAL}}} \\ &= \frac{[(y^2/x^2) + 1](R/S) + (y^2/z^2)}{[(y^2/x^2) + 1](R/S) + 1}. \end{aligned} \quad (3)$$

Since by definition the hyperfine grid is of higher resolution than the low-resolution dataset (i.e., $y > z$, $y/z > 1$, and thus $c_{\text{ratio}} > 1$), the PORTAL remapping method is *always* computationally more efficient than equivalent traditional remapping methods when the data processing system is considered as a whole. When the scientific algorithm is trivial, $S \rightarrow 0$, the methods are computationally equivalent in expense [see Eqs. (1) and (2)]. However, as the scientific algorithms become more complex relative to the remapping process (R/S becomes small), the computational performance differences increase and approach a limiting value of

$$\lim_{R/S \rightarrow 0} c_{\text{ratio}} = \left(\frac{y}{z} \right)^2, \quad (4)$$

which is dependent on the resolution ratio of the low-resolution projection space to the reference working projection space. Thus, for nontrivial scientific algorithms (the case where R/S becomes small), the PORTAL remapping method becomes more computationally advantageous as the reference working projection space resolution is increased to more accurately depict the full resolution of the satellite datasets. The accuracy and error characteristics of the PORTAL data fusion method are examined in more detail later in the error propagation subsection.

c. Software tools

The PORTAL data fusion method is the focal point of a group of software tools for manipulation of the satellite datasets by scientific application programs. In addition to the data fusion capabilities, an advantage of the system is that the scientific application programs have available all satellite datasets due to the previous generalization of the input data format. Scientific applications are no longer specific to the satellite datasets used in the original development (excluding non-object-oriented development contained within the scientific application codes—such as “hardwiring” instrument specifics such as sensor frequencies, field of view, etc.—that may or may not be application dependent). By providing a conceptually consistent framework defined with data-object tools using a self-describing data format, the future generality of these scientific applications is enhanced.

Several tools for data manipulation make up the PORTAL software library. The data fusion program (GDF2GDF) is at the heart of the system. Its basic function is to rectify one sensor satellite dataset to another reference satellite dataset using the PORTAL data fusion method.

Five software subsystem groupings are used within the PORTAL software package:

- 1) *format translators* that convert the original “raw” satellite datasets to GDF files,
- 2) *data fusion* program that merges multisensor-multispectral GDF files using the PORTAL remapping method discussed above,
- 3) *scientific applications* that use the GDF files as input and output,
- 4) *visualization* routines for routine displaying of the GDF file information in one of several methods,
- 5) *utilities* used for viewing the GDF file data structures for debugging, generation of statistics, and cursory quick-look information.

In the fivefold classification of the software groups, it should be noted that the visualization step is completely separated from the data fusion method. It is possible (and occurs often from our practical experience with the method) to merge the disparate datasets without producing image raster files or other related visualization products. Thus, processing of the multisensor-multispectral data is made more routine.

The focus of the overall software system is the data objects contained in the GDF files. Once the format translators operate without loss (i.e., without modification of the data values or associated data attributes) on the original raw satellite datasets and produce equivalent datasets in the GDF, the data flow is very easy between software subsystems. The multisensor data processing flow is functional in that original data are translated into the GDF file data structures, merged with one data fusion program, and then operated on

subsequently by the scientific applications. In addition to the main data processing flow, additional programs assist in visualization of the GDF files at any stage of analysis. Special utility programs help find processing errors or easily examine features of interest without going to the effort to write fully developed application programs.

Scientific applications are developed as a stand-alone subsystem in that the applications are the only software that is expected to change regularly (i.e., normal scientific algorithm development). The applications are based on template programs for ease of modification, and can open multiple GDF data files of disparate data types and combine the information in new ways. Output files are also GDF files, which enables multistage analysis of the results. Visualization of the datasets is removed from the analysis process.

Visualization of GDF files is done primarily through the generation of raster images for input into a locally developed image display system. The information within the GDF files can be remapped into any of the possible 11 types of predefined projection spaces available, including a raw projection space that is merely an identity operator mapping. This allows the researcher to select the most desirable projection for visualization purposes, not just the most convenient. Numerous standard image processing techniques are then available through the image display system for further analysis. Graphic map overlays can also be generated. Figure 5 is an example of how the visualization tools can be used to show the typical morning coverage from the DMSP F-8 and F-10 SSM/I sensors

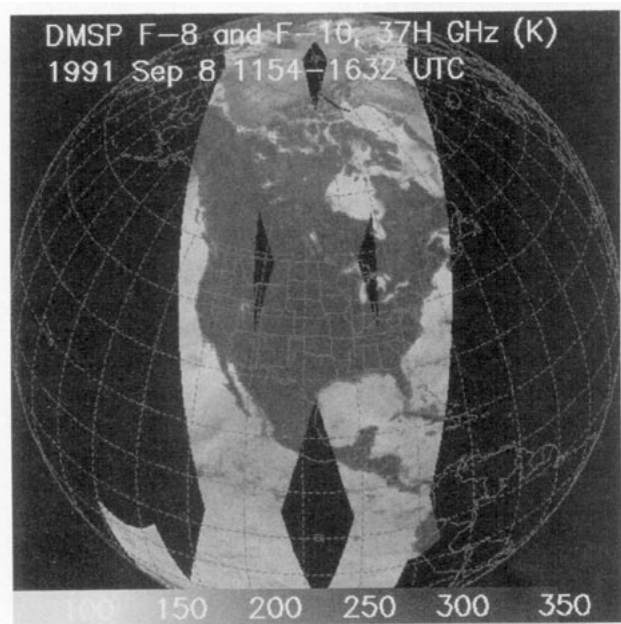


FIG. 5. Combined DMSP F-8 and F-10 SSM/I 37H-GHz imagery remapped to a satellite-view projection (from geostationary altitude) centered on 40°N, -97.5°E, 1154–1632 UTC 8 September 1991.

(37H-GHz channel) over the continental United States remapped to a specified satellite projection.

Several generic utilities are also provided to assist in some common data handling requirements.

GDF DUMP: A utility that reads any GDF file and translates the internal GDF data structures (including the data values) into human-readable form. This random access process is quite efficient. It functions much like a text editor that is unable to modify its contents. Additional capabilities also exist to optionally suppress the human-readable output and provide statistics on portions or on the entire dataset. This is especially useful for error checking and determining data value ranges. Comments, identification, and other miscellaneous metadata can be viewed with this utility.

GDFSECTOR: A utility that sectorizes an original dataset by earth-location ranges, scan-line number ranges, multispectral channel numbers, time ranges, etc.

GDFCOMP: A utility that generates composite GDF data files. This is extremely useful for generating climatological datasets from information obtained from a group of individual orbit passes or for averaging multisensor-multispectral data fusion results for more than one time period. The GDFCOMP routine also generates statistical information such as means, standard deviations, and extrema that are retained in an extended GDF file with any necessary additional metadata added. History records within the GDF header structure are updated to include the entire processing record of the final composite GDF data file.

d. Error propagation

Tracing of error sources becomes more important in the processing of the remote sensing data as measurements from remote sensing are increasingly utilized as a data source for numerical models, GISs, and other quantitative processing, analysis, conversion, and final product generation (Lunetta et al. 1991). This is especially true of merged large multisensor datasets where validation becomes particularly challenging (Kahn et al. 1991).

Error sources include the navigation used in the inverse image referencing method, which must be accurate or results will be degraded (Emery et al. 1989). For meteorological satellite datasets, most navigation errors are relatively small compared with the sensor field of view or can be corrected using methods that are currently available (e.g., Poe and Conway 1990). This may require some additional preprocessing of the data to reduce the navigation errors before attempting multisensor data fusion processing.

The data transformation processing flow, including sampling and regridding, may also introduce noise, in addition to the possibility of introduction of more fundamental changes to the input data affecting the quality of the final merged product (Davis et al. 1991). In the process of merging the multisensor-multispectral da-

tasets with PORTAL there is a significant simplification of the processing flow, thus reducing the complexities involved in tracing the error propagation path. This is an advantage if such a system is to confront the error propagation issues. It should be noted, however, that even with a simplified error propagation path, the error analysis of such a system is still a very challenging one.

Interpolation of the satellite imagery is an important factor to the overall accuracy of any data fusion system. The method used in PORTAL is a simple bilinear interpolation method, but could be extended with other methods that optimally interpolate sensor data based on, for instance, antenna gain patterns (Poe 1990). Other possible needs include the ability to compute the mean and variance of area estimates for understanding and modeling impacts of spatial errors, which is also important for GISs (Prisley et al. 1989).

An attempt was made to estimate error characteristics of the data fusion processing system regarding the resolution of the reference working projection space. Within the PORTAL data fusion method, the reference working projection space can be thought of as an "internal" grid, since it is not used in the external representation of the merged datasets. For a particular area, the resolution of the reference working projection space is related to the internal grid domain size. Data were processed for a typical GOES VISSR (infrared) image in which the data were merged to the lower-resolution data of a specific SSM/I orbit pass. The standard deviation of the GOES infrared data, σ , rectified to their corresponding grid cell within the SSM/I projection space was computed for each SSM/I field of view. An average GOES standard deviation value was calculated for internal grid domain sizes ranging from 32×32 to 2048×2048 (see Fig.

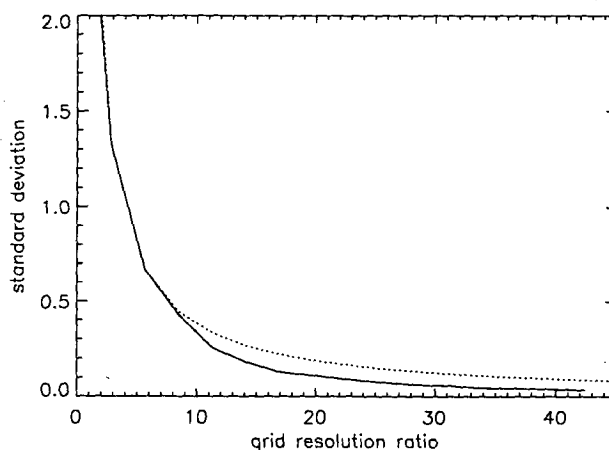


FIG. 6. GOES data statistics versus grid resolution ratio. Solid line: average standard deviation (K) of the GOES infrared data rectified to their corresponding grid cell within the SSM/I projection space versus the grid resolution ratio. Dashed line: standard deviation (K) versus the grid resolution ratio derived from theory using the $\sigma n^{-1/2}$ relationship.

6). The grid resolution ratio was defined as the ratio of the square root of infrared data pixel area ($4 \text{ km} \times 8 \text{ km}$) with respect to the square root of the internal grid cell area (the grid resolution would be the quantity x/z in the equivalent terminology of Fig. 3). Thus, grid resolution ratio values greater than 1.0 denote an internal grid that is more fine than the original dataset, and grid resolution ratios less than 1.0 would have an internal grid that would be less fine than the original dataset and would be the functional equivalent of an overbinning method (i.e., methods whose internal grid resolutions are lower than the original sensor data resolution). The values that are generated by the rectification process asymptotically approach the value of the highest grid resolution results, as expected. The theoretical $\sigma n^{-1/2}$ dependence (Beers 1957), however, is only approximate as shown in Fig. 6. Factors contributing to this behavior are replication of infrared data pixels to grid cells within the SSM/I projection space during the rectification process, the nonnormal data distribution of the GOES infrared data, and local spatial correlations due to sensor time integration of the measurements (Hillger et al. 1993). By comparing the average standard deviation due to the data fusion analysis procedure to the expected standard deviation of the signal that is to be observed, results such as in Fig. 6 provide a measure of the highest internal grid resolution requirement to accurately remap the datasets at resolutions comparable to the original sensor resolution. It should be noted that, in general, this error contribution is ignored in most data processing systems that use traditional remapping methods. The GOES data values that traditional remapping methods would provide within the SSM/I grid cells come from a fixed-resolution working reference projection space with the inherent assumption that the resolution of the working reference projection is high enough to sufficiently minimize resolution-related remapping errors. This analysis shows that a flexible internal gridding method can be used to help determine the maximum grid resolution required before significant data statistical properties are lost in the data transformation processing flow. The amount of statistical information lost would of course vary with sensor noise characteristics and on the specific data application.

3. Example

The multisensor-multispectral data fusion method is demonstrated using data from the GOES VISSR and DMSP SSM/I sensors to retrieve effective microwave surface emittance for clear sky conditions.

a. Satellite sensor descriptions

The GOES VISSR sensor is a multichannel visible and infrared instrument in geostationary orbit. The instrument is actually several detectors combined and

has the capability of operating in several modes (Clark 1983). When operating in the dwell sounding mode to produce temperature retrievals, the instrument is known as the VISSR Atmospheric Sounder (VAS) and can transmit up to 12 channels in the infrared. In our example, the instrument was operating in the VISSR multispectral imaging (MSI) mode in which selected channels of infrared imagery at (in this case) 12.7, 11.2, 6.7 μm (channels 7, 8, 10) with $4 \text{ km} \times 16 \text{ km}$, $4 \text{ km} \times 8 \text{ km}$, and $4 \text{ km} \times 16 \text{ km}$ nominal resolution, respectively, and visible imagery with $1 \text{ km} \times 1 \text{ km}$ nominal resolution are transmitted each half hour for the whole hemisphere.

The SSM/I on the DMSP F-8 and F-10 satellites is in a sun-synchronous near-polar orbit at an altitude of 833 km with a period of 102 min. The orbit has an inclination of 98.8° , which allows for overlapping coverage poleward of 50° latitude twice daily (Hollinger et al. 1987). The SSM/I has four frequencies (19.35, 22.235, 37.0, and 85.5 GHz) and dual polarization capabilities on all except the 22.235-GHz frequency, which records only the vertical polarization. The data are collected during the rearward 102° portion of the instrument rotation, which results in a conical scanning pattern 1394 km wide. The effective field of view for each channel is listed in Table 1. The data are sampled at 25-km intervals with 64 data elements per sensor scan line, except for the 85.5-GHz channels that are sampled at 12.5-km intervals and have 128 elements per sensor scan line (see Table 1). Geolocation errors for the SSM/I are approximately 7 km for data archived after July 1989 (Poe and Conway 1990).

b. Surface emittance retrieval method

The surface emittance retrieval method is similar to the method used in Jones and Vonder Haar (1990), but does not account for atmospheric attenuation in the infrared or microwave frequencies. It can be thought of as an effective surface emittance defined as

$$\epsilon_{\text{sfc}} = \frac{T_{B_{\text{MW}}}}{\bar{T}_{B_{\text{IR}}}} \quad (5)$$

(Ulaby et al. 1986), where $T_{B_{\text{MW}}}$ and $\bar{T}_{B_{\text{IR}}}$ are the mi-

TABLE 1. SSM/I instrument characteristics.

Channel number	Frequency (GHz)	Polarization (H or V)	Effective field of view* (km)	Number of scan-line elements
1	19.35	V	69×43	64
2	19.35	H	69×43	64
3	22.235	V	60×40	64
4	37.0	V	37×28	64
5	37.0	H	37×29	64
6	85.5	V	15×13	128
7	85.5	H	15×13	128

Adapted from Hollinger et al. (1987).

* Using a 3-dB limit.

crowave and mean infrared brightness temperatures. The mean infrared brightness temperature is given by

$$\bar{T}_{B_{IR}} = B_{IR}^{-1} \left[\frac{\sum_{ij} w_{ij} B_{IR}(T_{B_{IR,ij}})}{\sum_{ij} w_{ij}} \right], \quad (6)$$

where B_{IR} and B_{IR}^{-1} are the Planck function and its inverse, respectively, and w_{ij} is the sensor spatial weighting function to account for the SSM/I antenna pattern overlap defined for the relative SSM/I element and scan-line number, i and j (Jones and Vonder Haar 1990). The emittance values are “effective” emittances since no atmospheric corrections are applied to the radiances; thus, the values are approximate and are not the true emittance values (Comiso 1983). However, spatial inhomogeneities are rather well represented in these results since nearest-neighbor atmospheric corrections are correlated at these scales. Notice that the PORTAL data fusion method allows the antenna pattern correction across the SSM/I field of view to be trivially applied, other remapping methods that do not maintain the relative sensor field-of-view positions would have made the antenna pattern correction more difficult to implement.

c. Cloud clearing

The surface emittance is retrieved only for clear-sky regions since mid- and high-level clouds can bias the surface emittance results due to their effects on the microwave and infrared brightness temperatures (Grody 1983). An infrared threshold method is used to determine the clear-sky areas with a manually determined threshold value of 280 K used in the examples presented.

d. Procedure and results

The VISSR and SSM/I datasets are converted to GDF file representations for input into the data fusion program (GDF2GDF). The format translator programs are specific to each satellite dataset and include navigation software to earth-locate the datasets as well as sensor specific calibration information. The multiple resolution sampling of the SSM/I sensor is accounted for in this process by storing the sensor sample pattern information in a table within the GDF header data structure. Each channel thus has individual control of its sampling behavior and in this respect is independent of the other sensor channels. The sampling pattern is also independently controlled with regard to the scan-line sampling as well as with respect to the scan elements. Some inefficiencies exist however, since the missing elements that are not within the sampling pattern are still recorded and stored with missing data values. This was necessary to simplify the data I/O routines that read the datasets and provide the desired amount of software flexibility.

The entire sequence of data processing steps after translating the datasets into GDF files is shown in Figs. 7–10. Raster images presented in the figures are output from the visualization tools available within PORTAL. Figures 7–10 are *not* direct representations of the GDF file contents that are in their respective satellite sensor projection space. In a sense, we are interrupting the data fusion process at every possible stage and “viewing” the GDF file contents on a common projection space through the visualization software tools for comparison purposes. This is not a requirement of the data fusion process. The display and visualization of the GDF file contents are entirely independent of the data fusion analysis software. In fact, it is quite reasonable to assume that once the satellite datasets have been translated into GDF files that the next visualization step would occur during the display of the final results (Fig. 10). In addition, in an effort to simplify the demonstration example, only a single VISSR (11 μm) and SSM/I (37H GHz) channel combination are shown, when in reality all channels contained in the GDF files are coregistered after processing through the data fusion program.

The raw VISSR infrared data (Fig. 7) must be coregistered into the projection space of the SSM/I 37H-GHz channel (Fig. 8). The VISSR sensor was selected to be remapped into the SSM/I sensor projection space since the VISSR is a higher-resolution sensor and the mean infrared brightness temperature can be calculated relatively easily using Eq. (6). If the reverse were to occur (i.e., the SSM/I data remapped into the VISSR sensor projection space), the interpretation of the results would be muddled since the SSM/I field of views

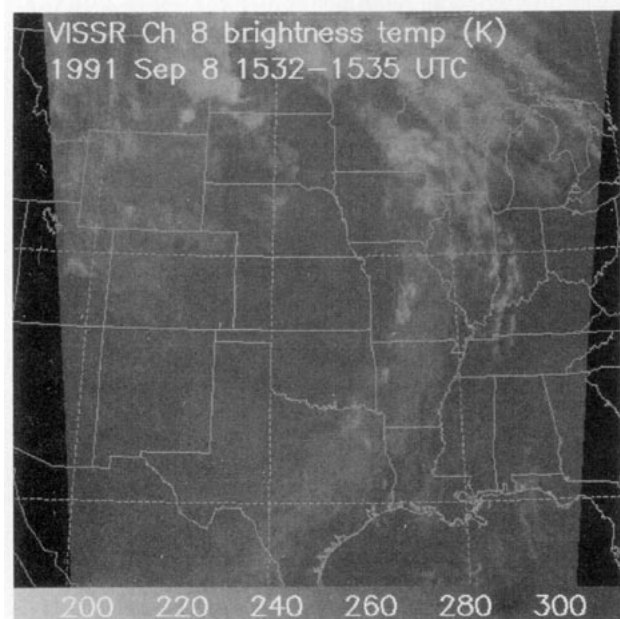


FIG. 7. VISSR channel 8 (11 μm) brightness temperature, 1532–1535 UTC 8 September 1991.

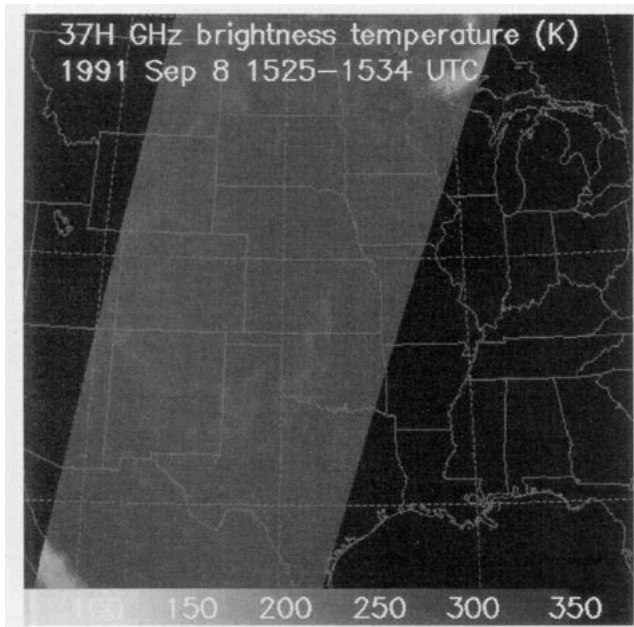


FIG. 8. SSM/I 37H-GHz brightness temperature, 1525-1534 UTC 8 September 1991.

(FOVs) are larger than the VISSR FOVs, and scientific questions would arise due to the effect of the SSM/I antenna gain functions on the basic assumptions of the effective surface emittance retrieval method. Such an approach of remapping low-resolution sensor data into higher-resolution data projection spaces is thus dubious, unless care is taken to specifically correct and interpret the results (Lee 1994; Poe 1990).

The respective GDF data files containing the translated “raw” VISSR and SSM/I datasets are merged with the data fusion program (GDF2GDF). The SSM/I earth-location information is used as a reference projection space to control the remapping of the VISSR satellite dataset. This is facilitated by an internal grid system in which the one-to-one mapping of the VISSR to SSM/I projection space is accomplished. For the example shown, the internal reference working projection space is a Lambert equal-area projection over the central United States. This was one of 11 working projection spaces defined within the PORTAL system (see Table 2). Examples presented in Figs. 7-10 are also displayed using the same projection space specification. Additional projection spaces could be easily added to PORTAL by including the projection’s mathematical definition and associated domain and range limits to the projection space driver subroutine.

A 512 × 512 grid of moderate resolution was used to merge the VISSR satellite data into the SSM/I projection space (Fig. 9). Thus, an equivalent traditional remapping method would also employ a reference working grid of 512 × 512 grid points; however, in the traditional remapping method case, any future processing would occur on both datasets within this 512

TABLE 2. Predefined PORTAL projections.

Projection
Lambert conformal
Azimuthal equidistant
Gnomonic
Lambert equal-area
Orthographic
Stereographic
Satellite view
Cylindrical equidistant
Mercator
Mollweide-type
Raw data projection

× 512 grid. The data fusion method used within PORTAL merges the VISSR satellite dataset into the SSM/I projection space. This achieves considerable savings regarding data storage and facilitates computation of geophysical parameters with the merged VISSR and SSM/I satellite datasets, since calculations are performed only for each SSM/I FOV instead of remapping the data into a large grid with many points, which would require many redundant calculations. For this example, there are only 64 × 137 (8768) SSM/I FOVs, instead of 512 × 512 (262 144) in the equivalent traditional remapping method. Therefore, subsequent scientific algorithms using the PORTAL merged datasets have nearly 30 times fewer data elements to process compared to the traditional remapping method, in addition to the already mentioned benefits from the

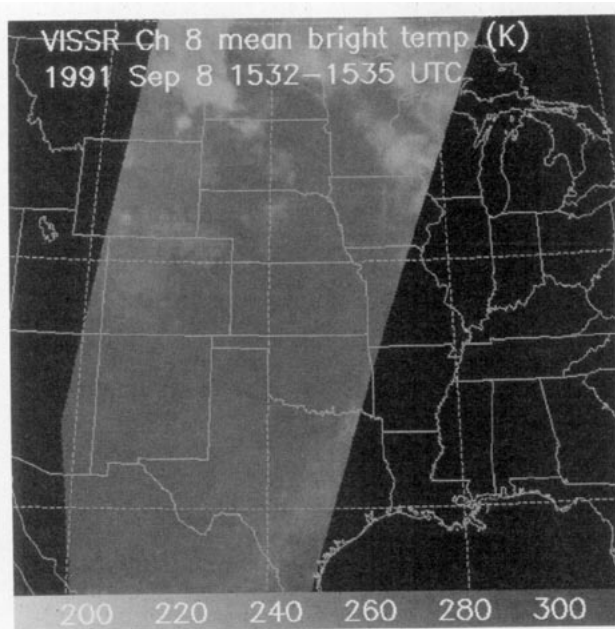


FIG. 9. VISSR channel 8 (11 μm) mean brightness temperature, 1532-1535 UTC 8 September 1991 coregistered to match the DMSP SSM/I projection.

flexible and extensible design concepts used within the PORTAL system. The results of the scientific algorithm, the effective microwave surface emittance for clear-sky areas based on Eqs. (5) and (6), are produced in the SSM/I sensor projection space (Fig. 10).

e. Composite results

Data fusion of multisensor–multispectral satellite datasets does not stop with a single combination of individual satellite orbit passes. Methods are required to generate composite results to investigate phenomenon that might not be obvious with only one pairing of data. By using the program GDFCOMP, a series of merged datasets can be composited over a longer time period. This was done with data from DMSP F-8 and F-10 SSM/I and GOES-7 VISSR that were processed to determine the spatial and temporal variability of the effective surface emittance during the period 30 July–7 October 1991 (~70 days) over the central United States (Jones et al. 1993). The mean 37H-GHz surface emittance composite for the total time period shows strong surface effects from lakes, rivers, and desert regions as expected (Neale et al. 1990) (Fig. 11), but several low surface emittance features are also present in the Texas panhandle and the Oklahoma regions, which are unrelated to natural water bodies and desert effects.

Relevant statistics are also generated for the composite time period and stored in an extended GDF file, which allows further flexibility with the data analysis since the GDF file is not an “end result” output file, but can also be used as input into the GDFCOMP

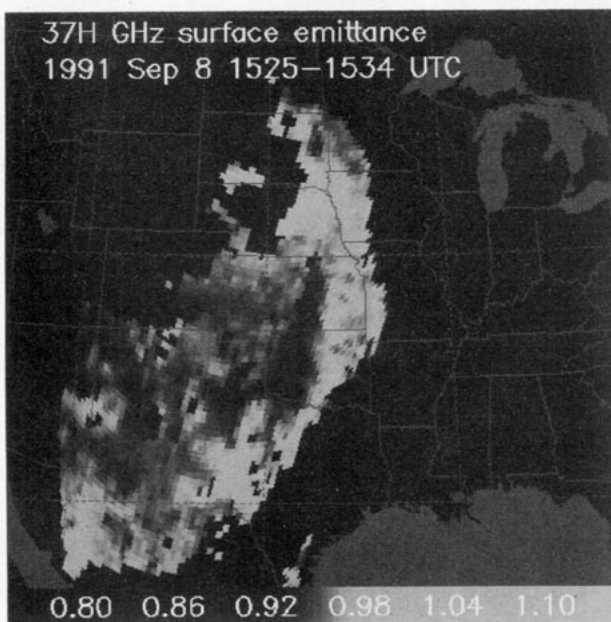


FIG. 10. The 37H-GHz surface emittance, 1525–1534 UTC, 8 September 1991.

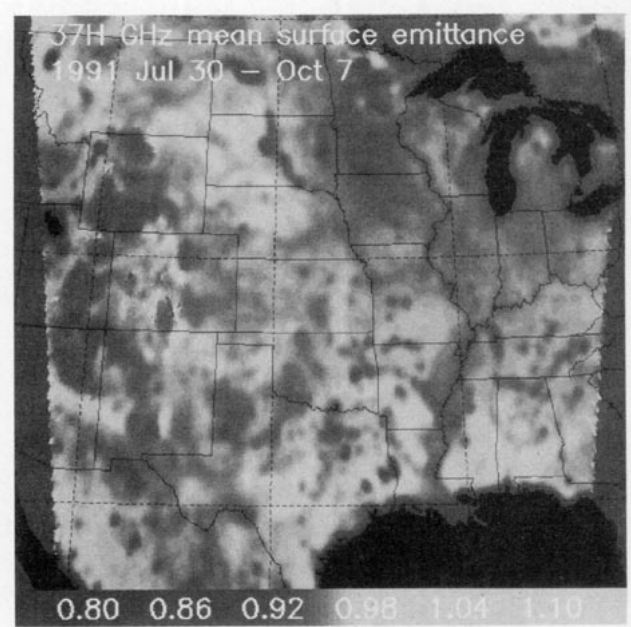


FIG. 11. The 37H-GHz mean surface emittance, 30 July–7 October 1991.

program, thus modularizing the data processing flow, similar to the modular concepts developed within Candis (Raymond 1988). The GDF files are acting in this sense as data objects, continuously flowing between data analysis programs. Such a data-object-based method would have significant benefits toward operational use of merged multisensor–multispectral satellite datasets. Data objects would be available for immediate use while at the same time providing continuity of the analysis with subsequent new merged datasets by a cyclical reintroduction of the GDF file as input into the original analysis program.

4. Summary and conclusions

We have introduced a new data fusion method that generates merged datasets in an original satellite projection space. Computational efficiencies are compared between this data fusion method and that of traditional remapping methods. While comparable processing times are needed to physically merge the datasets, results show significant performance gains on any subsequent analysis of the merged datasets since scientific algorithms operate within the original satellite projection space.

A self-describing generalized data format (GDF) is also discussed that reduces the complexity of the data fusion software system and modularizes the data processing flow to obtain significant improvements in terms of flexibility, extensibility, and application generality. The GDF described within this paper is a specific implementation of a data format suitable for efficient processing of large volumes of satellite data in

a multisensor-multispectral data fusion software system. The discussion of the GDF is meant to convey the structure of the overall system design and suggest data format characteristics that are desirable in such a system. More widely accepted generalized data formats such as CDF, netCDF, or HDF could provide similar functionality and will be considered in future work as efforts are made to standardize the implementation.

The two main concepts used within PORTAL, the new data fusion method and the generalized data format, allow for routine multisensor-multispectral data fusion processing of satellite datasets in an efficient and productive manner. While the specific focus of this paper has been on the effective use of multisensor-multispectral satellite datasets, the methods employed are general enough for *any* raster-based (i.e., scan-line-based) datasets [such as topography, radar (ground and satellite based), raster-based GIS information, and many other possible datasets]. The method presented has much to offer in terms of making multisensor-multispectral data fusion a tool-of-the-trade rather than leaving it in the realm of special intensive case studies. It has been enthusiastically used by many of our colleagues. The ability to vary the internal grid resolution ratio has implications toward tracing error propagation within the data fusion processing system and provides an opportunity for more understanding of the inherent errors in such a complex analysis system. The method is also able to provide statistical information on the data fusion processing method, which may impact results and their subsequent interpretation. Further work quantifying the error behavior within the data fusion system would also be of great use in subsequent interpretation of results from such a multisensor-multispectral data fusion processing system.

Future work (which previously would have not been possible without the ability to routinely merge multisensor-multispectral datasets) will consist of longer-term investigations into the effects of microwave surface emittance as derived with the multisensor-multispectral data fusion method, and in using the data fusion method as a means for importing remote sensing data into a research-quality mesoscale atmospheric model. Efforts are also planned to improve the data storage efficiency (e.g., using the method of Gilbert et al. 1992) and to extend the data fusion error analysis work with other satellite datasets including synthetic datasets with well-defined statistical properties.

Acknowledgments. We would like to thank John Forsythe, Don Hillger, and many others at CIRA for the fruitful discussions during the design and implementation of PORTAL. We would also like to thank the anonymous reviewers for their comments and many helpful suggestions. This work was supported by NOAA through Grant NA90RAH00077, and by the U.S. Army Center for Geosciences under ARO Grant

DAAL03-86-G-0190 to the Cooperative Institute for Research in the Atmosphere (CIRA) at CSU.

REFERENCES

- Barnes, S. L., 1980: SESAME news. *Bull. Amer. Meteor. Soc.*, **61**, 1401–1404.
- Beers, Y., 1957: *Introduction to the Theory of Error*. Addison-Wesley Publishing Co., 66 pp.
- Bernstein, R., C. Colby, S. W. Murphey, and J. P. Snyder, 1983: Image geometry and rectification. Vol. 1, *Manual of Remote Sensing*, 2d ed. R. N. Colwell, Ed., American Society of Photogrammetry, 873–922.
- Botts, M. E., 1992: The state of scientific visualization with regard to the NASA EOS Mission to Planet Earth. *Applied Information Systems Research Program (AISRP) Workshop II Meeting Proc.*, Boulder, CO, Laboratory for Atmospheric and Space Physics, University of Colorado, E132–E139. [Available from Information Systems Branch, Flight Systems Division, Office of Space Science and Applications, NASA Headquarters.]
- Campbell, W. J., and R. F. Crompt, 1990: Evolution of an intelligent information fusion system. *Photogram. Eng. Remote Sens.*, **56**, 867–870.
- , N. M. Short Jr., and L. A. Treinish, 1989: Adding intelligence to scientific data management. *Comput. Phys.*, **3**, 26–32.
- Clark, J. D., 1983: *GOES User's Guide*. U.S. Department of Commerce, Washington, D.C.
- Comiso, J. C., 1983: Sea ice effective microwave emissivities from satellite passive microwave and infrared observations. *J. Geophys. Res.*, **88**, 7686–7704.
- , P. Wadhams, W. B. Krabill, R. N. Swift, J. P. Crawford, and W. B. Tucker III, 1991: Top/bottom multisensor remote sensing of arctic sea ice. *J. Geophys. Res.*, **96**, 2693–2709.
- Davis, F. W., D. A. Quattrochi, M. K. Ridd, N. S.-L. Lam, S. J. Walsh, J. C. Michaelsen, J. Franklin, D. A. Stow, C. J. Johannsen, and C. A. Johnston, 1991: Environmental analysis using integrated GIS and remotely sensed data: Some research needs and priorities. *Photogram. Eng. Remote Sens.*, **57**, 689–697.
- Ehlers, M., 1991: Multisensor image fusion techniques in remote sensing. *ISPRS J. Photogram. Remote Sens.*, **46**, 19–30.
- Emery, W. J., J. Brown, and Z. P. Nowak, 1989: AVHRR image navigation: Summary and review. *Photogram. Eng. Remote Sens.*, **55**, 1175–1183.
- Frank, A. U., M. J. Egenhofer, and W. Kuhn, 1991: A perspective on GIS technology in the nineties. *Photogram. Eng. Remote Sens.*, **57**, 1431–1436.
- Franklin, S. E., and C. F. Blodgett, 1993: An example of satellite data fusion. *Comput. Geosci.*, **19**, 577–583.
- Gilbert, J. R., C. Moler, and R. Schreiber, 1992: Sparse matrices in MATLAB: Design and implementation. *SIAM J. Matrix Anal.*, **13**, 333–356.
- Grody, N. C., 1983: Severe storm observations using the Microwave Sounding Unit. *J. Climate Appl. Meteor.*, **22**, 609–625.
- Heymsfield, G. M., K. K. Gnosh, and L. C. Chen, 1983: An interactive system for compositing digital radar and satellite data. *J. Climate Appl. Meteor.*, **22**, 705–713.
- Hibbard, W., and D. Santek, 1989: Visualizing large data sets in the earth sciences. *Computer*, **22**, 53–57.
- Hillger, D. W., J. F. W. Purdom, and D. A. Molenaar, 1993: A noise level analysis of special multiple-spin VAS data during STORM-FEST. NOAA Tech. Rep., NESDIS 65, NOAA/NESDIS/RAMM Branch, Fort Collins, CO, 30 pp.
- Ho, D., and A. Asem, 1986: NOAA AVHRR image referencing. *Int. J. Remote Sens.*, **7**, 895–904.
- Hollinger, J., R. Lo, G. Poe, R. Savage, and J. Pierce, 1987: *Special Sensor Microwave/Imager User's Guide*. Naval Research Laboratory, Washington, DC, 120 pp.
- Jones, A. S., and T. H. Vonder Haar, 1990: Passive microwave remote sensing of cloud liquid water over land regions. *J. Geophys. Res.*, **95**, 16 673–16 683.

- , and —, 1992: PORTAL—A satellite remap/fusion system. Preprints, *Sixth Conf. on Satellite Meteorology and Oceanography and Eighth Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology*, Atlanta, GA, Amer. Meteor. Soc., J98–J101.
- , K. E. Eis, and T. H. Vonder Haar, 1993: Measurement of high resolution microwave surface emittance using advanced multisensor data fusion techniques. Preprints, *Topical Symp. on Combined Optical-Microwave Earth and Atmosphere Sensing*, Albuquerque, NM, IEEE/LEOS, 248–251.
- Kahn, R., R. D. Haskins, J. E. Knighton, A. Pursch, and S. Granger-Gallegos, 1991: Validating a large geophysical data set: Experiences with satellite-derived cloud parameters. *Proc. 23d Symp. on the Interface*, Computing Science and Statistics, 133–140.
- King, M., 1992: Editor's corner. *The Earth Observer*, **4**, EOS Project Office, Code 900, NASA Goddard Space Flight Center, 1–2.
- Kohn, N. M., A. L. Johnston, and C. G. Mohr, 1978: MUDRAS—Multiple Doppler Radar Analysis System. NOAA Tech. Memo. ERL/WPL-35, Wave Propagation Lab, Boulder, 170 pp.
- Lee, A. C. L., 1994: Filtering, sampling, and information content within satellite-derived multispectral or mixed-resolution imagery. *J. Atmos. Oceanic Technol.*, **11**, 401–421.
- Liou, K.-N., 1980: *An Introduction to Atmospheric Radiation*. Academic Press, Inc., 392 pp.
- Lunetta, R. S., R. G. Congalton, L. K. Fenstermaker, J. R. Jensen, K. C. McGwire, and L. R. Tinney, 1991: Remote sensing and geographic information system integration: Error sources and research issues. *Photogram. Eng. Remote Sens.*, **57**, 677–687.
- Mohr, C. G., L. J. Miller, R. L. Vaughan, and H. W. Frank, 1986: The merger of mesoscale datasets into a common Cartesian format for efficient and systematic analyses. *J. Atmos. Oceanic Technol.*, **3**, 143–161.
- NASA, 1991: *EOS Reference Handbook*. David Dokken, Ed., NASA Goddard Space Flight Center, NP-144, 147 pp.
- , 1992: *EOS Data and Information System (EOSDIS)*. Earth Science and Applications Division (Code SE), NASA Headquarters, Washington DC, 31 pp.
- NCSA, 1994: *Hierarchical Data Format (HDF) Reference Manual V3.3*. National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.
- Neale, C. M. U., M. J. McFarland, and K. Chang, 1990: Land-surface-type classification using microwave brightness temperatures from the Special Sensor Microwave/Imager. *IEEE Trans. Geosci. Remote Sens.*, **28**, 829–838.
- NSSDC, 1994: *CDF User's Guide, Version 2.4*. National Space Science Data Center, NASA/Goddard Space Flight Center.
- Oye, R., and E. Mueller, 1986: A media efficient common logical format for radar and lidar data. Preprints, *23d Radar Meteorology Conf.*, Snowmass, CO, Amer. Meteor. Soc., JP361–JP362.
- Paris, J. F., and H. H. Kwong, 1988: Characterization of vegetation with combined Thematic Mapper (TM) and Shuttle Imaging Radar (SIR-B) image data. *Photogram. Eng. Remote Sens.*, **54**, 1187–1193.
- Poe, G. A., 1990: Optimum interpolation of imaging microwave radiometer data. *IEEE Trans. Geosci. Remote Sens.*, **28**, 800–810.
- , and R. W. Conway, 1990: A study of the geolocation errors of the Special Sensor Microwave/Imager (SSM/I). *IEEE Trans. Geosci. Remote Sens.*, **28**, 791–799.
- Prisley, S. P., T. G. Gregoire, and J. L. Smith, 1989: The mean and variance of area estimates computed in an arc-node geographic information system. *Photogram. Eng. Remote Sens.*, **55**, 1601–1612.
- Purdum, J. F. W., and P. N. Dills, 1994: Cloud motion and height measurements from multiple satellites including cloud heights and motions in polar regions. Preprints, *Seventh Conf. on Satellite Meteorology and Oceanography*, Monterey, CA, Amer. Meteor. Soc., 408–411.
- Raymond, D. J., 1988: A C language-based modular system for analyzing and displaying gridded numerical data. *J. Atmos. Oceanic Technol.*, **5**, 501–511.
- Rew, R. K., and G. Davis, 1990: NetCDF: An interface for scientific data access. *IEEE Comput. Graphics Appl.*, **10**, 76–82.
- Star, J. L., J. E. Estes, and F. Davis, 1991: Improved integration of remote sensing and geographic information systems: A background to NCGIA initiative 12. *Photogram. Eng. Remote Sens.*, **57**, 643–645.
- Steffen, K., and A. J. Schweiger, 1990: Multisensor approach to sea ice classification for the validation of DMSP-SSM/I passive microwave derived sea ice products. *Photogram. Eng. Remote Sens.*, **56**, 75–82.
- Straka, J., J. Klokočník, H. Graßl, 1993: Navigation of satellite measurements without ground control points. *Int. J. Remote Sens.*, **14**, 1981–2004.
- Suomi, V. E., R. Fox, S. S. Limaye, and W. L. Smith, 1983: McIDAS III: A modern data access and analysis system. *J. Climate Appl. Meteor.*, **22**, 766–778.
- Treinish, L. A., 1989: An interactive, discipline-independent data visualization system. *Comput. Phys.*, **3**, 55–64.
- , and M. L. Gough, 1987: A software package for the data independent management of multi-dimensional data. *EOS Trans.*, **68**, 633–635.
- Ulaby, F. T., R. K. Moore, and A. K. Fung, 1986: *Microwave Remote Sensing Active and Passive*. Vol. 3, *From Theory to Applications*, Addison-Wesley Publishing Co., 1097 pp.
- Unidata, 1993: *NetCDF User's Guide, Version 2.3*. Unidata Program Center, Boulder, CO, 118 pp.
- Welch, R., and M. Ehlers, 1987: Merging multiresolution SPOT HRV and Landsat TM data. *Photogram. Eng. Remote Sens.*, **53**, 301–303.
- Wolberg, G., 1990: *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, CA, 318 pp.