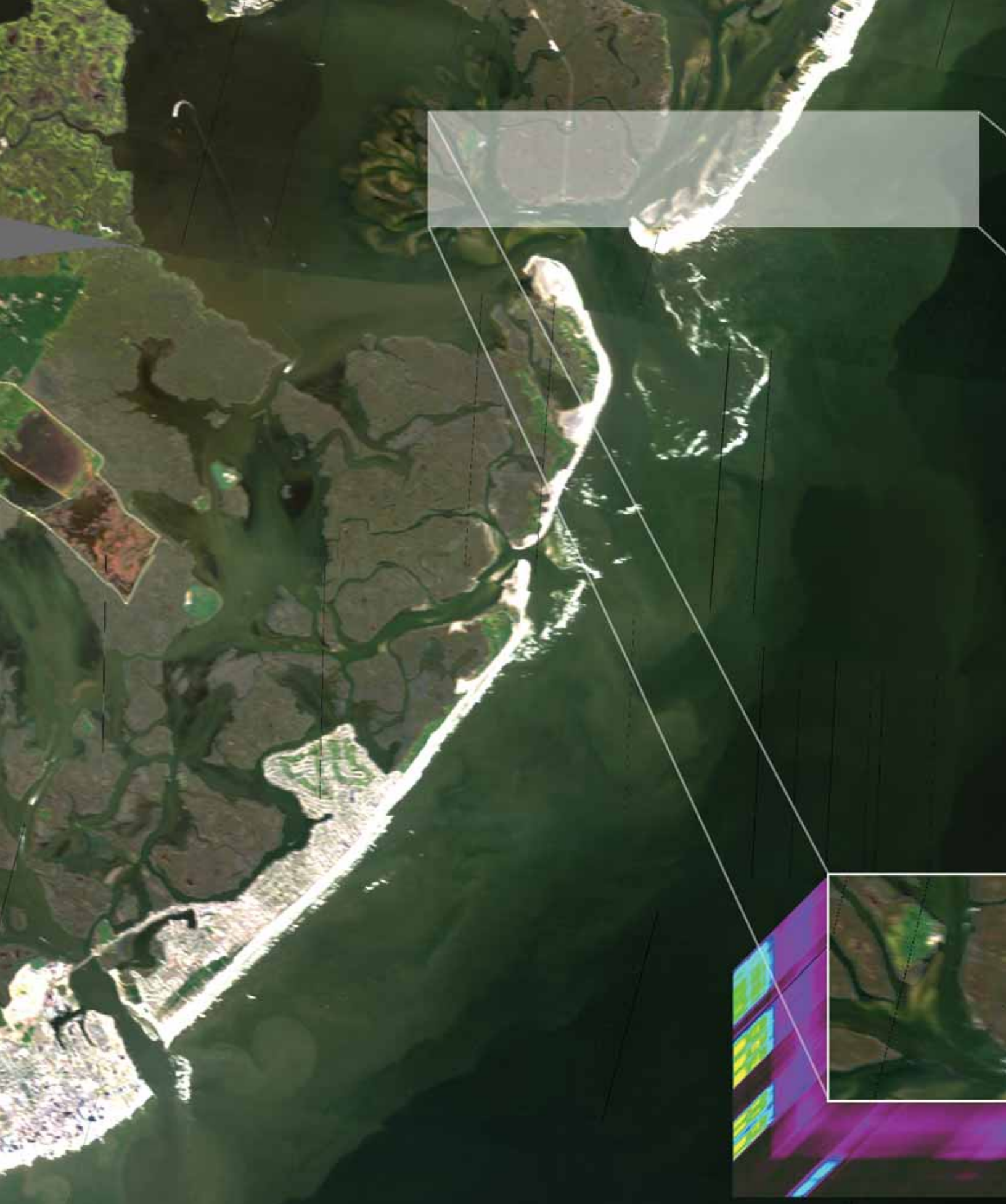

From Meters to Kilometers

A LOOK AT OCEAN-COLOR SCALES OF VARIABILITY, SPATIAL COHERENCE, AND THE NEED FOR FINE-SCALE REMOTE SENSING IN COASTAL OCEAN OPTICS

BY W. PAUL BISSETT, ROBERT A. ARNONE, CURTISS O. DAVIS, TOMMY D. DICKEY, DANIEL DYE, DAVID D.R. KOHLER, AND RICHARD W. GOULD, JR.

INTRODUCTION

The physical, biological, chemical, and optical processes of the ocean operate on a wide variety of spatial and temporal scales, from seconds to decades and from micrometers to thousands of kilometers (Dickey et al., this issue; Dickey, 1991). These processes drive the accumulation and loss of living and nonliving mass constituents in the water column (e.g., nutrients, phytoplankton, detritus, sediments). These mass constituents frequently have unique optical characteristics that alter the clarity and color of the water column (e.g., Preisendorfer, 1976). This alteration of the ocean color, or more specifically the change in the spectral "water-leaving radiance," $L_w(\lambda)$, has led to the development of optical techniques to sample and study the change in biological and chemical constituents (Schofield et al., this issue). Thus, these optical techniques provide a mechanism to study the effects of underlying biogeochemical processes. In addition, because time- and space-dependent changes in $L_w(\lambda)$ may be measured remotely, optical oceanography provides a way to sample ecological interactions over a wide range of spatial and temporal scales.

The question often posed by scientists trying to resolve problems involving the temporal and spatial variation of oceanic properties is: "What is the optimal time/space sampling frequency?" The obvious answer is that the sampling frequency should be one half the frequency of the variation (i.e., Nyquist frequency) of the property of interest. However, therein lies the rub for the oceanographer: the range of the relevant scales is large, and the range of available resources and/or actual engineering capabilities to sample all relevant scales is often small. Hence, the decisions affecting resource allocation become critical in order to maximize the total data information in both quantity and quality. While these scientific resource decisions are rarely discussed in explicit terms of cost-benefit analysis, such discussions should be integral parts of the scientific design of instruments, platforms, and experiments aimed at resolving oceanic processes.

The practical examples of this problem in remote sensing include: "What is the optimal repeat coverage frequency?" and "What is the optimal Ground Sample Distance (GSD) or pixel size of the data?" For the optical oceanographer, there is also the issue of optimal spectral coverage needed to resolve the optical constituents of interest (Chang et al., this issue). The sum of these considerations feed into the sensor, deployment platform, and deployment schedule decisions. For polar-orbiting and geo-stationary satellites that cost hundreds of millions of dollars, as well as airborne sensors that have smaller upfront costs but higher deployment costs, the decision of sampling frequency directly impacts the scientific use of the data stream, and what processes may be addressed with data streams collected by these sensors. These scientific cost-benefit analyses extend beyond the cost in dollars because the typical lifetime and replacement cycle of these sensors is on the order of years to decades, and a poorly designed sensor package is very difficult to replace.

W. Paul Bissett *(pbissett@flenvironmental.org) is Research Scientist, Florida Environmental Research Institute, Tampa, FL.* **Robert A. Arnone** *is Head, Ocean Sciences Branch, Naval Research Laboratory, Stennis Space Center, MS.* **Curtiss O. Davis** *is at Remote Sensing Division, Naval Research Laboratory, Washington, DC.* **Tommy D. Dickey** *is Professor, Ocean Physics Laboratory, University of California, Santa Barbara, Goleta, CA.* **Daniel Dye** *is at Florida Environmental Research Institute, Tampa, FL.* **David D.R. Kohler** *is Senior Scientist, Florida Environmental Research Institute, Tampa, FL.* **Richard W. Gould, Jr.** *is Head, Ocean Optics Center, Naval Research Laboratory, Stennis Space Center, MS.*

In 2001, the Office of Naval Research (ONR) sponsored the Hyperspectral Coastal Ocean Dynamics Experiment (HyCODE) (Dickey et al., this issue), which presented the opportunity to study the question of scales of variability in remote-sensing data. Hyperspectral airborne sensors were deployed on several platforms at various altitudes. This coverage was supplemented by numerous space-borne, remote-sensing satellites. The airborne instruments included two versions of the Portable Hyperspectral Imager for Low-Light Spectroscopy (PHILLS 1 and PHILLS 2) (Davis et al., 2002) operating at an altitude of less than 10,000 feet and 30,000 feet, respectively, as well as the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor operating at 60,000 feet. These sensors provided hyperspectral data at 2 m, 9 m, and 20 m GSDs, respectively. The satellite data collected included the multi-spectral images from Sea-viewing Wide Field-of-view Sensor (SeaWiFS), Moderate Resolution Imaging Spectroradiometer (MODIS), Fengyun 1 C (FY1-C), Oceansat as well as the multi-spectral polarimeter Multiangle Imaging SpectroRadiometer (MISR) sensor and sea surface temperature (SST) sensor Advanced Very High Resolution Radiometer (AVHRR). These collections provided a wealth of remote-sensing and field data during a spatially and temporally intense oceanographic field campaign, and they offered the ability to begin to address the issue of optimal sampling scales for the coastal ocean.

The use of these multiple remote-sensing data streams requires the calibration, validation, and atmospheric correction of the sensor signals to retrieve estimates of $L_w(\lambda)$, or "remote sensing reflectance," $R_{rs}(\lambda)$, a normalized measure of the $L_w(\lambda)$. Our goals in this paper are to illuminate some of the issues of remote sensing spatial scaling in the nearshore environment and attempt to derive some understanding of appropriate

sampling scales in the nearshore environment. We will focus on the data collected by a single sensor (PHILLS 2) to reduce uncertainties in the analysis that may result from the different data processing techniques applied to each of the individual sensors' data.

METHODS

The PHILLS 2 was deployed seven times (July 21st, 23rd, 27th, 31st a.m., 31st p.m., and August 1st and 2nd) during the 2001 HyCODE LEO-15 field program, and each mission generated nearly 4,000 square kilometers of spectral data at 9 m resolution (Figure 1). For this discussion on spatial scaling, we have chosen to focus on a single PHILLS 2 image from July 31st, as the coverage provided by these data approximates the total spatial extent of a satellite sensor, at a much higher spatial resolution, which allows us to explore scaling issues within a single image cube. The calibration of the sensor (Kohler et al., 2002a) and the specific corrections for the window as well as the atmospheric correction of these data are described elsewhere (Kohler et al., 2002b). The data values are given in $R_{rs}(\lambda)$, units of 1/sr (Mobley, 1994). This single 124-band data cube from July 31, 2001 represents 15 GB of raw data. This data was calibrated, atmospherically corrected, and geo-rectified for the analyses presented here.

The engineering issues surrounding the collection, storage, and transmission of higher spatial and spectral resolution systems are fairly cost intensive. If this image was collected from space, it would require over three hours to transmit the data to a ground station over an X-band downlink (for reference, a polar-orbiting satellite has approximately an 11-minute transmission window). One of the easiest ways to reduce data density is to reduce spectral resolution. However, reducing spectral resolution also reduces the biogeochemical information that may be derived from optical data. To look at the impacts of spectral resolution reduction on the ability to discern spatial variability in the spectral $R_{rs}(\lambda)$ data, the hyperspectral data were reduced in spectral resolution to approximate the SeaWiFS bands. This was accomplished by multiplying the $R_{rs}(\lambda)$ by the SeaWiFS wavelength response function (Figure 2). This created an 8-band image, with band centers located at 412, 443, 490, 510, 555, 670, 765, and 865 nm. These data are used to illuminate the different multispectral and hyperspectral data streams to resolve information variability in the nearshore environment.

The autocorrelation function has previously been used in time-series studies to determine the optimal time frequency of sam-

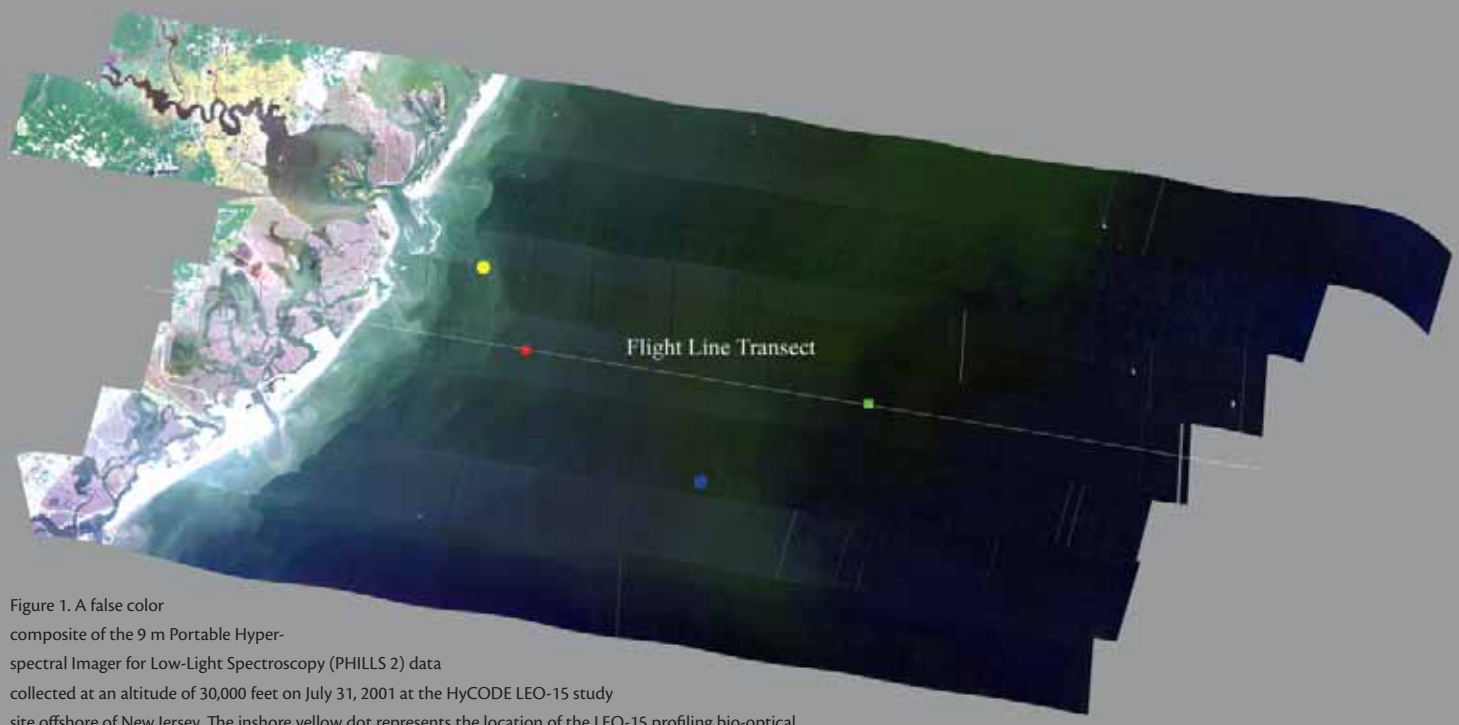

Figure 1. A false color composite of the 9 m Portable Hyperspectral Imager for Low-Light Spectroscopy (PHILLS 2) data collected at an altitude of 30,000 feet on July 31, 2001 at the HyCODE LEO-15 study site offshore of New Jersey. The inshore yellow dot represents the location of the LEO-15 profiling bio-optical node. The offshore blue dot is the location of the UCSB OPL (University of California, Santa Barbara, Ocean Physics Laboratory) bio-optical mooring. The inshore small red box and the offshore small green box represent regions of interest (ROIs) where the variance of the SeaWiFS Band 5 R_{rs}, PC1 (SW) and PC1 (Hyp) were approximately the same, even though the mean was significantly different (see text and Table 1). The size of these boxes represents a mean ground sampling distance (GSD) of 441 m (49 pixels on a side for a total of 2401 pixels equal to approximately 0.2 km^2). The white line represents the transect data used in the variable GSD study. Its selection was driven by the desire to use a single flight line of data for the variance calculation (see text).

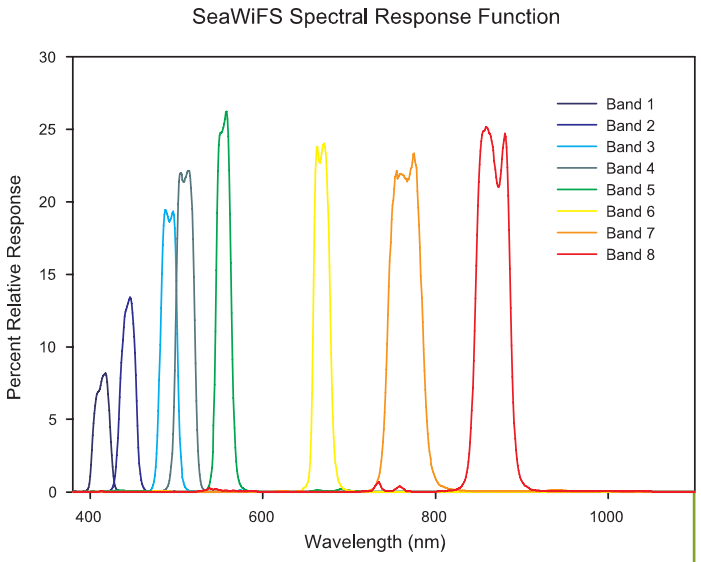

Figure 2. To evaluate the effect of reduced spectral resolution on spatial variability, a reduction in the spectral resolution of the hyperspectral data was performed so as to approximate that of SeaWiFS bands. Shown are the SeaWiFS wavelength response functions used to transform the hyperspectral PHILLS 2 data into a simulated SeaWiFS-type data product.

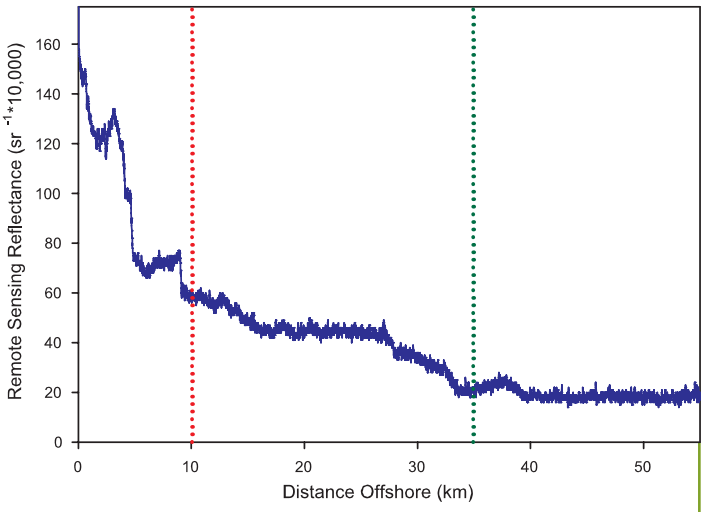

Figure 3. The Simulated SeaWiFS Band 5 R_{rs} values (sr^{-1}*10,000) along the sampling line transect as shown in Figure 1. The vertical green and red lines denote the respective locations of offshore and inshore regions of interest from which the variance threshold for the GSD analysis was determined.

pling (e.g., Abbott and Letelier, 1998; Chang et al., 2002; Dickey et al., 2001), which led us to attempt a spatial autocorrelation to examine spatial variability of $R_{rs}(555)$ along the transect shown in Figure 1 (Figure 3). However, results indicated a trend in this data record, with higher intensities of $R_{rs}(555)$ nearshore. Autocorrelation studies require the mean of any subsample of a record to approximate the mean of the total record. Any attempt to calculate a decorrelation scale from this transect would result in a value for the decorrelation scale that had a direct proportional relationship to the total length of the transect. The statistical reason for this is that data from the transect does not represent a stationary function (i.e., the mean changes as the transect length increases and, therefore, the sample variance will increase with domain size) (Chilès and Delfiner, 1999) (see Statistics Review box). This suggests that the measure (decorrelation scale) may be an improper statistic to use to describe the optimal spatial sampling frequency for coastal-ocean data sets.

While not statistically explored here, the change in mean (and thus, variance) appears to be different over cross-shore distances compared to along-shore distances. There are other statistical methods for estimating spatial variability, including those that determine the anisotropy in the directionality of the variance calculation (i.e., the variance is different in different directions) (Curran, 1988; Dale et al., 2002). Variance ellipses have been used to describe the variance in altimeter-derived velocities in the nearshore environment (e.g., Strub and James, 2000). Of particular interest may be the use of semivariogram or variograms developed in the soil research community to describe "roughness" in the topology of spatial measurements (Curran, 1988). These have been used with satellite ocean remote-sensing data to describe larger scales of interest in chlorophyll distributions (e.g., Yoder et al., 1987). However, many of these methods require interpretations that are difficult to definitively relate to geophysical parameters, i.e., "sills" and "nuggets" in variograms. Here, we are interested in determining an optimal sampling size that is more easily discussed in terms of this scene of interest and in terms of the sensor capabilities. In other words, we would like the scene itself to describe the optimal GSD based on the ability of the sensor and hyperspectral data to resolve distinct, homogeneous waters. The more rigorous application of 2-D variance analyses is the subject of a follow-on study.

For this study, we derived another method of estimating spatial variability, one that focuses on the ability to separate the linearly additive noise of the image from the "real" geophysical detail of the scene. Linearly additive noise refers to the interference derived from the noise of the sensor as well as any noise generated from the atmosphere or processing algorithms. If the noise is stable and linear, then the true signal of interest may be retrieved from a sample of a population, provided the sample size is sufficiently large. In this case, we would expect that the standard deviation of the $R_{rs}(\lambda)$ signal to be a proxy for the total noise, and that over any homogenous region of the scene it should be constant, regardless of the magnitude of the signal. Thus, any pixel in a homogenous region of interest would be equal to the mean value of the region ± some random component.

This analysis would fail if there were multiplicative noise in our data. Multiplicative noise occurs when the noise (i.e., standard deviation) is a direct function of the intensity (mean) of the signal. By assuming (and confirming) that the noise of the scene is truly random and linear, any increase in the standard deviation would thus be generated by a change in real geophysical properties within the region of interest. Therefore, an increase in a region's standard deviation above a background random noise-generated standard deviation would suggest a nonhomogenous region of interest, i.e., one with real differences in the region's geophysical properties. Put simply, a region of interest with a standard deviation greater than a background noise-generated standard deviation must contain regions of distinctly different optical constituents.

Most studies of spatial variation in radiance fields focus on the variance within a single channel, or perhaps a combination of channels. In this study, we wish to assess if there is any additional information to be retrieved from the continuous spectrum of reflectance data, as opposed to using only one or two bands individually. The question of how to use the entire hyperspectral data simultaneously to identify homogeneous regions of optical properties is an active area of research; as a first step, we would like to be able to determine if the full spectrum offers any ability over single or multichannel data to separate water masses into distinct optical regions. One approach to using the full spectrum simultaneously is to first linearly transform the n-dimensional spectral data (where n is the number of wavelengths) into a variance minimizing coordinate system. When the "proper" or root vectors (eigenvectors) of this new coordinate system are orthogonal to each other, this type of transformation is called a Principal Component Analysis (PCA). A PCA allows the user to focus on the vectors that describe the most variance (information) using the entire spectral and image space, rather than focusing only on the variance in the image at a single wavelength A PCA is a powerful way to look for patterns.

STATISTICS REVIEW

When analyzing any data set, a good place to start is by calculating the data set's mean (a measure of the central tendency) and variance (a measure of the dispersion or variability). The mean is given by the following equation:

$$\mu = \frac{\sum_{i=1}^{n} X_i}{N}$$

where the capital Greek letter sigma (Σ) means summation over all values, X_i, in the population, divided by the total number of values, N, in the population. The variance is given by:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{N}$$

It is easily seen that the greater the separation of the individual values, X_i, are from the mean, μ, the larger the variability of the population represented in the data set grows. The square root of the variance is called the standard deviation. When the population is normally distributed around its mean, the standard deviation provides a measure that is easily conceptualized as a distance away from the mean. The standard deviation may also be used to produce a confidence interval in populations that are normally distributed. In such data sets, one would expect 68 percent of the population to fall with 1 standard deviation (1σ) around the population's mean. The probabilities that any member of the population would fall within 2 and 3σ are approximately 95 and 99 percent, respectively.

In spatial data analysis, one is frequently interested in how a sample at one spot co-varies or correlates with the same measure of a sample in another location. Autocovariance and autocorrelation are simply measures of the covariance and correlation of the values of a single variable for all pairs of points separated by a given spatial lag (Dale et al., 2002). An estimate of the autocovariance for samples at a distance d is given by:

$$\overline{Cov} = \frac{\sum_{i=1}^{n-d} (X_i - \mu)(X_{i+d} - \mu)}{N - d}$$

The autocorrelation is given by dividing *Cov* by σ^2. The value of these statistics in describing the data set of interest depends on the validity of the underlying assumptions. A trend in the spatial data (similar to Figure 3) violates the assumption of stationarity, i.e. the estimate of the mean and the autocorrelation are constant with respect to distance along the record, and negates the effectiveness of the autocovariance and autocorrelation in the overall analysis.

Other measures, such as sample variograms and Pair Quadrat Variance (PQV), focus only on the change with lag distance. For a transect of n contiguous or equally spaced intervals (quadrats), a sample variogram for a given distance d is given by (Dale et al., 2002):

$$\hat{\gamma}(d) = \frac{\sum_{i=1}^{n-d} (X_i - X_{i+d})^2}{N - d}$$

Note that this equation is omnidirectional. If the sample variogram is constant with respect to direction, it is referred to as isotropic. If the variogram changes with respect to the direction with which it was calculated, then it is referred to as anisotropic. It is clear from Figures 1 and 6 that there appears to be a directionality component to the along shore and cross shore variance, and thus this image would be consider anisotropic.

REFERENCES

Dale, M.R.T., P. Dixon, M.-J. Fortin, P. Legendre, D.E. Myers, and M.S. Rosenberg, 2002: Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25, 558-577.

It should be noted that great care must be used in analyzing a PCA transformation of a hyperspectral image. There will be an equal number of eigenvectors as there are spectral channels, but frequently only the first 10 eigenvectors are necessary to describe >99 percent of the total variance in the scene. However, the number of eigenvectors needed to describe the total variance in the scene is completely image dependent. If there is a large amount of spectral variation in the scene, then more eigenvectors will be needed to describe the majority of the scene variance. If there is a small amount of spectral variation, then a smaller number of eigenvectors will be required. As an example, many open-ocean images have been found to only need the first three eigenvectors to describe 98 percent of the scene dependent variance (e.g., Mueller, 1976). For these ocean images, a common error in PCA is to assume that only three spectral channels are needed to describe the scene dependent variance. It must be understood that the eigenvector is a measure of the variance across all bands simultaneously, and therefore requires the user to recognize that it is a hyperspectral vector itself, which could not have been generated without the full spectral data set. The easiest way to see the impacts of all of the wavelengths on the eigenvector is to square the PCA eigenvectors to calculate each channel's percentage contribution to the description of scene's spectral variation.

We seek to use the hyperspectral data to separate homogenous water masses, and we believe that there is additional information in the full spectrum of the radiance field, rather than in any single channel or combination of channels. In order to test this belief, we will compare the spatial variability of three images created from the same hyperspectral data set. The first image is a single simulated SeaWiFS band (Band 5). The second is an image of the eigenvalues of the first eigenvector created from a PCA (referred to as PCA SW) of a simulated two band SeaWiFS image (Bands 3 and 5). These two bands were selected for this multispectral test, since they are used in many common SeaWiFS chlorophyll algorithms (O'Reilly et al., 1998). The third image is an image of the eigenvalues of the first eigenvector created from a PCA (referred to as PCA Hyp) of the hyperspectral image. We used the Environment for Visualizing Images (ENVI) software package from Research Systems, Inc., to accomplish a PCA of the hyperspectral data. The first eigenvector (PC1) of both the hyperspectral and two band images described >95 percent of the variance of the images; PCA Hyp PC1 = 95.6 percent and PCA SW PC1 = 99.0 percent. The second and third eigenvector of the PCA Hyp accounted for 2.9 percent, and 0.7 percent of the image's spectral variance, respectively. The total variance described by the remaining eigenvectors for PCA Hyp is 1.43 percent. There are only two eigenvectors for the PCA SW, and the second accounts for 1 percent of the variance.

The first three eigenvectors from the PCA Hyp as well as the percentage contribution from each spectral channel to each eigenvector, is shown in Figure 4. It is clear that while there are some dominant channels in the first eigenvector (i.e., approximately 560 nm in PC1), it peaks at only approximately 6 percent, which means that the other wave-

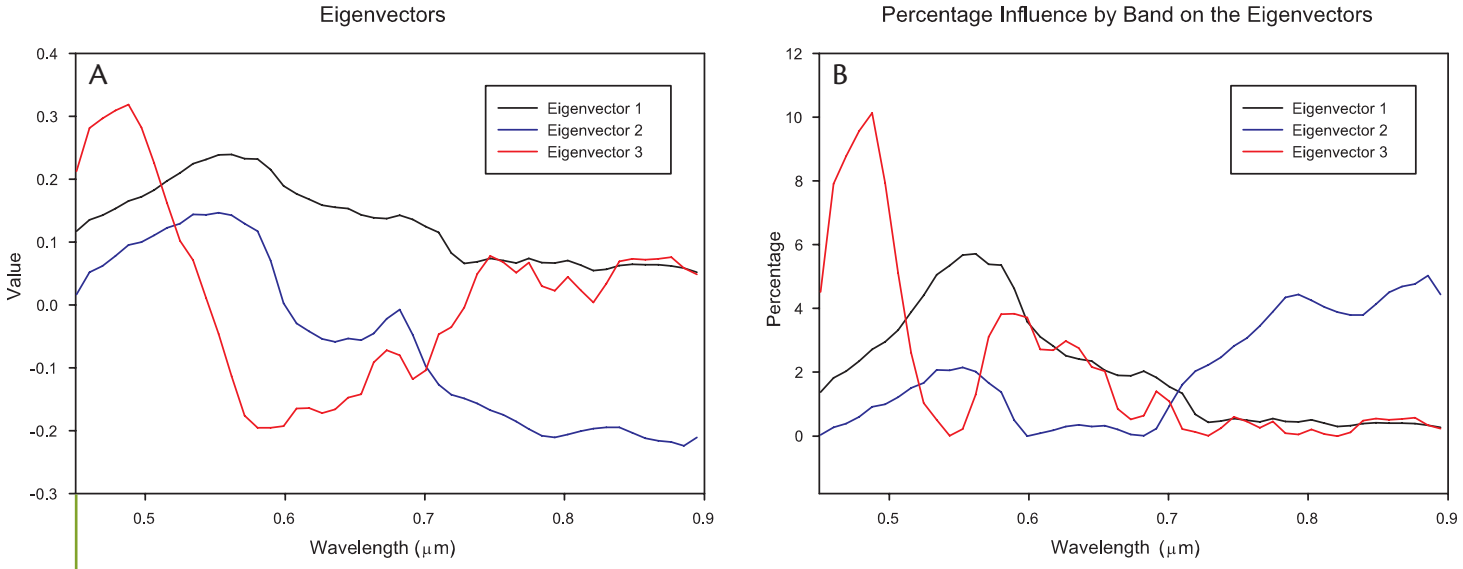

Figure 4. So as to evaluate the entire spectral data set, Principal Component Analysis (PCA) was used to reduce the dimensionality of the image. PCA is a method of maintaining nearly all of the characteristics of the original data set while reducing the number of parameters needed to describe the data. This is accomplished by reprojecting those data along orthogonal axes that are positioned to best describe the variance of the data (eigenvalues and eigenvectors). The first three eigenvector principal components are displayed in (A). The influence that each spectral band had on the first three principal components is displayed in (B).

lengths contribute 94 percent of the influence on the variance described by this vector (Figure 4A). Therefore, it would not be accurate to say a single channel would describe 95.6 percent of the variance in this image. A more accurate statement would be that the spectral shape that describes the most variance in this image is demonstrated in the first eigenvector.

Two regions of interest (ROIs) in the visually homogenous areas of the imagery were selected to confirm the hypothesis of linear noise (which should be applicable to the PCA because it is a linear transformation of the hyperspectral data) and to generate a test standard deviation value. These two ROIs were approximately 441 m on a side (49 pixels), (approximately 0.2 km^2, 2401 pixels); one region was inshore while the other was offshore (Figure 1). Table 1 gives the mean and standard deviations for the two ROIs from the simulated SeaWiFS Band 5 as well as the PC1 for PCA SW and PCA Hyp. Note this standard deviation is not normalized by the mean (e.g., Mahadevan and Campbell, 2002; Mahadevan and Campbell, in press) because we are trying to separate random noise of the sensor and processing from the real geophysical changes in the image. Theoretically, any homogenous region of the same size should have a similar standard deviation; otherwise, some real feature of interest has been included within the study region. As the ROIs were selected with an eye to a perfectly homogenous region, we allow for some error in our selection criteria. Table 1 also provides the test standard deviation, σ_t, for each of the GSD calculations.

Image Type	Region of Interest, ROI	Mean of ROI	Standard Deviation of ROI	Standard Deviation Used in Analysis, σ_t
Simulated SeaWiFS Band 5	Inshore	52.77	1.54	2.0
	Offshore	24.55	1.85	
PC1 of Simulated SeaWiFS Bands 3 and 5	Inshore	13.38	1.79	2.2
	Offshore	-21.68	1.98	
PC1 of Hyperspectral Cube	Inshore	111.08	6.44	8.0
	Offshore	-12.11	7.46	

Table 1. The mean and standard deviation from the simulated single-band image (SeaWiFS Band 5) as well as the first Principal Component Analysis eigenvalue images from simulated dual band (SeaWiFS Band 3 and 5), and hyperspectral data. Also, included is the test standard deviation, σ_t, used for the optimal GSD calculation.

Next, at every pixel along the transect line (Figure 1), a new ROI_i was created with a minimum size of 3 X 3 pixels, or 27 X 27 m (729 m^2). The mean and standard deviation, σ, of each region was calculated, and σ_i was compared against σ_t. If σ_i was less than σ_t, then ROI_i increased in size by two pixels in each of the along-track and cross-track directions, while remaining centered on pixel i, and the mean and standard deviations were recalculated. This procedure continued until σ_i was greater than σ_t, at which point the size of the previous non-failing ROI_i was recorded. The size of the ROI_i should then equate to the maximum size of a region with homogenous optical properties.

RESULTS AND DISCUSSION

The results of this approach in describing the spatial variability of this coastal environment may be found in Figure 5. Here, the largest GSD of the ROI_i that has a standard deviation greater than or equal to σ_t is plotted as a function of the position along the transect for three images: single band (Figure 5A), dual band PC1 (Figure 5B), and hyperspectral PC1 (Figure 5C). It can be seen that the size of the GSD increases when moving from onshore to offshore. The optimal GSD for each data set increases rapidly out of the surf zone to an average of approximately 100 m within 200 m of the shore. By about 10 km, the optimal GSD grows to >1 km. The average and median optimal GSD for all vary between 150 and 200 m out to 5 km, with the average GSD growing to approximately 1 km beyond approximately 12 km from the shoreline.

The variability inshore for each GSD calculation is driven primarily by intensity differences, probably resulting from the sediments suspended during the passage of the weather front. This variability is represented in Figure 6, as a false color composite of the PCA Hyp PC1 eigenvalues rendered in density slices. Clearly, there is a tremendous amount of spatial variability inshore, which decreases as we move offshore. As we move offshore past 20 km, the optimal GSD increases for each test. However, beyond this point there are significant differences between the SW Band 5 and PCA SW and the PCA Hyp. The optimal average GSD and median GSD grow to approximately 2 km and approximately 1.5 km, respectively, for SW Band 5 as the water masses become more homogeneous with respect to this wavelength. The average and median GSD for the PCA SW and PCA Hyp are less, as the additional bands of information provide improved ability to delineate water-mass types. There is some additional geophysical structure between 28 and 40 km that reduces the optimal GSD back to the levels seen nearshore for all three tests. Once offshore more than 40 km, the optimal GSD grows to > 6 km for the Band 5 test, and > 4 km for the PCA SW. These larger GSDs approach the scale of chlorophyll distributions described by others in the coastal environment using multispectral data (e.g., Yoder et al., 1987). However, the PCA Hyp drops back to

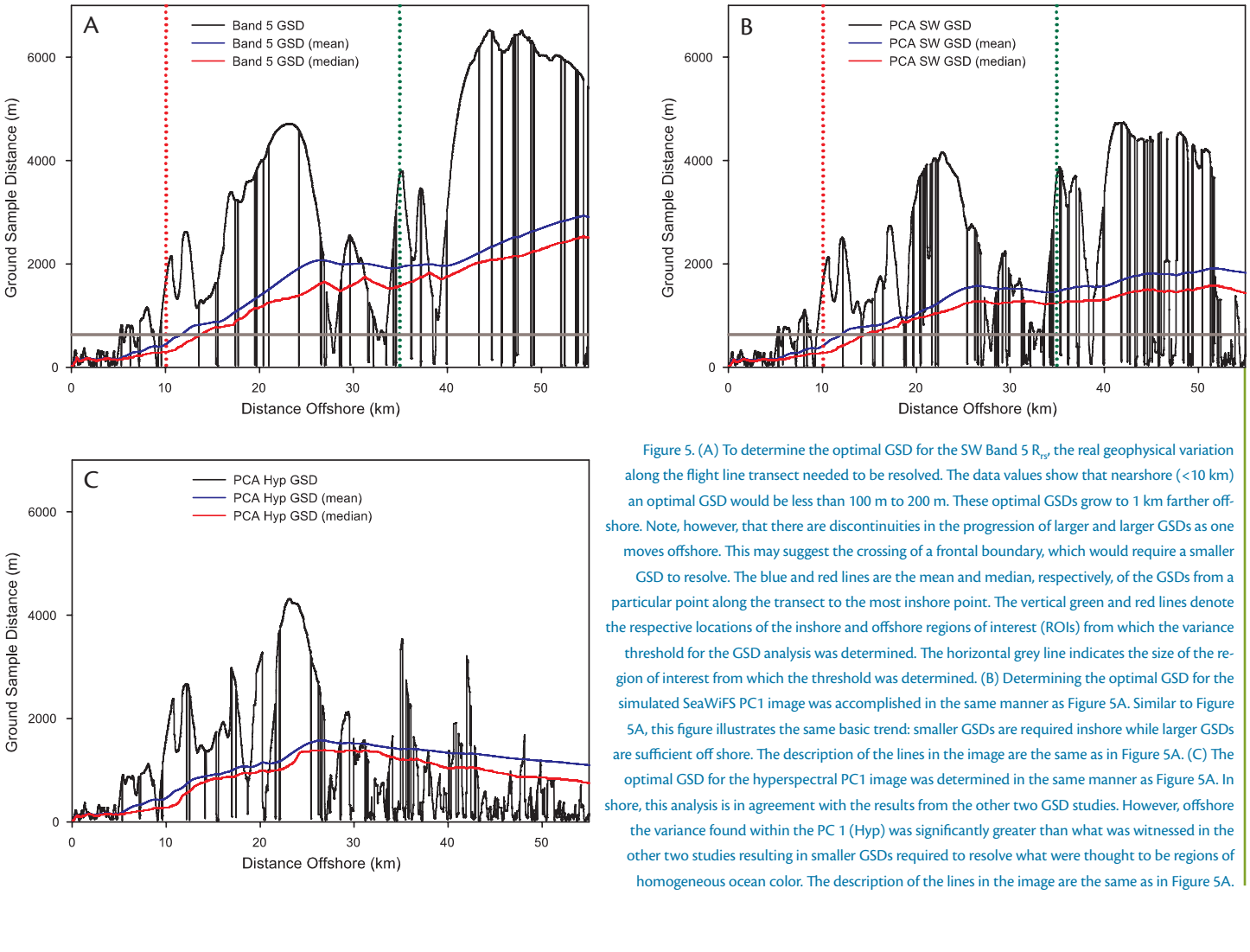

Figure 5. (A) To determine the optimal GSD for the SW Band 5 R_{rs}, the real geophysical variation along the flight line transect needed to be resolved. The data values show that nearshore (<10 km) an optimal GSD would be less than 100 m to 200 m. These optimal GSDs grow to 1 km farther offshore. Note, however, that there are discontinuities in the progression of larger and larger GSDs as one moves offshore. This may suggest the crossing of a frontal boundary, which would require a smaller GSD to resolve. The blue and red lines are the mean and median, respectively, of the GSDs from a particular point along the transect to the most inshore point. The vertical green and red lines denote the respective locations of the inshore and offshore regions of interest (ROIs) from which the variance threshold for the GSD analysis was determined. The horizontal grey line indicates the size of the region of interest from which the threshold was determined. (B) Determining the optimal GSD for the simulated SeaWiFS PC1 image was accomplished in the same manner as Figure 5A. Similar to Figure 5A, this figure illustrates the same basic trend: smaller GSDs are required inshore while larger GSDs are sufficient off shore. The description of the lines in the image are the same as in Figure 5A. (C) The optimal GSD for the hyperspectral PC1 image was determined in the same manner as Figure 5A. Inshore, this analysis is in agreement with the results from the other two GSD studies. However, offshore the variance found within the PC 1 (Hyp) was significantly greater than what was witnessed in the other two studies resulting in smaller GSDs required to resolve what were thought to be regions of homogeneous ocean color. The description of the lines in the image are the same as in Figure 5A.

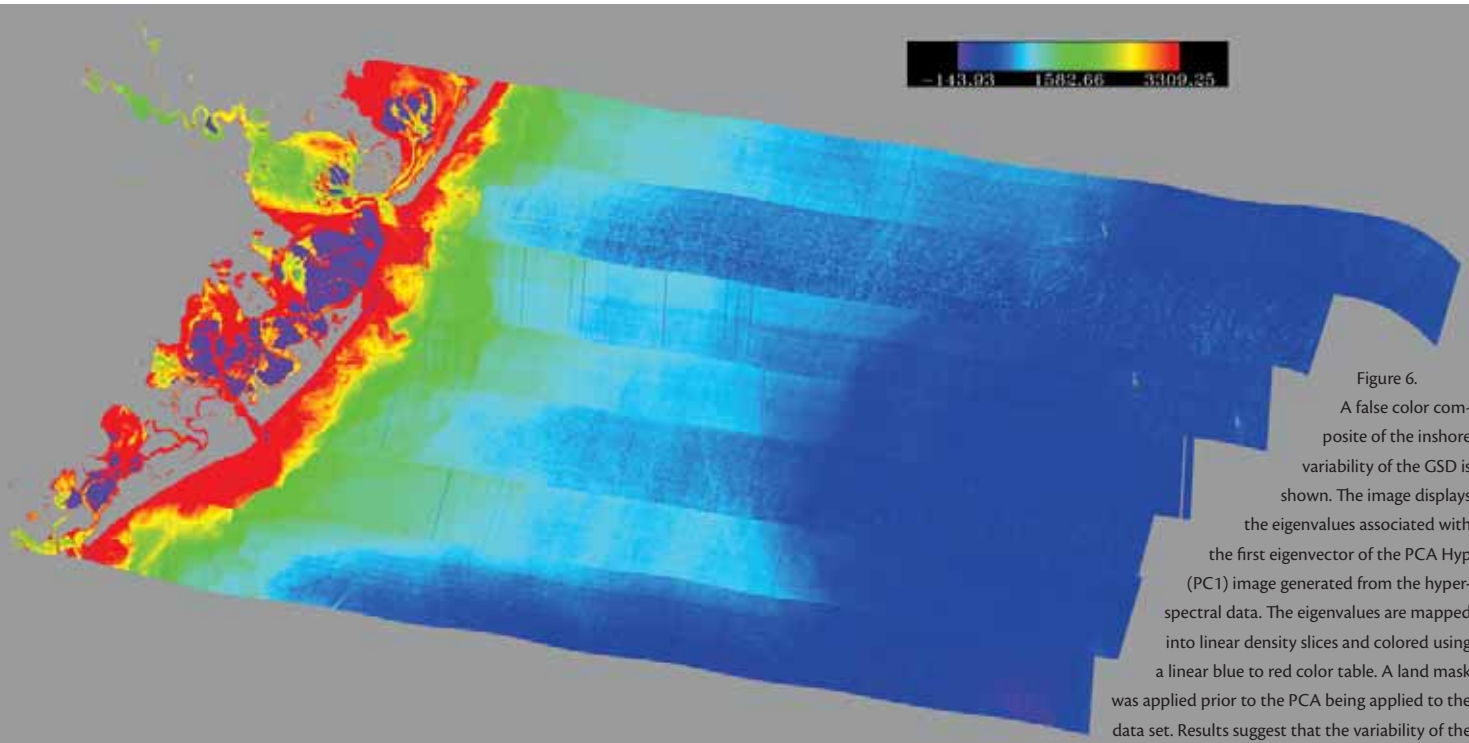

Figure 6. A false color composite of the inshore variability of the GSD is shown. The image displays the eigenvalues associated with the first eigenvector of the PCA Hyp (PC1) image generated from the hyperspectral data. The eigenvalues are mapped into linear density slices and colored using a linear blue to red color table. A land mask was applied prior to the PCA being applied to the data set. Results suggest that the variability of the water color is greater as one approaches the shore.

levels seen nearshore, suggesting the hyperspectral data offers additional information with which to separate features in otherwise homogeneous-appearing waters.

In many respects, these statistical results confirm what many coastal oceanographers intuitively understand. The closer to shore one approaches, the more variable the color of the water. In the optically deep waters off the coast of New Jersey, the optical features are driven by the wind and tidal mixing of sediments as well as allochthonous inputs of sediments, nutrients, and organic material from rivers and estuaries. In addition, far field dynamics drive coastal jets that also bring in allochthonous material into this location (Chant et al., in press), which are tidally mixed with nearshore waters. This drives the variability of the optical signal to a very high level over small spatial distances. As we move offshore into the deeper waters of the shelf, the impacts of tidal oscillations are less important. The change in water-mass optical characteristics is driven by the interactions of larger-scale physical features (i.e., mean currents) and weather patterns. These larger-scale processes tend to homogenize water masses over kilometer scales, and as a result, the GSD needed to adequately resolve the real horizontal geophysical boundaries within these homogenous waters grows in size.

In the nearshore environment (less than 10 km from the shore in this example), each of these tests yield approximately the same result (Figure 7). It would appear from this result that to adequately describe the geophysical features in the nearshore, the GSD must be < 200 m. Many important biochemical processes occur within 10 km of the shore. River discharges of nutrients and organic matter have their greatest influences in this nearshore region, and the cycling of these materials within the nearshore environment may have large impacts on estimates of the fate of biogeochemical elements (e.g., carbon, at the terrestrial or ocean boundary). In addition, the input of fresh water has its greatest impact on baroclinicity in the nearshore environment. The use of optical tracers for salinity (Coble et al., this issue) may actually improve the understanding and prediction of coastal circulation, a requirement for any study on the sources and fate of biogeochemically relevant materials. These results suggest that color studies at the LEO-15 site may require GSDs approximately 100 m to resolve biogeochemical processes from ocean-color data.

The optimal GSD is also a function of the information content of the data set. The single-band data set shows less variability in its standard deviation than the PC1 of the dual-band data set, which in turn shows less variability in the standard deviation than the PC1 of the hyperspectral data set. This effect results in lower mean and median optimal GSDs as the number of bands used in the analysis increases. This result suggests that additional bands add information that may be used to discriminate optically different water masses, and perhaps retrieve estimates of different optically active constituents. In particular, beyond 40 km, there is a real divergence between the GSDs of the PCA Hyp and those derived from the simulated SeaWiFS data set. Unsurprisingly, it also suggests that the optimal GSD for delineating the spectral variances in upwelling radiance signals may be a function of the total number of bands sampled.

A PCA reduces the dimensionality of a data set of interest by rotating the coordinate system into one that minimizes the variance across the entire data space. In the hyperspectral image cube of Figure 1, a single eigenvector (Figure 4) accounts for most of the variance in this image. While this eigenvector (PC1) may appear similar to a water-leaving radiance vector in high-scattering green waters, great care must be used in ascribing real geophysical properties to eigenvectors. Figure 4B does show why the three techniques were so similar, particularly in the nearshore region, as the wavelengths around 560 nm influenced the variance of the PC1 the most. PCA is a tool to describe scene-dependent variance, and in this paper, we focus on the information content of spectral data over small homogeneous regions of water color, and intensity across a large scene of interest. The spatial scale of homogeneous regions often depends on the total number of bands used to describe that homogeneous region, particularly when moving away from the shallow water regions impacted by high-energy mixing. It does not, however, necessarily suggest that ecological parameters of interest vary over these same scales. The determination of variance of ecological-relevant material would

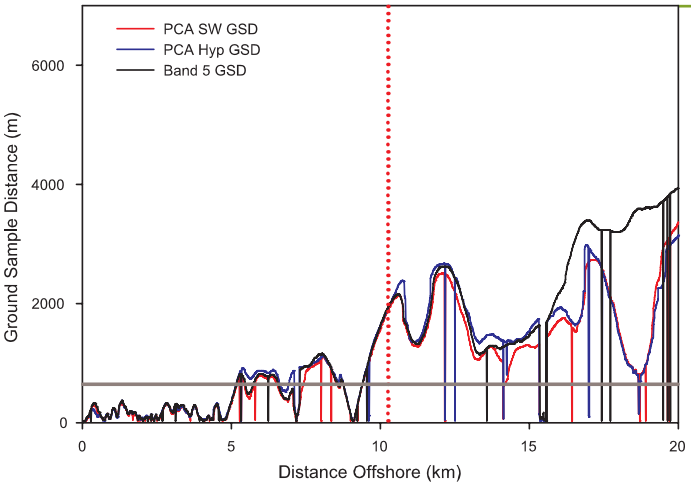

Figure 7. The inshore GSD for SW Band 5, PCA SW, and PCA Hyp. The similarity within this region of the GSD trends is striking, and suggests that variance within the inshore region may be driven by the concentrations of suspended matter. The vertical red line denotes the location of the inshore ROI from which the variance threshold for the analysis was determined. The horizontal grey line indicates the size of the region of interest from which the threshold was determined.

depend greatly on the algorithms that invert color and intensity into optically active mass constituents (i.e., chlorophyll, colored dissolved organic matter [CDOM], sediments).

Ocean-color research and applications are frequently more concerned with the products derived from $R_{rs}(\lambda)$ estimates (e.g., total absorption, total scattering, diver visibility, total chlorophyll concentration), rather than the R_{rs} data itself. It may be that the appropriate spatial sampling frequency for these products is different than the sampling frequency determined from the spatial variations in the radiance fields. However, using the products produced from these same radiance fields to determine spatial sampling frequencies may produce vary different scaling results, strictly due to the method of product generation. In this paper, we show how the noise of the sensor and atmospheric processing approximates a linear transform function, and we expect that the variance in radiance data above the background linear noise represents true difference in ocean color. However, many remote-sensing product calculations are nonlinear transforms of the R_{rs} data (e.g., Lee and Carder, 2002; O'Reilly et al., 1998). A nonlinear transform of the R_{rs} data will alter the mean and variance statistics in ways that may alter scaling results shown here.

We specifically did not include an analysis using a ratio of a simulated Band 3 to Band 5 because our primary assumption is that the sum of environmental and sensor noise is a linearly additive component of the total signal, and the standard deviation of this noise component of the signal would be constant across varying levels of intensity. If this assumption is true (and it would appear so from this analysis), then a linear addition of noise to a downward trending numerator would yield an increasing variance estimate of a ratio product. In fact, for the ROIs shown in Figure 1, the Band 3:5 ratio showed a significant difference in standard deviation, versus the similar standard deviations calculated for each band in each ROI. This violates our primary assumption, and suggests that studies using ratio analyses should attempt to delineate the differences in the variances of biogeochemical estimates that result from variance of the data versus the mathematical variance created by the application of the algorithm.

This work suggests that future studies on the optimal sampling frequency in the spatial domain of remote-sensing data begin with the R_{rs} data itself. Furthermore, the optimal sampling frequency may be a function of the total number of wavebands available for analysis. This work does not definitively suggest that variations of optically-active constituents may be retrieved at the same spatial resolution as the variation in the total R_{rs} vector. However, it does suggest that spatial variations in ocean color depend on the number of channels used to described differences between homogenous regions. If these additional channels can be used to discriminate additional biological, chemical, and physical information, then the hyperspectral ocean color signal will yield a greater ability to identify, study, and predict important ecological processes in the coastal environment.

New algorithms are being developed that focus on relatively continuous spectral data rather than on the ratio of multispectral channels. These new algorithms use a variety of techniques to take advantage of the greater degrees of freedom that the hyperspectral data stream offers to the ocean-color scientist; algorithms are in development to retrieve standard oceanographic products, such as total chlorophyll and CDOM, as well as new products such as bathymetry, bottom type, and water-column Inherent Optical Properties (IOPs) (e.g., Dierssen et al., 2003; Hoge et al., 2003; Lee et al., 1998; Lee et al., 1999; Lee and Carder, 2004; Louchard et al., 2003; Mobley et al., 2002). In addition, hyperspectral approaches may also yield information on phytoplankton speciation, which might allow for the remote identification of harmful algal blooms (HABs) (Roesler et al., in press). These hyperspectral imagery analysis techniques offer the potential to dramatically increase our ability to retrieve coastal zone information from ocean color data streams and specifically address critical issues in coastal-zone management. As we move from multispectral to hyperspectral data products, we may also find a need for higher spatial resolution data to better describe the changes in the nearshore coastal environment.

SUMMARY AND CONCLUSIONS

There are probably many methods of determining the optimal spatial sampling frequency in the coastal zone. However, when a statistical approach is used, care must be taken to use a method that is applicable to the sample, and to ensure the rigorous assumptions of stationary functions are not violated. Otherwise, unclear results are obtained that may lead to an inefficient scientific design of a remote sensing sensor or experiment, for example, for the data discussed here, the assumptions inherent to using the autocorrelation function were violated, indicating that autocorrelation analysis is the wrong tool for this coastal data set.

The results described here suggest that the spatial resolution required for offshore studies may be dependent on the spectral resolution of the data stream. At LEO-15 between 1 and 10 km, a 50- to 200-m GSD appears sufficient for single-band, dual-band, and hyperspectral-band data. Within 1 km of the shore, an even higher resolution sensor might be needed to resolve the wind and tidally impacted features. In the optically deep offshore waters of LEO-15, bottom effects do

not impact R_{rs}. However, in optically shallow areas, the spatial heterogeneity of the bottom may further reduce the GSD required to resolve the optical constituents near the coast (see Philpot et al., this issue). Offshore of 10 km, there is a significant difference in the ability to discriminate optical boundaries using the single or dual band data compared to the hyperspectral data. This suggests that hyperspectral data may be better able to delineate optically distinct regions in offshore coastal waters and that scaling studies may be dependent on the total number of spectral channels used in the analysis.

Nonlinear transforms of R_{rs}, typical of algorithms for products such as chlorophyll concentration, may alter variance calculations; optimal scaling results for these derived products may depend in part on the transform of the data, or the algorithm used. Care must be used when deriving statistics on nonlinear transforms to avoid missing real differences in ocean color and biogeophysical boundaries.

The real goal in optical oceanography is to use optics to identify interesting oceanic features as well as describe their time-dependent change. These features range from individual, HAB-forming phytoplankton species to the delineation of terrestrial outflow plumes from background coastal waters. The features may also include bottom characteristics, such as seagrass beds or coral formations. The future of hyperspectral imagery will depend on the ability to retrieve these optically distinct constituents from the $R_{rs}(\lambda)$ data streams. One of the first steps in retrieving biogeochemical information from imagery data is to resolve the spatial variations distinguishable in these hyperspectral scenes. The next step will be to focus on the development of new algorithms that use the entire water-leaving spectrum to assess and monitor the coastal zone environment.

ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research. We would like to thank Sharon DeBra and Mubin Kadiwala for their help in editing and preparing this paper. Lastly, the review and comments of Dr. Ricardo Letelier were greatly appreciated.

REFERENCES

Abbott, M.R., and R.M. Letelier, 1998: Decorrelation scales of chlorophyll as observed from bio-optical drifters in the California Current. *Deep Sea Research II*, 45, 1,639-1,667.

Chang, G.C., T.D. Dickey, O.M. Schofield, A.D. Weidemann, E. Boss, W.S. Pegau, M.A. Moline, and S.M. Glenn, 2002: Nearshore physical processes and bio-optical properties in the New York Bight. *Journal of Geophysical Research*, 107(C9), 3,133.

Chant, R.J., S. Glenn, and J. Kohut, in press: Flow reversals during upwelling conditions on the New Jersey inner shelf. *Journal of Geophysical Research*.

Chilès, J.-P., and P. Delfiner, 1999: *Geostatistics: Modeling Spatial Uncertainty*. John Wiley and Sons, Inc., New York, 720 pp.

Curran, P.J., 1988: The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment*, 24(3), 493-507.

Dale, M.R.T., P. Dixon, M.-J. Fortin, P. Legendre, D.E. Myers, and M.S. Rosenberg, 2002: Conceptual and mathematical relationships among methods for spatial analysis. *Ecography*, 25, 558-577.

Davis, C.O., J. Bowles, R.A. Leathers, D. Korwan, T.V. Downes, W.A. Snyder, W.J. Rhea, W. Chen, J. Fisher, W.P. Bissett, and R.A. Reisse, 2002: Ocean PHILLS hyperspectral imager: Design, characterization, and calibration. *Optics Express,* 10(4), 210-221.

Dickey, T., S. Zedler, X. Yu, S.C. Doney, D. Frye, H. Jannasch, D. Manov, D. Sigurdson, J.D. McNeil, L. Dobeck, T. Gilboy, C. Bravo, D.A. Siegel, and N. Nelson, 2001: Physical and biogeochemical variability from hours to years at the Bermuda Testbed Mooring site: June 1994 March 1998. *Deep Sea Research II*, 48, 2,105-2,140.

Dickey, T.D., 1991: The emergence of concurrent high-resolution physical and bio-optical measurements in the upper ocean and their applications. *Reviews of Geophysics*, 29(3), 383-413.

Dierssen, H., R. Zimmerman, R. Leathers, T. Downes, and C. Davis, 2003: Ocean-color remote sensing of seagrass and bathymetry in the Bahamas Banks by high-resolution airborne imagery. *Limnology and Oceanography*, 48(1, part 2), 444-455.

Hoge, F.E., P.E. Lyon, C.D. Mobley, and L.K. Sundman, 2003: Radiative transfer equation inversion: Theory and shape factor models for retrieval of oceanic inherent optical properties. *Journal of Geophysical Research*, 108(C12), 3,386.

Kohler, D.D.R., W.P. Bissett, C.O. Davis, J. Bowles, D. Dye, J. Britt, J. Bailey, R. Steward, O.M. Schofield, M. Moline, S. Glenn, and C. Orrico, 2002a: Characterization and calibration of a hyperspectral coastal ocean remote sensing instrument. In: Proceedings of the AGU/ASLO Ocean Sciences 2002 Meeting, Honolulu, HI, February 11-15. American Geophysical Union, Washington, D.C.

Kohler, D.D., W.P. Bissett, C.O. Davis, J. Bowles, D. Dye, R.G. Steward, J. Britt, M. Montes, O. Schofield, and M. Moline, 2002b: High resolution hyperspectral remote sensing over oceanographic scales at the LEO 15 field site. In: *Ocean Optics XVI*, conference held in Santa Fe, NM, November 18-22. Office of Naval Research, Washington, D.C.

Lee, Z., K.L. Carder, C.D. Mobley, R.G. Steward, and J.S. Patch, 1998: Hyperspectral remote sensing for shallow waters: 1. A semianalytical model. *Applied Optics*, 37(27), 6,329-6,338.

Lee, Z., K.L. Carder, C.D. Mobley, R.G. Steward, and J.S. Patch, 1999: Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization. *Applied Optics*, 38(18), 3,831-3,843.

Lee, Z.P., and K.L. Carder, 2002: Effects of spectral band numbers on the retrieval of water column and bottom properties from ocean color data. *Applied Optics*, 41(12), 2,191-2,201.

Lee, Z.P., and K.L. Carder, 2004: Absorption spectrum of phytoplankton pigments derived from hyperspectral remote-sensing reflectance. *Remote Sensing of Environment*, 89, 361-368.

Louchard, E., R. Reid, F. Stephens, C. Davis, R. Leathers, and T. Downes, 2003: Optical remote sensing of benthic habitats and bathymetry in coastal environments at Lee Stocking Island, Bahamas: A comparative spectral classification approach. *Limnol. Oceanogr.*, 48(1, part 2), 511-521.

Mahadevan, A., and J.W. Campbell, 2002: Biogeochemical patchiness at the sea surface. *Geophysical Research Letters*, 29(19), 1,926.

Mahadevan, A., and J.W. Campbell, eds., in press: Biogeochemical variability at the sea surface: How it is linked to process response times. Pp. 215-226 in *Handbook of Scaling Methods in Aquatic Ecology: Measurement, Analysis, Simulation*. CRC Press, Boca Raton, FL.

Mobley, C.D., 1994: *Light and Water*. Academic Press, San Diego, CA, 592 pp.

Mobley, C.D., L. Sundman, C.O. Davis, M. Montes, and W.P. Bissett, 2002: A look-up-table approach to inverting remotely sensed ocean color data. In: *Ocean Optics XVI*, conference held in Santa Fe, NM, November 18-22. Office of Naval Research, Washington, D.C.

Mueller, J.L., 1976: Ocean color spectra measured off the Oregon coast: Characteristic vectors. *Applied Optics*, 15(2), 394-402.

O'Reilly, J.E., S. Maritorena, B.G. Mitchell, D.A. Siegel, K.L. Carder, S.A. Garver, M. Kahru, and C. McClain, 1998: Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical Research*, 103, 24,937-24,953.

Preisendorfer, R.W., 1976. *Hydrologic Optics*, Volumes 1-6, National Oceanic and Atmospheric Administration, Honolulu, Hawaii. NTIS PB-259 793. National Technology Information Service, Springfield, VA, 1,727 pp.

Roesler, C.S., S.M. Etheridge, and G.C. Pitcher, in press: Application of an Ocean Color Algal Taxa Detection Model to Red Tides in the Southern Benguela. In: Proceedings of the Tenth International Conference for Harmful Algal Blooms.

Strub, P.T., and C. James, 2000: Altimeter-derived variability of surface velocities in the California Current System: 2. Seasonal circulation and eddy statistics. *Deep Sea Research II*, 47, 831-870.

Yoder, J.A., C.R. McClain, J.O. Blanton, and L.Y. Oey, 1987: Spatial scales in CZCS-chlorophyll imagery of the southeastern U.S. continental shelf. *Limnol. Oceanogr.*, 32(4), 929-941.