NASA Technical Memorandum 104566, Vol. 32

SeaWiFS Technical Report Series

Stanford B. Hooker, Elaine R. Firestone, and James G. Acker, Editors

Volume 32, Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms

Janet W. Campbell, John M. Blaisdell, and Michael Darzi

August 1995

NASA Technical Memorandum 104566, Vol. 32

SeaWiFS Technical Report Series

Stanford B. Hooker, Editor
NASA Goddard Space Flight Center
Greenbelt, Maryland

Elaine R. Firestone, Technical Editor
General Sciences Corporation
Laurel, Maryland

James G. Acker, Technical Editor
Hughes STX
Lanham, Maryland

Volume 32, Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms

Janet W. Campbell
University of New Hampshire
Durham, New Hampshire

John M. Blaisdell and Michael Darzi
General Sciences Corporation
Laurel, Maryland

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

1995

J.W. Campbell, J.M. Blaisdell, and M. Darzi

Abstract

The level-3 data products from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) are statistical data sets derived from level-2 data. Each data set will be based on a fixed global grid of equal-area bins that are approximately $9 \times 9\,\mathrm{km}^2$. Statistics available for each bin include the sum and sum of squares of the natural logarithm of derived level-2 geophysical variables where sums are accumulated over a binning period. Operationally, products with binning periods of 1 day, 8 days, 1 month, and 1 year will be produced and archived. From these accumulated values and for each bin, estimates of the mean, standard deviation, median, and mode may be derived for each geophysical variable. This report contains two major parts: the first (Section 2) is intended as a users' guide for level-3 SeaWiFS data products. It contains an overview of level-0 to level-3 data processing, a discussion of important statistical considerations when using level-3 data, and details of how to use the level-3 data. The second part (Section 3) presents a comparative statistical study of several binning algorithms based on CZCS and moored fluorometer data. The operational binning algorithms were selected based on the results of this study.

1. INTRODUCTION

The level-3 data processing stage is the first stage in which data from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) are spatially and temporally averaged. Prior to this stage, a standard set of geophysical variables will be derived for individual pixels. These level-2 variables include chlorophyll concentration, a diffuse attenuation coefficient, and water-leaving radiances in the visible bands of SeaWiFS.

In generating level-3 data products, pixels containing valid level-2 data will be mapped to a fixed spatial grid whose resolution elements are $9 \times 9\,\mathrm{km}^2$. These square grid elements or *bins* are arranged in rows beginning at the South Pole. Each row begins at 180° longitude and circumscribes the Earth at a given latitude. There are 5,940,422 bins for each level-3 data set. Within each bin, statistics will be accumulated for time periods of 1 day, 8 days (often referred to as the *weekly product*), 1 month, and 1 year. There will be a global level-3 data product archived for each day, 8-day period, calendar month, and calendar year of the SeaWiFS mission.

The level-3 data products may be used to derive the mean, standard deviation, and other statistical measures for the standard level-2 variables, and for certain other variables, such as primary productivity, which are functions of level-2 variables. The Coastal Zone Color Scanner (CZCS) North Atlantic monthly composite chlorophyll images (Esaias et al. 1986 and Feldman et al. 1989) are examples of monthly means derived from level-3 CZCS data.

The purpose of binning data is to create reduced-volume data sets appropriate for use in climate and basin-scale biogeochemical models. By averaging data over time periods of several days or longer, problems of missing data can be overcome. Although temporal and spatial resolutions are reduced, compared with the level-2 data, the resulting smoothed level-3 means are effective in depicting seasonal patterns on regional and basin scales.

There are important statistical considerations that involve the use of level-3 data. Users should be aware of these considerations, especially in situations where level-3 data are used in models to derive other variables. For example, to use a mean chlorophyll concentration (level-3 variable) in an algorithm to derive mean primary productivity might result in significantly biased results. Recommended procedures for using level-3 variables in models are presented in this report.

The remainder of this report is divided into two parts. The first part (Section 2) is intended to serve as a guide for users of level-3 data products. Section 2.1 is an overview of the processing from level-0 to level-3. Section 2.2 contains a discussion of the important statistical considerations involved in using level-3 data, and Section 2.3 provides the equations to be used to compute the mean, standard deviation, median, and mode of each level-3 variable. Equations for computing statistics of *level-4* variables, derived from level-3 variables, are given in Section 2.4.

The second part (Section 3) documents a statistical study based on CZCS data and moored fluorometer data which compared alternative binning algorithms. Results of this study were the basis for the selection of the binning algorithm used. Three color plates compare the results of alternative binning algorithms applied to seven representative CZCS scenes.

In addition, there are three appendices providing details for statisticians and programmers who may wish to write codes to bin data. Appendix A explains the procedure used for mapping pixels to bins based on the center latitude and longitude of the pixel, and for determining the latitude and longitude coordinates of a bin. Appendix B contains details of the weighting scheme used for weighting data from different orbits (times). Appendix C contains three pseudocodes that reveal how data are accumulated spatially (Space Binner Code), temporally (Time Binner Code) and how means, standard deviations, and other statistics are calculated from the binned data (Bin Data Interpreter Code).

2. USERS' GUIDE

2.1 Overview of Data Processing

As the name would suggest, the level of a data product refers to the amount of processing that has been applied to the data. Certain conventions have been adopted to describe the major levels of processing.

2.1.1 Level-0 Data

Data recorded on board the satellite and subsequently broadcast to ground receiving stations are called level-0 data. Data broadcast directly (without being recorded) are also considered level-0 data. The recorded data provide either local area coverage (LAC) or global area coverage (GAC). This classification refers to the spatial resolution of the data. In SeaWiFS LAC data, the spatial resolution is 1.1 km at nadir (directly beneath the satellite), and pixels are contiguous.

The GAC data are comprised of individual pixels having the same spatial resolution as LAC data (1.1 km), but the pixels are spaced at 4.4 km intervals. The GAC data are created on board the satellite by selecting every fourth pixel on every fourth scan line. This subsampling reduces the volume of data required to provide global coverage. A comparative study of alternative GAC sampling algorithms was reported by McClain et al. (1992).

Only a limited amount of LAC data will be recorded on board SeaWiFS. However, LAC data will be continuously broadcast as high-resolution picture transmission (HRPT) data to sites around the world which operate licensed ground-receiving stations. All HRPT data will be LAC data.

2.1.2 Level-1a Data

The level-1a products include the raw image data and all instrument and spacecraft telemetry, as in the level-0 data, together with appended instrument calibration and navigation data. In addition, instrument telemetry and selected spacecraft telemetry are reformatted and also appended.

Approximately 40 minutes of contiguous level-1 data are produced on the daylight portion of each orbit. Operationally, this 40-minute *swath* may be subdivided into two or more level-1 *scenes*. The division may occur when the sensor tilt is changed, i.e., so each scene would nominally have a constant sensor tilt, or other criteria, e.g., maximum scan lines per scene, may dictate further subdivisions of the swath.

The level-1a data can be used to calculate calibrated radiances in units of $\mathrm{W\,m^{-2}\,\mu m^{-1}\,sr^{-1}}$ in the 8 spectral bands of SeaWiFS. This radiance received at the satellite altitude is solar radiation backscattered from the Earth's atmosphere, ocean, clouds and land. Water-leaving radiance (the signal of interest) usually comprises less than 10% of the total signal.

2.1.3 Level-2 Data

Geophysical properties of the ocean and atmosphere derived from level-1a data are considered level-2 data. Level-2 data correspond to the original pixel positions; there is no remapping. Each level-2 scene corresponds to a level-1 scene and vice versa; there is no change in the geographical coverage of each scene for operational products.

Before computing level-2 data, pixels are eliminated if they contain clouds, sun glint, or other abnormalities. For pixels that pass these screens, an atmospheric correction algorithm (Gordon et al. 1983 and Gordon and Castaño 1987) is applied to subtract the atmospheric scattering components from the total radiance, and thus derive the water-leaving radiances in bands 1–5. Then, bio-optical algorithms (Clark 1981 and Gordon and Morel 1983) are applied to the water-leaving radiances to derive in-water properties.

Standard variables currently planned for computation are:

$L_{WN}(\lambda_i)$ normalized water-leaving radiances in the bands $i = 1$–5,

$L_a(\lambda_i)$ atmospheric aerosol radiances in the bands $i = 6$–8,

$\tau_a(865)$ aerosol optical thickness at 865 nm (band 8),

PIG CZCS-like pigment concentration ($\mathrm{mg\,m^{-3}}$),

CHL chlorophyll *a* concentration ($\mathrm{mg\,m^{-3}}$), and

K_{490} diffuse attenuation coefficient at 490 nm ($\mathrm{m^{-1}}$).

2.1.4 Level-3 Data

The level-3 data are statistical data products derived by binning level-2 GAC data. This is the first stage at which data are both spatially and temporally averaged. A level-3 product will be produced for each day, 8-day period (*week*), calendar month, and calendar year of the SeaWiFS mission. The 8-day periods are started from the first day of each calendar year. Thus, there will be 46 *weeks* per calendar year, with the last *week* having only 5 or 6 days instead of 8.

Each data product will contain statistics derived by mapping level-2 data to a fixed global grid whose resolution elements (called *bins*) are approximately $9 \times 9\,\mathrm{km}^2$. The bins are arranged in rows beginning at 180° longitude and circumscribing the Earth eastward at a given latitude. There are 5,940,422 bins for each level-3 data product. Appendix A contains details related to the gridding scheme, and the precise areal coverage and geographic location of each bin.

Statistical data provided with the level-3 data products will allow users to calculate the mean, standard deviation, median, and mode for each level-2 variable listed above. The procedures are described in Section 2.3, and pseudocodes for programming implementation are detailed in Appendix C.

In addition to level-2 variables, statistical data will also be provided for the ratio:

$$\mathrm{IC_K} \equiv \frac{\mathrm{CHL}}{K_{490}} \qquad (1)$$

calculated at each pixel in the level-2 data set (but not saved as a level-2 variable). This ratio, which appears in several primary productivity algorithms (Balch et al. 1992, Platt and Sathyendranath 1988, Eppley et al. 1985, Smith and Baker 1978, and Bannister 1974), may be regarded as the integral chlorophyll (units of $\mathrm{mg\,m^{-2}}$) integrated over the upper optical depth. The rationale for including this as a level-3 variable will be presented in Section 2.2.

In addition to the level-3 data products, a number of standard level-3 image products will be produced. These will include standard mapped images, which are equirectangular projections of means derived from the level-3 statistical data, and reduced resolution images intended for browsing purposes.

2.1.5 Level-4 Data

In this report, variables derived from level-3 data will be called *level-4* variables. It is anticipated that level-3 data will be used as input to biogeochemical models where the goal of the modeling is to estimate global fluxes of key elements such as carbon and nitrogen. In such applications, it is important that the level-4 variable represent a spatial-temporal mean, e.g., the average daily, weekly, or monthly carbon flux. The practice of substituting means into models to produce spatial-temporal means can result in significantly biased results. This will be discussed further in Section 2.2.

The methods used to produce the level-3 SeaWiFS data have been designed to overcome this problem for a large class of level-4 variables. Procedures for computing unbiased estimates of the mean of level-4 variables will be discussed in detail in the following sections.

2.2 Statistical Considerations

The question of how to bin SeaWiFS data revolved around certain statistical issues. Many of the issues or questions raised had come to light through the experience of binning CZCS data into daily, monthly, and yearly composites. There were several proposed ways to *average* data, and results would be significantly different depending on the method chosen. It was further recognized that the choice of method should depend on how level-3 SeaWiFS data are to be used. The practice of using level-3 means in equations to derive *level-4* means was inappropriate, and, therefore, this issue had to be addressed as well.

Following is a discussion of four major issues and the summary of the decisions related to each. In many instances, decisions were based on a statistical analysis of CZCS data and moored fluorometer time-series data. The results of the statistical study are presented in Section 3. The four issues were:

1. Should statistics be computed for CHL or for log(CHL)? What about other level-2 variables?
2. What is the best method for estimating level-4 variables?
3. What statistics should be saved for each sampling domain?
4. Should the temporal statistics give equal weight to all data falling within the sampling domain? Or, should some accommodation be made to compensate for the uneven temporal distribution of data?

2.2.1 CHL vs. log(CHL) Statistics

Chlorophyll measurements tend to be lognormally distributed, i.e., log(CHL) is normally distributed, in large data sets of satellite or ship data (Fig. 1). Lognormal distributions occur commonly in biological processes where the rate of change of a variable is proportional to its size (Aitchison and Brown 1957 and Crow and Shimizu 1988). One of the first issues addressed, therefore, was whether or not statistics should be computed for CHL or for log(CHL). The same question was also addressed for other variables.

It is fairly common practice to log-transform CHL measurements before using them in other derivations. For example, Chelton and Schlax (1991) used log-transformed data in comparing time averages of chlorophyll data. The CZCS pigment algorithm was derived by a linear regression of log(CHL) versus log-transformed radiance ratios, and CZCS pigment images are usually scaled according to the logarithm of pigment. The mean derived by first averaging log-transformed data and then inverting the transform is the geometric mean. Is the geometric mean preferable to the arithmetic mean?

It was agreed at the outset that the arithmetic mean is the appropriate mean for most biogeochemical applications. The mean chlorophyll concentration, for example, represents the mean biomass per unit volume which will subsequently be multiplied by total volume (depth × area) to estimate regional or global biomass. However, the sample mean derived from small samples might be a poor estimator of the true population mean.

Let X be a lognormally distributed variable (Fig. 2), and let $\overline{X}$ denote the true mean of X within a sampling domain. In the context of the SeaWiFS data processing,

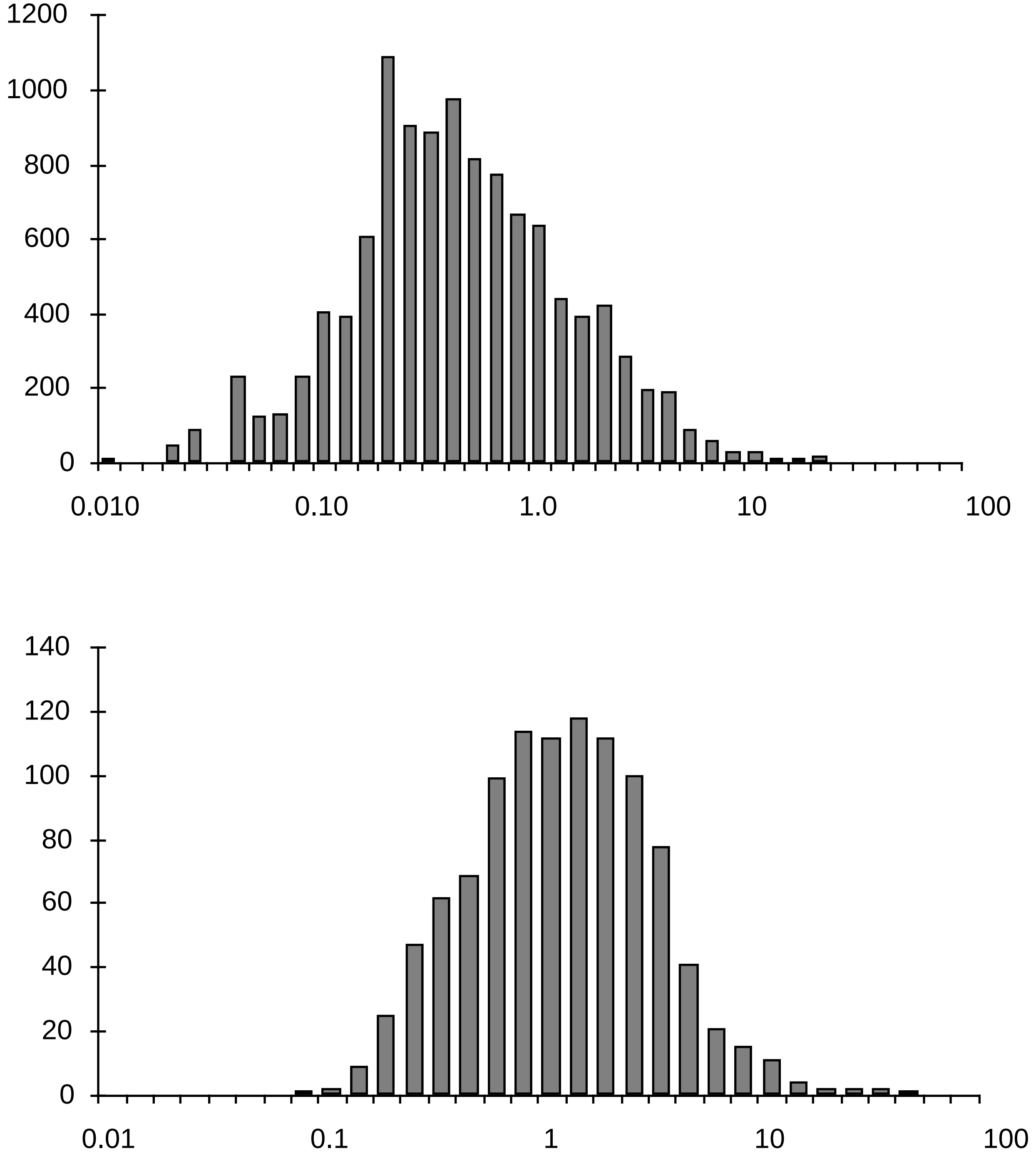


Fig. 1. Histograms of chlorophyll concentration derived from *in situ* measurements. The top panel displays 11,176 measurements from the world ocean collected by C.S. Yentsch, 1956–86. The bottom panel displays 1,047 surface measurements from the northwest Atlantic continental shelf, Marine Resources Monitoring, Assessment, and Prediction (MARMAP), 1978–82. (Campbell and O'Reilly 1988)

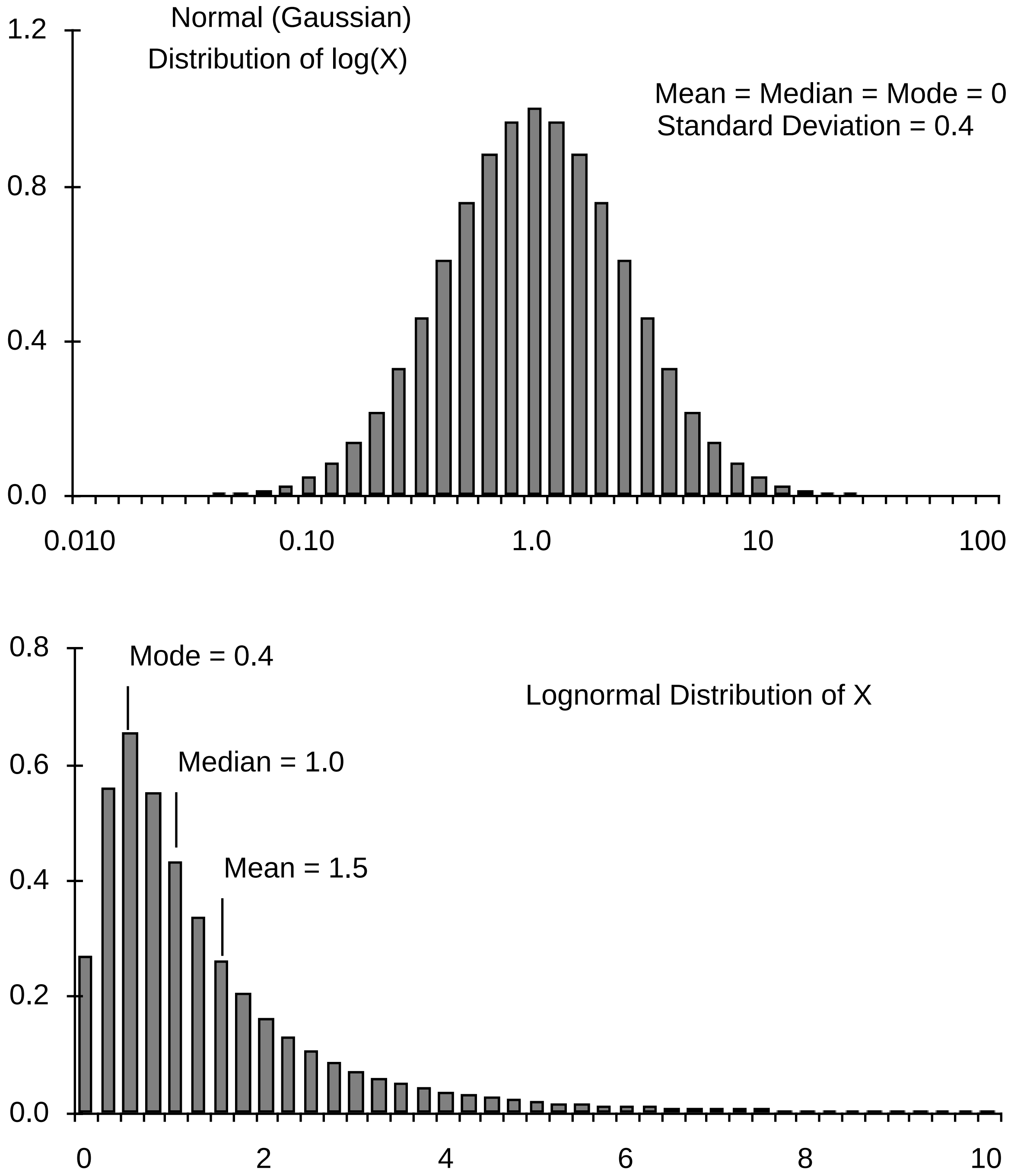


Fig. 2. The lognormal distribution: The top panel displays a histogram of log(X), where log(X) is normally distributed with mean 0 and standard deviation 0.4. The bottom panel shows the corresponding histogram of the lognormal variable, X.

"sampling domain" refers to a specific bin and averaging period; X is the level-2 variable, and $\overline{X}$ its level-3 equivalent. The question is: what is the best method for estimating $\overline{X}$ given a sample of n measurements (pixels): $X_1, \ldots, X_n$?

In the case of a lognormal distribution, the sample mean (or arithmetic average):

$$\overline{X}_{\text{avg}} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2}$$

tends to underestimate the true population mean when sample sizes are small (Baker and Gibson 1987). The higher the variance of the underlying distribution, the more this is true. The reason for this is that small samples tend to miss high values which occur much less frequently than low values. However, the high values have a significant influence on the mean of the distribution. For example, much of the biological production in the ocean occurs in localized areas such as upwelling zones, and in transient blooms of relatively short duration. A sample that misses these areas and blooms would significantly underestimate global or regional production.

Sample sizes involved in binning GAC data will be small. Since the GAC data have a 4 km spacing between pixels, at most 9 pixels from a single orbital pass can fall into an 9 × 9 km bin. The average sample size will be closer to four in data sets derived from a single orbital pass. Although sample sizes will increase with longer averaging periods, the variance will also increase. Thus, there was concern that small sample sizes and large variances might make the arithmetic average a poor estimator for level-3 means.

The practice of transforming data first, computing the mean, m_x, of log-transformed data

$$m_x \;=\; \frac{1}{n}\sum_{i=1}^{n} \ln\bigl(X_i\bigr) \tag{3}$$

and then estimating the mean of X as

$$\overline{X}_{\text{geom}} \;=\; e^{m_x} \tag{4}$$

gives the geometric mean. In the case of a lognormal variable, the geometric mean is the median of the distribution. For any distribution that is positively skewed, the geometric mean will underestimate the population mean.

Studies have shown that the maximum likelihood estimator for a lognormal mean

$$\overline{X}_{\text{mle}} = e^{\left(m_x + \frac{1}{2}s_x^2\right)} \tag{5}$$

performs better than either of the other two when variances are large and sample sizes small (Baker and Gibson 1987). In (5), m_x is the sample mean of $\ln(X)$, given by (3), and s_x^2 is the sample variance given by

$$s_x^2 \;=\; \frac{1}{n}\sum_{i=1}^{n}\Bigl[\ln\bigl(X_i\bigr) - m_x\Bigr]^2. \tag{6}$$

Note that this is not the more commonly used unbiased estimator which uses a divisor of $n-1$ instead of n. However, this is the maximum likelihood estimator for the variance of a normal random variable. In order for (5) to be the maximum likelihood estimator for $\overline{X}$, m_x and s_x^2 must be maximum likelihood estimators for the mean and variance of $\ln(X)$ (Crow and Shimizu 1988).

In the statistical study presented in Section 3, the three estimators, $\overline{X}_{\text{avg}}$, $\overline{X}_{\text{geom}}$, and $\overline{X}_{\text{mle}}$, were compared using CZCS data and a time series of moored fluorometer data (Medeiros and Wirick 1992). Results obtained for both time and space averages were:

1. The sample mean, $\overline{X}_{\text{avg}}$ (2), and the maximum likelihood estimator, $\overline{X}_{\text{mle}}$ (5), gave equivalent results.
2. The geometric mean or median, $\overline{X}_{\text{geom}}$ (4), was systematically less than the other two.

The same results were obtained for other standard CZCS variables: K_{490} and normalized water-leaving radiances $L_{WN}(\lambda_i)$. Thus, based on their performance as estimators of the mean, $\overline{X}_{\text{avg}}$ and $\overline{X}_{\text{mle}}$ were regarded as acceptable estimators for the true population mean, $\overline{X}$.

2.2.2 Estimating Level-4 Variables

It is not possible to prescribe a general method for estimating level-4 variables. The appropriate method will depend on the nature of the relationship involved, i.e., whether it is linear or nonlinear, and the form it takes.

Let $Y = f(X)$ be a relationship that defines the variable Y as a function of the level-2 variable X, and let $\overline{Y}$ be the level-3 equivalent of Y. That is, $\overline{Y}$ represents the true mean of Y within a sampling domain. In general, X may be a vector of level-2 variables, i.e., Y may be a function of more than one level-2 variable.

The problem that motivates this issue is that $\overline{Y}$ is not, in general, equal to $f\bigl(\overline{X}\bigr)$. *Substitution of the mean of X into the function is only legitimate for linear functions. In general, the mean of a function of several variables is not equal to the function of the means.*

For any general function, the only way to obtain an accurate estimate of the true mean, $\overline{Y}$, would be to compute $Y_i = f(X_i)$ at each pixel in the level-2 data, and then determine its average using either the arithmetic average, $\overline{Y}_{\text{avg}}$, or the maximum likelihood estimate, $\overline{Y}_{\text{mle}}$. In this case, the function $Y = f(X)$ would be a level-3 variable computed by averaging over pixels in the level-2 data. An example is IC_K (1) which will be computed in this way.

It is not possible or practical to anticipate the many functions or mathematical relationships that may be applied to SeaWiFS data. Thus, there needed to be guidelines and methods for using level-3 data to obtain accurate estimates of the mean of level-4 variables.

The decision was made to use the maximum likelihood estimation (MLE) method instead of the more common arithmetic average (AVG) method because the MLE method provides a way to estimate the mean (and other statistics) for a large class of level-4 variables of the form

$$Y = AX^B \tag{7}$$

where A and B are constants, and X is a single variable, i.e., not a vector.

For variables in this class, $\ln(Y)$ is linearly related to $\ln(X)$

$$\ln(Y) = \ln(A) + B\ln(X). \tag{8}$$

Therefore, the mean and variance of $\ln(Y)$ can be estimated as

$$m_y = \ln(A) + Bm_x \tag{9}$$

and

$$s_y^2 = B^2 s_x^2 \tag{10}$$

where m_x and s_x^2 are the mean and variance of $\ln(X)$ derived from the level-3 statistics saved for X.

According to the MLE method, the mean of Y is then given by

$$\overline{Y}_{\text{mle}} = e^{\left(m_y + \frac{1}{2}s_y^2\right)}. \tag{11}$$

It should be noted that if (5) proves to be an accurate estimator for the mean of X, then (11) will be an accurate estimator for the mean of Y. There is no loss of accuracy since (8)–(10) are exact relationships (not approximations).

The procedures for estimating the variance and other statistics of level-3 and level-4 variables are described in more detail in Sections 2.3 and 2.4 and in Appendix C. The equations used are based on MLE methods for estimating parameters of a lognormal distribution, and hence, they are referred to as *MLE estimators.* As will be shown, the MLE estimator is a robust estimator for the mean. That is, it generally performs well even when the underlying distribution is not lognormal. Indeed, the MLE method was not selected on the basis of an assumed lognormal distribution, but because it performed well compared with the arithmetic average (AVG estimator), and because it provided a method for estimating the mean of level-4 variables of the form given by (7).

An example of such a function is the euphotic depth, which is commonly defined as the 1% light-penetration depth (Kirk 1983). Using the level-2 variable K_{490} and applying Beer's Law, this depth may be defined as

$$Z_e \equiv -\frac{\ln(0.01)}{K_{490}} \tag{12}$$

which represents the 1% light-penetration depth at $\lambda = 490$ nm. If the mean of K_{490} based on level-3 data is used to estimate the mean euphotic depth, this will yield a biased estimate of the mean euphotic depth. However, the MLE method allows for an accurate estimate of the mean Z_e based on the saved statistics of $\ln(K_{490})$.

The equations proposed by Morel and Berthon (1989) for deriving integral euphotic chlorophyll, $\langle\text{Chl}\rangle_{\text{tot}}$, from satellite-derived chlorophyll (or pigment) also take the form of (7). Several algorithms for estimating integral productivity (Smith et al. 1982, Platt 1986, and Morel and Berthon 1989) involve the product of $\langle\text{Chl}\rangle_{\text{tot}}$ and photosynthetically available radiation (PAR) at the surface, PAR(0). The mean of this product can be derived as the product of the means of $\langle\text{Chl}\rangle_{\text{tot}}$ and PAR(0) since the two variables are uncorrelated. Thus, these algorithms may be applied to level-3 data using the saved statistics of standard level-2 variables.

2.2.3 Statistics Saved for Each Domain

Another issue that was raised concerned the choice of statistics to save for each sampling domain. Given that $\overline{X}_{\text{mle}}$ (5) is to be used for estimating the mean of the level-2 data in each domain, the statistics saved must include the sum and sum of squares of the natural logarithm of each variable. In addition, counts of the number of pixels contributing to the sums and similar ancillary information should also be saved.

Beyond this, further questions regarding what statistics to save are motivated by the concern expressed earlier as to how level-4 variables will be estimated. Two alternatives exist: either a) sufficient information is provided in the level-3 data to allow estimation of these variables using saved statistics of other variables or b) the variables should be computed at each pixel of level-2 data and their statistics saved as part of the level-3 data set. The latter is more costly from the standpoint of the storage required to add additional level-3 variables. As stated earlier, the MLE method permits the former choice for variables of the form given in (7).

There are other level-4 variables which cannot be calculated using only the saved statistics of the standard level-2 variables. Any variable that is a function of two or more level-2 variables would require additional information on the covariances between level-2 variables. An example of this is the variable IC_K (1) which appears in several primary productivity algorithms (Balch et al. 1992, Platt and Sathyendranath 1988, Eppley et al. 1985, Smith and Baker 1978, and Bannister 1974). To apply the MLE method, one must estimate the mean and variance of the natural logarithm of IC_K

$$\ln\left(\text{IC}_\text{K}\right) = \ln\left(\text{CHL}\right) - \ln\left(K_{490}\right). \tag{13}$$

The mean of $\ln\left(\text{IC}_\text{K}\right)$ is simply the difference between the means of $\ln(\text{CHL})$ and $\ln\left(K_{490}\right)$, but the variance of

$\ln\bigl(\mathrm{IC_K}\bigr)$:

$$\begin{aligned}\mathrm{var}\Bigl[\ln\bigl(\mathrm{IC_K}\bigr)\Bigr] &= \mathrm{var}\Bigl[\ln\bigl(\mathrm{CHL}\bigr)\Bigr] + \mathrm{var}\Bigl[\ln\bigl(K_{490}\bigr)\Bigr] \\ &\quad - 2\,\mathrm{cov}\Bigl[\ln\bigl(\mathrm{CHL}\bigr), \ln\bigl(K_{490}\bigr)\Bigr]\end{aligned} \tag{14}$$

involves the covariance, denoted by "cov" in (14), between ln(CHL) and $\ln\bigl(K_{490}\bigr)$, as well as their variances. The CZCS algorithms for CHL and K_{490} in Section 3 resulted in a nonlinear relationship between ln(CHL) and $\ln\bigl(K_{490}\bigr)$. Thus, their covariance varied from sample to sample. For this reason, it was decided to compute the variable $\mathrm{IC_K}$ (1) at each pixel in the level-2 data and save statistics of $\ln\bigl(\mathrm{IC_K}\bigr)$ as part of the level-3 data.

2.2.4 Weighting of Temporal Statistics

After each level-2 scene is generated, valid level-2 data from individual pixels will be binned. Sums and sums of squares accumulated at this stage are called spatial statistics, i.e., no temporal averaging is involved since data from the same scene are regarded as simultaneous. Spatial statistics from the same day will be combined into daily products, from the same 8-day period into *weekly* products, and so forth. The daily, weekly, monthly, and longer-term products will become the level-3 data, and the spatial statistics pertaining to individual scenes will be discarded.

On a given day, there may be two sets of spatial statistics for the same bin. Two sets might occur within the same orbit on different tilt segments, i.e., before and after a change in the sensor's tilt, or from different orbits in high-latitude areas where swaths overlap. In the case of two sets from the same orbit, only one set will be used. The set having the better sun-target viewing geometry will be selected. However, two sets of spatial statistics from different orbits will receive the same treatment as spatial statistics from different days. The same algorithms, called *temporal binning algorithms*, will be used to combine data separated by time gaps regardless of the size of the time gap.

Let N be the number of sets of spatial statistics (orbits) contributing to a temporal mean; let t_i be the time at which the ith set was acquired; and let n_i be the number of pixels contributing to the ith set, where $i = 1, \ldots, N$. In considering the temporal binning algorithms, a major concern was the fact that the times are unevenly distributed, and that the sample size (hence precision) varies from one time to another. Samples sizes will vary between 1 and 9, depending on where the bin lies relative to the ground track. Time gaps occur because of clouds, sunglint, and other factors.

The methods used to compensate for unevenly distributed data generally involve a scheme for weighting data. The alternative is to use simple composite statistics (unweighted data), which was the method used to create level-3 CZCS data such as the North Atlantic monthly composites (Feldman et al. 1989 and Esaias et al. 1986). These monthly composites have served as useful products for a number of scientific investigations (Campbell and Aarup 1992, Yentsch 1990, and Lewis et al. 1988), but some of the spatial patchiness in these data sets is an artifact of the uneven temporal distribution of data.

Chelton and Schlax (1991) have made a strong case for the superiority of optimal interpolation methods as compared to simple composite averages for deriving temporal means of irregularly spaced data. Such methods, known as *kriging* in the geostatistics literature (Journal 1989), require the use of correlation functions which must be determined a priori. When applied to satellite data, the methods could require both temporal and spatial correlation functions.

The advantage of optimal interpolation methods is that they allow estimates to be based on data that lie outside the domain (bin and time interval) being estimated. The disadvantage is their computational complexity. Data must be *deseasonalized* before applying the optimal interpolation method. That is, seasonal trends must be estimated and subtracted from the data. Therefore, at least a year of data must be collected before optimal interpolation methods can be applied. This is not compatible with the plan to generate level-3 data products along with the level-2 data processing.

It was decided not to apply optimal interpolation methods in the level-3 binning process. However, the binned statistics will be useful in applying optimal interpolation methods during post-processing. As an example, daily composite statistics might be used in deriving weekly and monthly means using optimal interpolation methods.

The question was, therefore, whether to use simple composite statistics (all data within a given domain are given equal weight) or to develop a weighting scheme that could be implemented easily at the time the level-2 data are processed. In general, a decision to use weighted versus unweighted statistics should depend on the distribution of the data *vis-a-vis* any trends that might exist. Simple unweighted statistics are recommended in the case where there is no trend (either spatial or temporal), or where the trend is impractical to estimate. The latter is the case for the spatial statistics. These will be unweighted sums and sums of squares of the pixels falling within each bin because it is impractical to estimate spatial trends for each bin.

In the case of weekly and monthly statistics, there may be significant trends that call for weighted sums. If simple composite (unweighted) statistics are used, each of the N sets of spatial statistics will, in effect, be weighted by its sample size, n_i. Thus, for example, a data set having $n_i = 9$ would be much more heavily weighted than one with $n_i = 1$. Trends may be lost in this process. Alternatively, a temporal mean might be calculated as the average of N spatial means, regardless of the number of pixels contributing to the spatial means. However, this would give too much weight to a data set with $n_i = 1$ compared to one with $n_i = 9$. This concern reflects the belief that precision is a function of sample size.

As a compromise to these two alternative approaches, it was decided to apply a weight of $\sqrt{n_i}$ to the spatial mean at time t_i, where n_i is the number of pixels falling in the bin at time t_i. This is effected by applying the weight

$$w_i = \frac{1}{\sqrt{n_i}} \qquad (15)$$

to the sums and sums of squares associated with the spatial statistics for time t_i. Details of the weighting scheme are given in Appendix B.

2.3 Protocols for Level-3 Statistics

The level-3 data products available for each day, *week*, month, and year of the SeaWiFS mission will allow users to compute the mean, standard deviation, median, and mode of each level-3 variable in each bin. The level-3 variables consist of level-2 variables, and in addition, the variable $\mathrm{IC_K}$ (1).

For each level-3 variable X, the level-3 data consists of a pair of sums for each bin

$$S_1 = \sum_{i=1}^{N} \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \ln\bigl(X_{ij}\bigr) \qquad (16)$$

and

$$S_2 = \sum_{i=1}^{N} \frac{1}{\sqrt{n_i}} \sum_{j=1}^{n_i} \Bigl[\ln\bigl(X_{ij}\bigr)\Bigr]^2 \qquad (17)$$

where X_{ij} is the jth observation of X at time t_i. Each observation corresponds to a pixel in the level-2 data. The number n_i is the number of pixels at time t_i containing valid level-2 data.

In addition, the following statistics are saved for each bin:

b bin index number (range: $1, \ldots, 5{,}940{,}422$),
N total number of orbits contributing data,
n total number of pixels contributing data, and
W sum of weights.

For the latter two quantities, their formulation is as follows:

$$n = \sum_{i=1}^{N} n_i \qquad (18)$$

and

$$W = \sum_{i=1}^{N} \sqrt{n_i}. \qquad (19)$$

In addition to the above variables, there will be a 16-bit time distribution variable T whose bits indicate whether data were available (bit = 1) or absent (bit = 0) in time intervals (days, two-day intervals, or months) covered by the averaging period. That is, each bit of the 16-bit number represents a time interval within the averaging period, and if a bit is set to 1, it indicates data were available during that interval.

2.3.1 The Mean and Variance of ln(X)

To estimate statistics for the variable X, the first step is to calculate the mean and variance of $\ln(X)$. These are given by

$$m_x = \frac{S_1}{W} \qquad (20)$$

and

$$s_x^2 = \frac{S_2}{W} - m_x^2. \qquad (21)$$

2.3.2 The Mean and Other Statistics of X

The mean of X is estimated by

$$\overline{X}_{\mathrm{mle}} = e^{m_x + \frac{1}{2}s_x^2} \qquad (22)$$

and the standard deviation by

$$\mathrm{SD}_X = \overline{X}_{\mathrm{mle}}\sqrt{e^{s_x^2} - 1}. \qquad (23)$$

The median or geometric mean may be estimated by

$$\overline{X}_{\mathrm{med}} = e^{m_x} \qquad (24)$$

and the mode (most frequent value) by

$$\overline{X}_{\mathrm{mod}} = e^{m_x - s_x^2}. \qquad (25)$$

The above equations are based on the MLE method which was demonstrated to be valid for means of CZCS data and moored fluorometer data. Equations (22)–(25) are based on an assumed lognormal distribution of X within the sampling domain. For a discussion of the underlying assumptions and robustness of the estimators see Section 3.3.

2.4 Protocols for Level-4 Statistics

As defined earlier, a variable, $Y = f(X)$, which is a function of one or more level-3 variables, is called a *level-4* variable. Here, guidelines are given for computing statistics of several classes of level-4 variables. It is not possible to specify protocols for all level-4 variables, in general, because the procedures depend on the function $f(X)$.

2.4.1 Computing Statistics for $Y=A+BX$

If Y is a linear function of X, then the mean of Y is given by the same linear function of the mean of X

$$\overline{Y}_{\mathrm{mle}} = A + B\overline{X}_{\mathrm{mle}}. \qquad (26)$$

The same is true for the median and mode of Y. The standard deviation of Y is scaled by the factor B

$$\mathrm{SD}_Y \;=\; B\big(\mathrm{SD}_X\big). \tag{27}$$

2.4.2 Computing Statistics for $Y{=}AX^B$

The MLE method was chosen because it provides a robust method for estimating the mean of level-4 variables of this form. To use the MLE method, one must first estimate the mean and variance of $\ln(Y)$. These statistics, m_y and s_y^2, can then be substituted into (22)–(25), in place of m_x and s_x^2, to estimate the mean, standard deviation, median, and mode of Y.

Let $Y = f(X)$ be a function of this form where X is a single level-3 variable. Its natural logarithm is a linear function of $\ln(X)$ (8). If m_x and s_x^2 are statistics of $\ln(X)$ derived from the level-3 data sets by (20) and (21), respectively, then the mean and variance of $\ln(Y)$ are, respectively:

$$m_y \;=\; \ln(A) \;+\; Bm_x \tag{28}$$

and

$$s_y^2 \;=\; B^2 s_x^2. \tag{29}$$

Statistics of $Y = f(X)$ can be derived by substituting $m_y = m_x$ and $s_y^2 = s_x^2$ into (22)–(25).

2.4.3 Statistics for Other Functions

So far the only considerations were functions of a single variable X. In general, if Y is a function of two or more level-3 variables, knowledge of the covariances between the level-3 variables is required to derive statistics for Y. It was initially recommended that a covariance matrix be saved as part of the level-3 statistics, but the storage costs were considered too high. Subsequently, it was decided to save statistics of $\mathrm{IC_K}$ because this function appears frequently in primary productivity algorithms.

Another situation involving a function of several level-2 variables occurs when a regional bio-optical algorithm is applied to derive better estimates of the CZCS-like pigment concentration. For example, suppose the standard (*global*) CZCS-like pigment algorithm is

$$\mathrm{PIG} \;=\; \mathrm{A_g}\left[\frac{\mathrm{L_{WN}(\lambda_i)}}{\mathrm{L_{WN}(\lambda_j)}}\right]^{\mathrm{B_g}} \tag{30}$$

where $L_{WN}(\lambda_i)$ and $L_{WN}(\lambda_j)$ are the normalized water-leaving radiance in bands i and j, and the wish is to compute pigment according to an alternative algorithm

$$\mathrm{PIG_r} \;=\; A_r\left[\frac{L_{WN}(\lambda_i)}{L_{WN}(\lambda_j)}\right]^{B_r} \tag{31}$$

using regionally-derived parameters, A_r and B_r. In this situation, it is possible to use the saved level-3 statistics for PIG to estimate statistics for $\mathrm{PIG_r}$. Substituting the means of $L_{WN}(\lambda_i)$ and $L_{WN}(\lambda_j)$ into (31) is not recommended.

The recommended procedure is, first, to estimate the mean and variance of $\ln\big(\mathrm{PIG}\big)$ according to (20) and (21). These statistics can be denoted by m_g and s_g^2, respectively. The mean of $\ln\big(\mathrm{PIG_r}\big)$ is then given by

$$m_r \;=\; \ln\big(A_r\big) \;+\; \frac{B_r}{B_g}\Big(m_g - \ln\big(A_g\big)\Big) \tag{32}$$

and the variance of $\ln\big(\mathrm{PIG_r}\big)$ is

$$s_r^2 \;=\; \frac{B_r^2 s_g^2}{B_g^2}. \tag{33}$$

These statistics can then be substituted into (22)–(25), replacing $m_r = m_x$, and $s_r^2 = s_x^2$, to obtain the statistics for $\mathrm{PIG_r}$.

This flexibility is the primary reason that the MLE method was chosen over the more commonly used estimation methods, e.g., arithmetic averages, for estimating spatial and temporal means. As shown in Section 3, the MLE estimator for the mean proved to be equivalent to the arithmetic average for spatial averages of CZCS data, and, in most situations, for temporal averages of moored fluorometer data. The statistical study detailed in Section 3 provides empirical evidence to support the use of the MLE method, as well as theoretical results which explain its success and, in some instances, failure for certain data sets.

3. EMPIRICAL BASIS

In 1992–93, a study was conducted to address statistical questions related to level-3 binning algorithms for SeaWiFS data. The questions addressed and recommendations derived from this study have been presented in Section 2 of this report. Here, the actual results of this study are presented. Results pertaining to spatial binning algorithms are presented in Section 3.1, followed by results pertaining to temporal binning algorithms in Section 3.2. Following the presentation of results, Section 3.3 contains a discussion of the major conclusions. Questions concerning the equivalence of the MLE and AVG methods are addressed in this section, and specific situations are described when the two methods would and would not be equivalent.

3.1 Spatial Statistics

The first step in creating level-3 data involves averaging data from a single orbital pass. This is considered the spatial binning step, because the data involved are regarded as simultaneous.

Three questions related to spatial binning were addressed:

1. How should level-2 data be averaged to provide the best estimate of their mean?
2. How should *level-4* means be estimated?
3. What statistics should be saved?

These are the first three questions presented and discussed in Sections 2.2.1–2.2.3.

3.1.1 Methods

Full-resolution CZCS data were used to address the aforementioned questions. The procedure was to use the full-resolution data to define the *true* mean of each variable within $9 \times 9\,\mathrm{km}^2$ bins and to compare other estimates of the mean against the *true* mean.

Seven scenes were selected as representative of the full range of variability in CZCS data. Details of these scenes are given in Table 1. The level-1 data were processed according to standard algorithms using the DSP `ANLY2DBL` code [Rosenstiel School of Marine and Atmospheric Science (RSMAS) 1990]. (The version of `ANLY2DBL.EXE` used in processing CZCS data was created 19 April 1990, and modified 18 September 1991). The resulting level-2 variables involved in this study were:

$L_{WN}(\lambda_i)$	normalized water-leaving radiances in bands $i = 1$–3,
CHL	pigment concentration (*chlorophyll*), and
K_{490}	diffuse attenuation coefficient at $\lambda = 490\,\mathrm{nm}$.

The normalized water-leaving radiances are radiances corrected for variations in solar zenith angle across the scan. All radiances are corrected to correspond to a solar zenith angle of zero. Details of the algorithms used may be found in Gordon et al. (1988).

The algorithm for K_{490} was

$$K_{490} \;=\; 0.022 \;+\; 0.088 \left[\frac{L_W(\lambda_1)}{L_W(\lambda_3)} \right]^{-1.491} \tag{34}$$

where $L_W(\lambda_i)$ is the non-normalized water-leaving radiance in band i. The quantity CHL was derived using a bifurcated algorithm that involved two ratio formulas:

$$\mathrm{CHL}_{13} \;=\; 1.130 \left[\frac{L_W(\lambda_1)}{L_W(\lambda_3)} \right]^{-1.705} \tag{35}$$

and

$$\mathrm{CHL}_{23} \;=\; 3.327 \left[\frac{L_W(\lambda_2)}{L_W(\lambda_3)} \right]^{-2.44} . \tag{36}$$

According to this algorithm, CHL was equal to CHL_{13} except when both formula values exceeded $1.5\,\mathrm{mg\,m^{-3}}$, in which case, CHL was equal to CHL_{23}. The CHL_{13} ratio was employed in all of the scenes analyzed, whereas the CHL_{23} ratio was employed in only three of the seven scenes.

After the scenes were processed to standard level-2 data, pixels in each scene were sorted into $9 \times 9\,\mathrm{km}^2$ bins oriented in rows perpendicular to the ground track of the satellite. Based on an instantaneous field-of-view (IFOV) angle of 0.865×10^{-3} radians (0.496°) and a sensor altitude of 955 km (and ignoring tilt), the spatial resolution of pixels at nadir is 0.825 km. The maximum number of pixels that fit into a $9 \times 9\,\mathrm{km}^2$ bin was 121 (11×11). This occurred only within ±300 pixels of nadir where pixels have spatial resolutions $\leq 0.9\,\mathrm{km}$.

3.1.1.1 Estimators of the Mean

Only cloud-free bins containing 121 pixels were used for the analysis. All estimators were evaluated using both full-resolution (LAC) data and 4 km resolution (GAC) data. The latter were obtained by subsampling every fifth pixel on every fifth line (since $5 \times 0.825 \approx 4\,\mathrm{km}$). Thus, LAC estimators were based on 121 level-2 observations, whereas for GAC data, the number of observatons (pixels falling in these bins) ranged from 4–9.

The estimators compared were:

AVG	arithmetic average (2) based on LAC data,
AVG4	arithmetic average based on GAC data,
MLE	maximum likelihood estimator (5) based on LAC data,
MLE4	maximum likelihood estimator based on GAC data,
MED	geometric mean or median estimator (4) based on LAC data, and
MED4	geometric mean or median estimator based on GAC data.

For each bin, the AVG estimator based on LAC data ($n = 121$) is given in (2) and was considered the *true* mean. In this equation, X_i is the ith observation or *realization* of the variable X $\big[$equal to $L_{WN}(\lambda_1)$, $L_{WN}(\lambda_2)$, $L_{WN}(\lambda_3)$, CHL, or $K_{490}\big]$, and n is the number of observations (pixels) falling in a bin. The true mean was computed for each variable and each bin having $n = 121$ valid observations. The other estimators of the mean were compared with $\overline{X}_{\mathrm{avg}}$ to determine how well they performed.

3.1.1.2 Standard Level-2 Variables

Let $\mathbb{X} \equiv \big[L_{WN}(\lambda_1),L_{WN}(\lambda_2),L_{WN}(\lambda_3),\mathrm{CHL},K_{490}\big]$ refer to the vector of standard variables, and let $Y = f(\mathbb{X})$ be any function that is derived from one or more of the standard variables.

The arithmetic mean of the function based on LAC data (n=121)

$$\overline{Y}_{\mathrm{avg}} \;=\; \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{37}$$

Table 1. CZCS scenes used for the analysis of spatial statistics. The scenes are listed in increasing order of mean pigment (see Fig. 1). The number of lines listed were for the whole scene, and the number of bins given is the number of $9 \times 9\,\mathrm{km}^3$ bins containing data. Time is given in Greenwich Mean Time (GMT) in the (left-to-right) order of hour, minutes, and seconds. (Note: In Tables 2 and 3, the number of bins listed is the number of bins containing $n = 121$ pixels. Only these cloud-free bins were used to define *true* means in the images.)

ID	*Orbit*	*Date*	*Time*	*Tilt*	*Location*	*Lines*	*Bins*
1	1,200	19 Jan 79	1:56:27	20°	Northwestern Pacific	1,023	3,186
2	218	9 Nov 78	0:52:23	0	Northwestern Pacific	1,023	2,964
3	1,029	6 Jan 79	16:38:13	−14	South Atlantic	1,023	3,087
4	1,016	5 Jan 79	18:33:31	6	Eastern Tropical Pacific	2,376	8,266
5	1,452	6 Feb 79	7:19:17	8	Indian Ocean	1,584	7,475
6	971	2 Jan 79	12:31:21	−2	Northwest of Africa	1,023	1,040
7	1,386	1 Feb 79	12:45:19	20	Southwest of Africa	1,584	4,020

was considered its true mean, where $Y_i = f(X_i)$ is the function calculated at pixel i. This defined the AVG estimator for $\overline{Y}$. Similarly, the AVG4, MLE and MLE4 estimators for the mean of Y were defined by substituting Y_i for X_i in the appropriate equations. In addition to these estimators, the FNC (*function*) estimator was defined as

$$\overline{Y}_{\mathrm{fnc}} = f\big(\overline{X}_{\mathrm{avg}}\big) \tag{38}$$

where $\overline{X}_{\mathrm{avg}}$ is the arithmetic average of X. This would be the result of calculating the function using level-3 means. It was called FNC when $\overline{X}_{\mathrm{avg}}$ was the AVG estimator, and FNC4 when $\overline{X}_{\mathrm{avg}}$ was the AVG4 estimator.

Functions that were investigated were as follows:

$\mathrm{IC_K}$ integral pigment (1) within the upper optical depth,

Z_e 1% light depth, and

$Y_{A,B}$ pigment *algorithm* $A\big(L_{WN}(\lambda_1)/L_{WN}(\lambda_3)\big)^B$, where $A = 1$ and $B = -1$, -2, and -3.

3.1.1.3 Relative Errors

For each bin, the relative error in an estimate of the mean, $\overline{X}_{\mathrm{est}}$, was defined as a percentage of the true mean $\overline{X}_{\mathrm{avg}}$

$$\mathrm{ERROR} = \frac{\overline{X}_{\mathrm{est}} - \overline{X}_{\mathrm{avg}}}{\overline{X}_{\mathrm{avg}}} \times 100\% \tag{39}$$

where $\overline{X}_{\mathrm{est}}$ was the estimate based on the MLE, MED, AVG4, MLE4, or MED4 estimator. Similarly, relative errors in estimates of the mean of a function, $\overline{Y}_{\mathrm{est}}$, were defined as a percentage of $\overline{Y}_{\mathrm{avg}}$, where $\overline{Y}_{\mathrm{est}}$ was the estimate based on the MLE, FNC, AVG4, MLE4, or FNC4 estimator.

3.1.2 Results

In Table 1, the scenes are listed in order of increasing mean pigment. In presenting results, scenes will be identified by the number (order) found in column 1 of this table.

3.1.2.1 Pigment Distributions

The pigment means and coefficients of variation (CV) for the seven scenes are compared in Fig. 3. Histograms of log(CHL) are shown in Fig. 4, where the abscissa is the 8-bit image value V, which is related to the logarithm (base 10) of pigment as

$$\log\big(\mathrm{CHL}\big) = -1.4 + 0.012\,V. \tag{40}$$

The distributions of log(CHL) shown in Fig. 4 appear to be either single normal distributions, e.g., scene 1, or mixtures of normal distributions, e.g., scene 3. Thus, CHL is approximately lognormally distributed within each scene or within portions of each scene.

In scenes 4, 6, and 7, the bifurcated CHL algorithm resulted in a discontinuity at $\mathrm{CHL} = 1.5\,\mathrm{mg\,m^{-3}}$ ($V = 132$). Values to the left of $V = 132$ have been calculated according to CHL_{13} (35), whereas values to the right were calculated according to CHL_{23} (36). This is an artifact of the CZCS pigment algorithm, which will be avoided when defining the SeaWiFS CHL algorithm. In scenes 6 and 7, CHL was recalculated using the CHL_{13} algorithm for all pixels. The resulting CHL distributions are shown in Fig. 5.

3.1.2.2 Comparison of Estimators

Representative results for estimators of CHL are shown in Figs. 6 and 7. Each point in these scatter plots corresponds to a bin in scene 4, the scene with the highest overall variance. The scales are log-log. In Fig. 6, the MLE, MED, MLE4, and MED4 estimates are plotted against the AVG estimate. The patterns shown here are typical of those observed in all the scenes analyzed. In all scenes, the MLE estimator was nearly identical to the AVG estimator, whereas the MED estimator underestimated AVG. There was no discernible difference between the MLE4 versus AVG and MED4 versus AVG plots. Both contained substantially more scatter than the plots involving MLE and MED estimates.

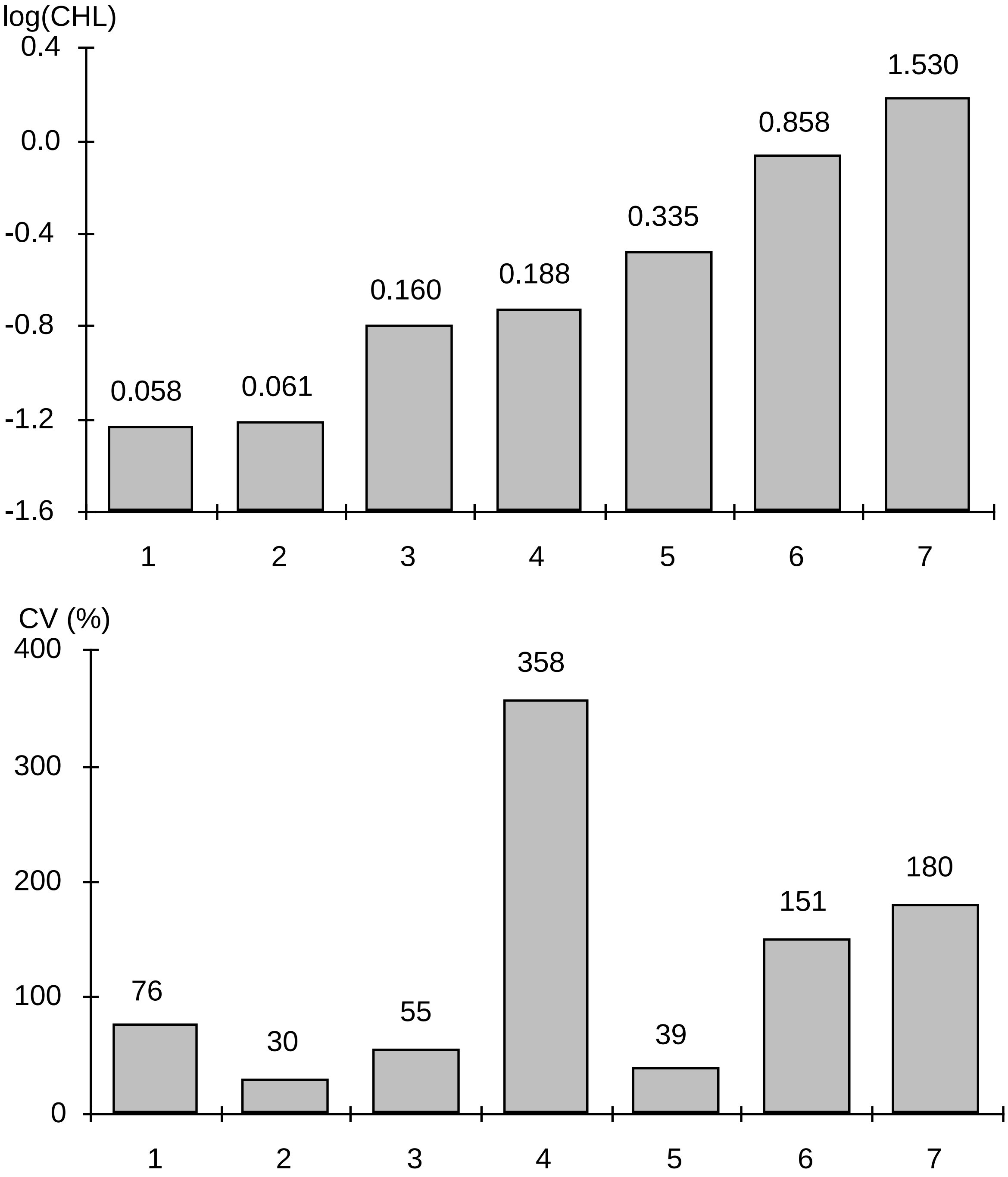
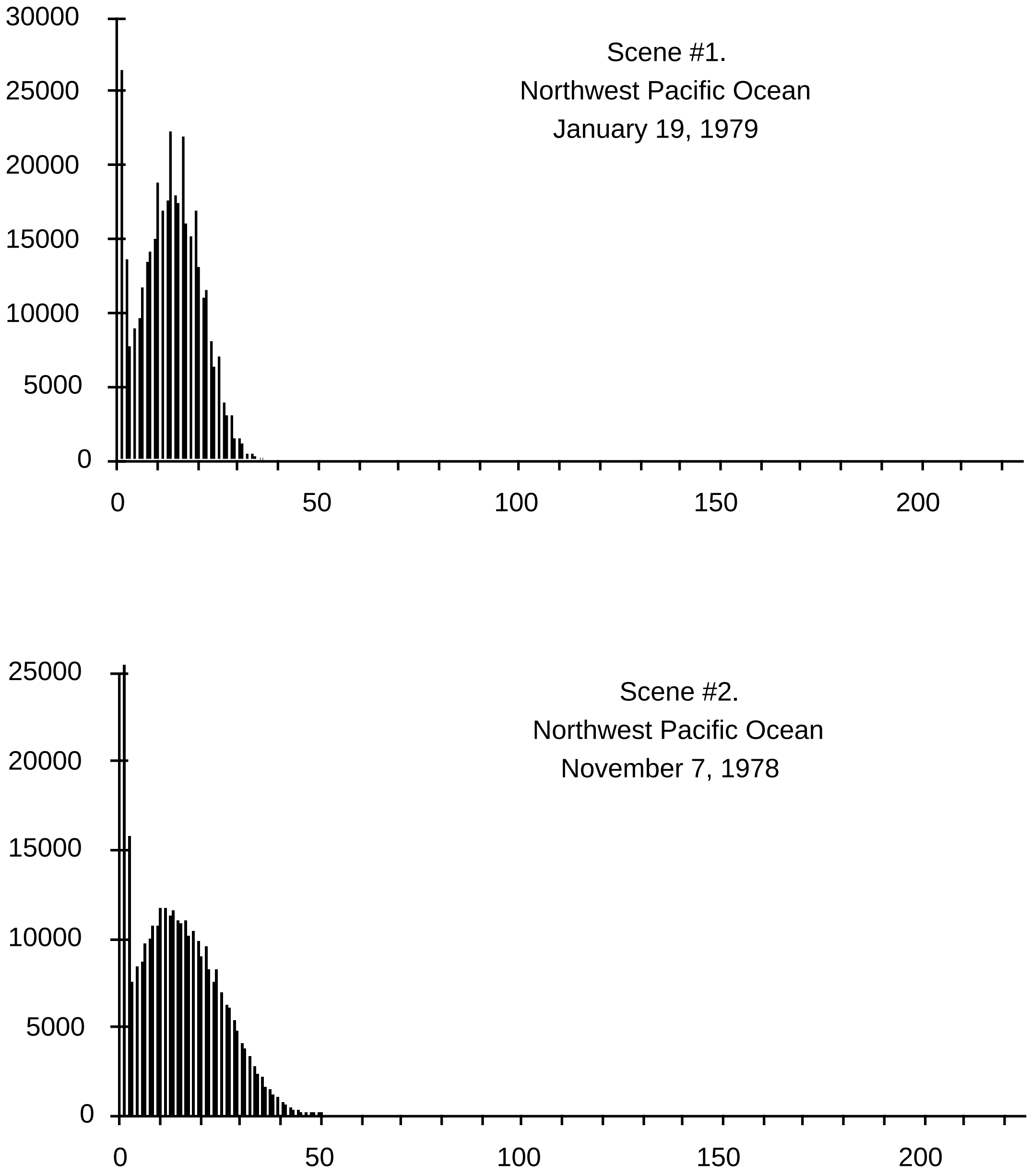
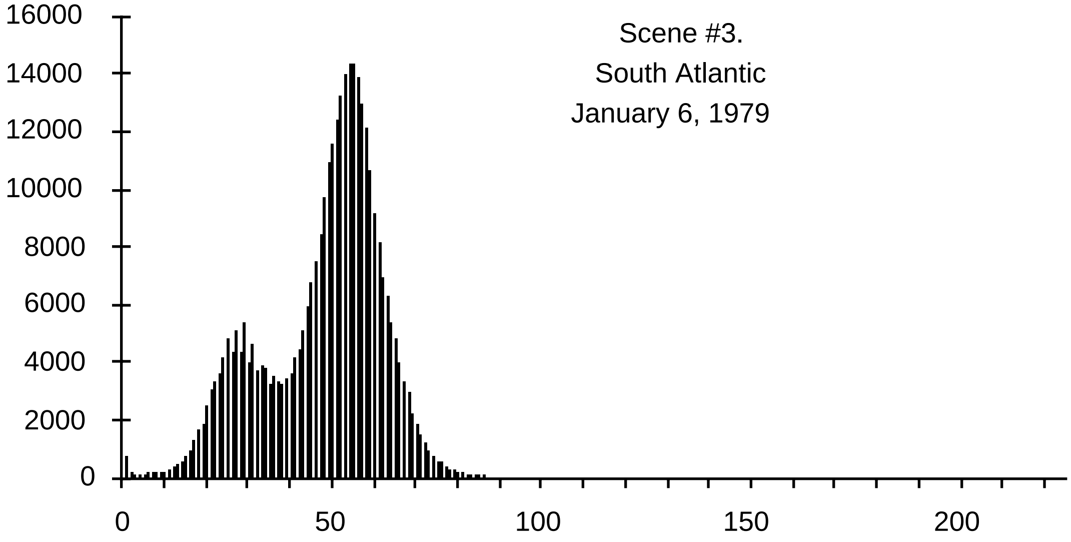


Fig. 3. The mean pigment (upper panel) and coefficient of variation (lower panel) for the seven CZCS scenes used in this analysis. The scenes are ordered from lowest to highest mean pigment. The numbers appearing above each bar are the mean pigment (mg m^{-3}) and coefficient of variation (standard deviation expressed as a percentage of the mean).


Scene #1.
Northwest Pacific Ocean
January 19, 1979

Scene #2.
Northwest Pacific Ocean
November 7, 1978

Fig. 4. Pigment histograms of seven CZCS scenes used in this analysis. The abscissa is the image value V which is linearly related to the logarithm of pigment: log(CHL) = −1.4 + 0.012(V).



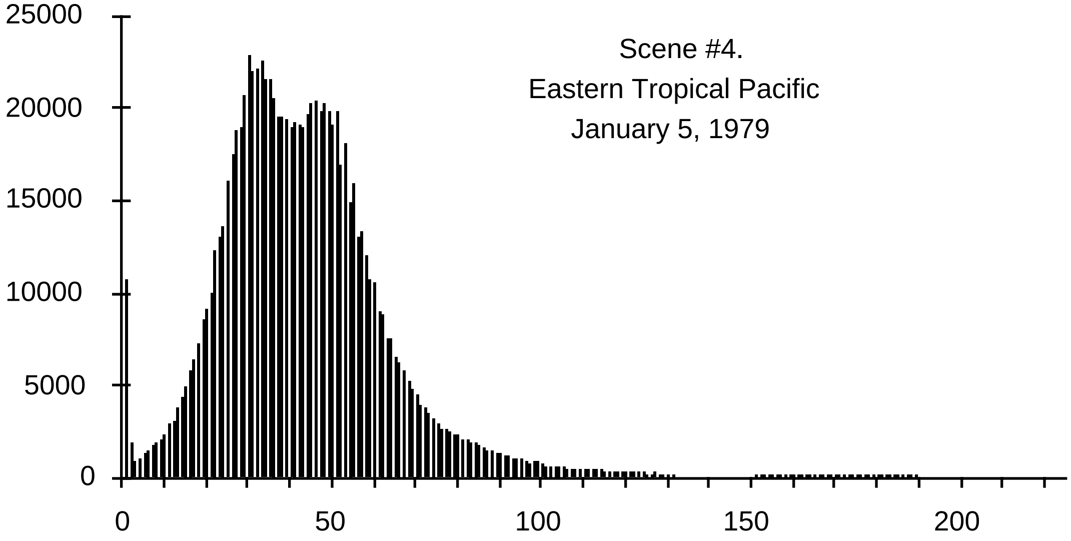
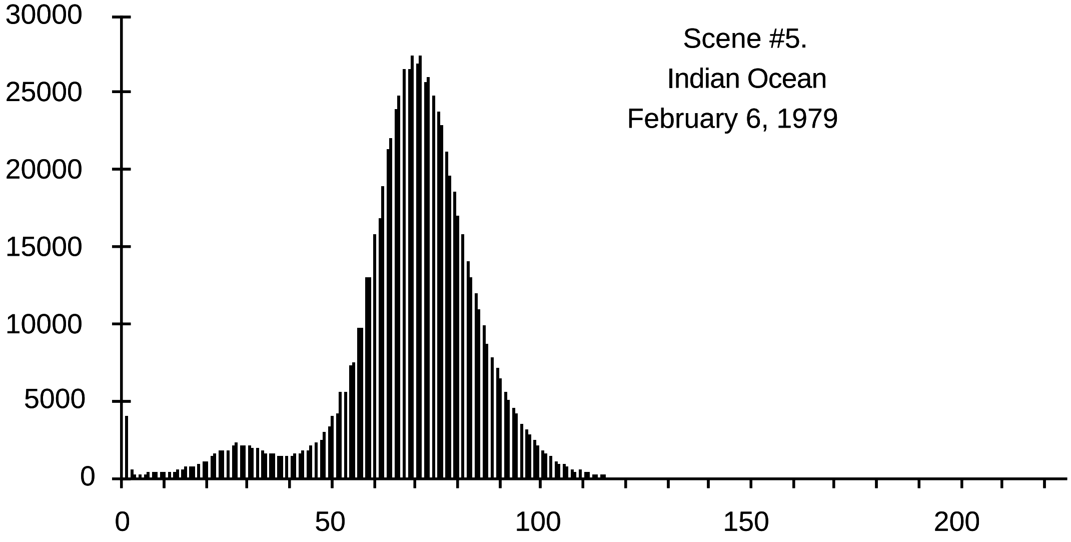


Fig. 4. (cont.) Pigment histograms of seven CZCS scenes used in this analysis.



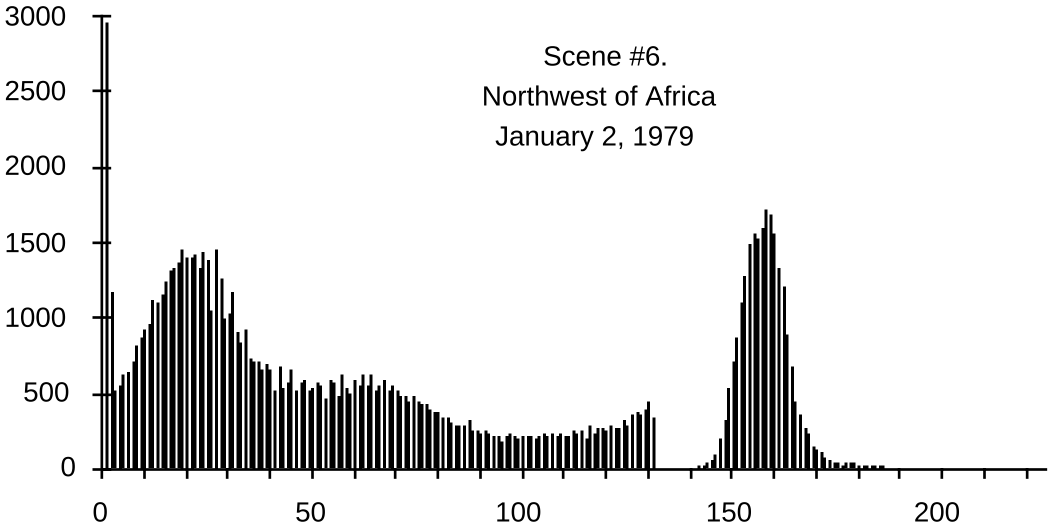


Fig. 4. (cont.) Pigment histograms of seven CZCS scenes used in this analysis.

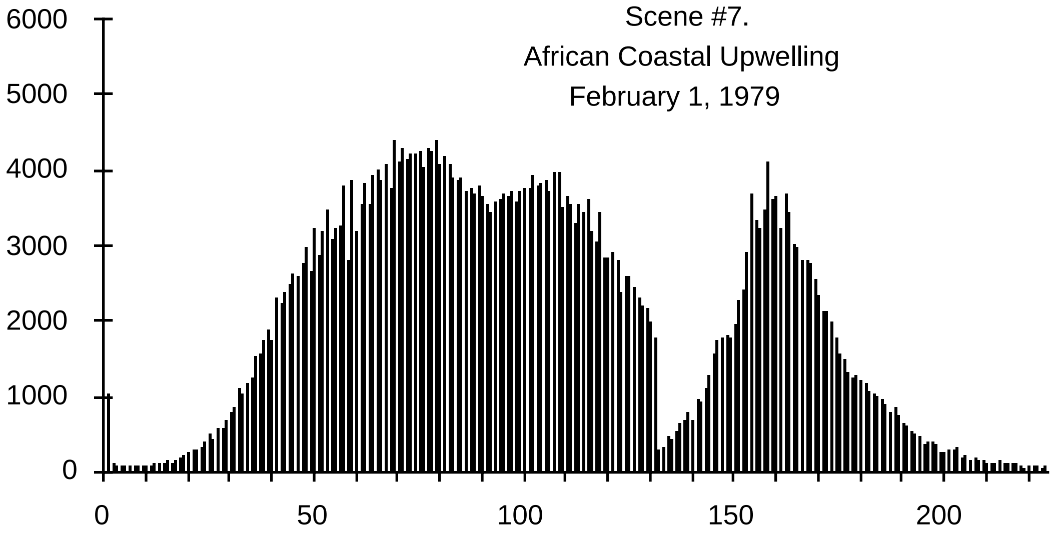


Fig. 4. (cont.) Pigment histograms of seven CZCS scenes used in this analysis.

In Fig. 7, the AVG4 estimator is compared with the AVG and MLE4 estimators. In the AVG4 versus AVG plot, the scatter is strictly the result of sample size differences; whereas in the AVG4 versus MLE4 plot, the scatter is the result of differences between the estimators. It is clear from these comparisons that the errors associated with GAC estimators were predominantly the result of their reduced sample size. When two GAC estimators were compared, e.g., AVG4 versus MLE4 in Fig. 7, the two agreed as well as the corresponding LAC estimators.

Color Plates 1 and 2 show level-3 mean CHL images for the seven scenes. That is, each pixel in these images is a bin in the level-3 data. Plate 1 compares the AVG and MLE estimators, and Plate 2 compares the AVG4 and MLE4 estimators. Difference images are shown in Plate 3. Differences between the MLE and AVG estimators seemed to be spatially organized with the largest differences located along fronts and coastlines. In contrast, there were no obvious spatial patterns in the differences between MLE4 and AVG4 estimators.

The combined histograms of relative errors (39) in CHL estimators from all seven scenes are shown in Figs. 8 and 9, and summarized in Table 2a. In all but a few cases, the MLE estimator differed from the AVG estimator by less than 1%; whereas, the MED estimator consistently underestimated the mean CHL. Its bias or average error was −2.1%, and 95th percentile range was −11% to −1%.

All three GAC estimators had a tendency to underestimate the true mean CHL. Errors associated with the AVG4 estimator are strictly the result of reducing sample sizes from $n = 121$ in the AVG estimator to $n \leq 9$ in the AVG4 estimator. The error histograms for AVG4 and MLE4 are remarkably similar. Their biases were −0.76% and −0.75%, respectively, and their 95th percentile range was −19% to +18%. The MED4 tended to underestimate the true mean as did the other GAC estimators, but the MED4 was a poorer estimator indicated by its larger† negative bias (−2.60%).

In the LAC error histograms (Fig. 8), true differences in the performance of the estimators may be seen; whereas, in the GAC histograms (Fig. 9), errors associated with reduced sample size are added to errors or differences between estimators. Differences between GAC estimators

$$\mathrm{DIFF1} \; = \; \frac{\mathrm{MLE4} - \mathrm{AVG4}}{\mathrm{AVG4}} \times 100\% \qquad (41)$$

and

$$\mathrm{DIFF2} \; = \; \frac{\mathrm{MED4} - \mathrm{AVG4}}{\mathrm{AVG4}} \times 100\% \qquad (42)$$

were examined. Here, a distinction is made between *errors* (39) which are relative to the *true* mean (AVG) and *differences*, (41) and (42), which are relative to AVG4, another estimate of the mean.

Histograms of DIFF1 and DIFF2 are shown in Fig. 10. These results for GAC estimators are very similar to the patterns seen when comparing LAC estimators (compare Fig. 10 with Fig. 8). The AVG4 and MLE4 estimators agree, as well as the AVG and MLE estimators; differences between the two methods of estimating the mean CHL are negligible. Likewise, differences between the MED4 and AVG4 estimators followed the same pattern as differences between the MED and AVG estimators. In both cases, the geometric mean underestimated the arithmetic average. The large *errors* in AVG4, MLE4, and MED4 (Fig. 9) were clearly dominated by the sample size effect.

The patterns seen in Figs. 8–10 for CHL estimators are similar to those that are obtained for other variables. GAC error histograms for the other variables (comparable to Fig. 9) are shown in Figs. 11–14, and summaries

† In referring to biases, the terms *larger* and *smaller* refer to the magnitude or absolute value of the bias.

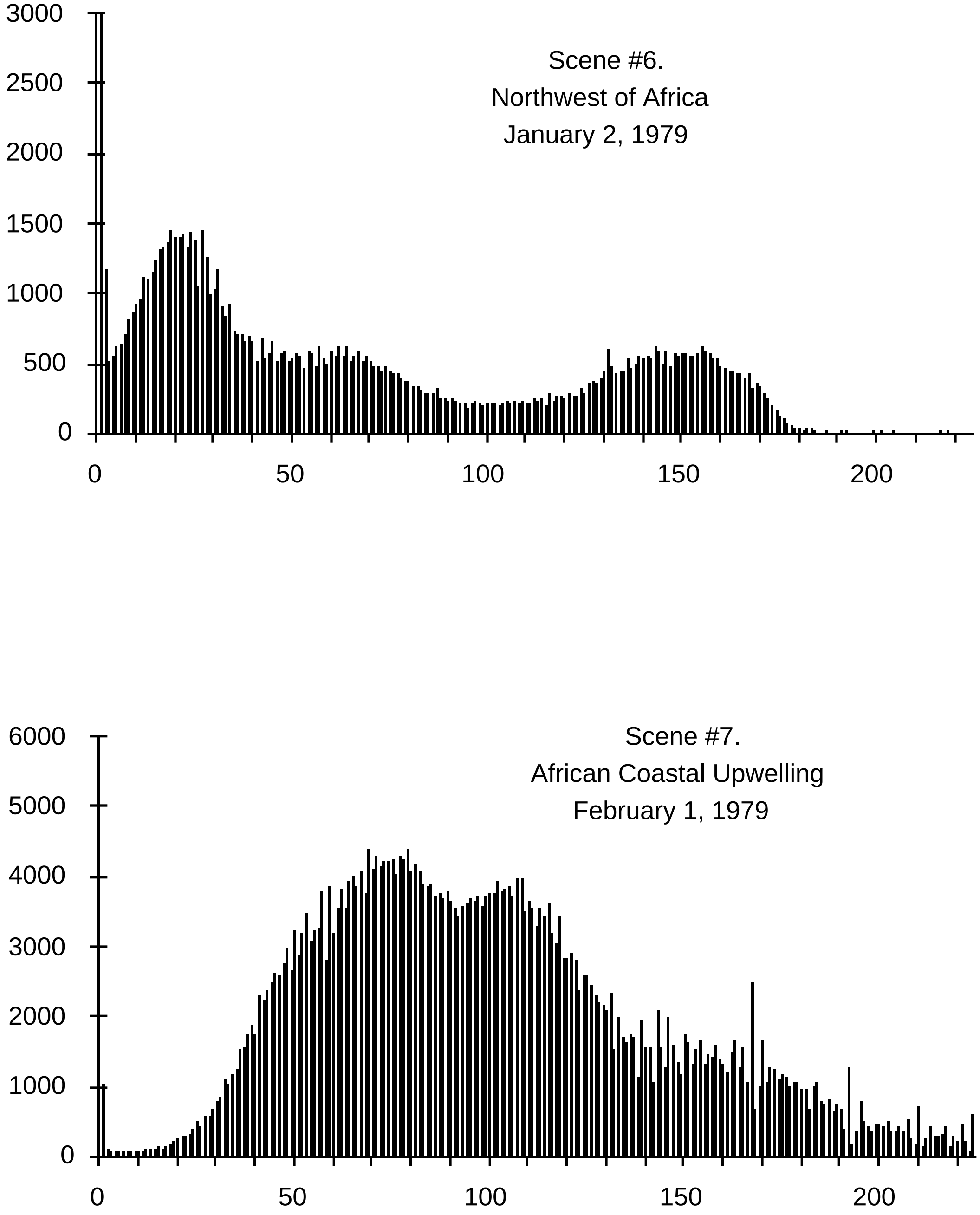


Fig. 5. CHL histograms for scenes 6 and 7 derived using CHL13 algorithm only.

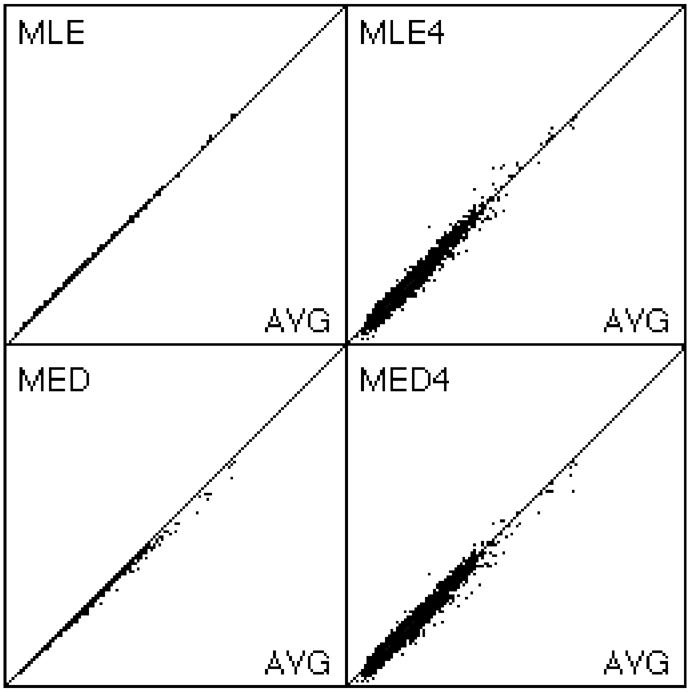


Fig. 6. In these scatter plots, four estimators of the mean are compared with the *true mean* (AVG) defined as the arithmetic average of all pixels in a bin (sample size = 121). The level-2 data used were CZCS-derived pigment values from scene 4 (see Table 1). Like the AVG estimator, the MLE and MED estimators are based on full-resolution (LAC) data, whereas the MLE4 and MED4 estimators are based on 4 km subsampled (GAC) data. The scales on each plot are log-log where the range is from 0.04 (V=0) to 45 (V=255), where V is the 8-bit image value [see (40)].

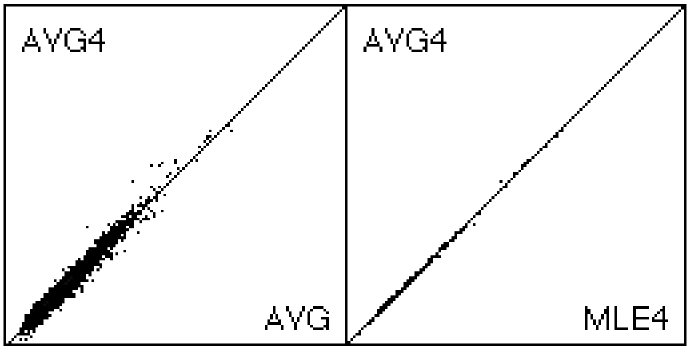


Fig. 7. In these scatter plots, the ordinate (AVG4) is the arithmetic average based on 4 km subsampled (GAC) data for the same scene as in Fig. 6. The plot on the left compares this estimator with the average based on full-resolution (LAC) data. The scatter in this plot is strictly the result of sample size differences. The AVG4 has less precision since its sample size is reduced from $n = 121$ (LAC) to $n \leq 9$ (GAC). The plot on the right compares the AVG4 and MLE4 estimators. Like the MLE and AVG estimators (Fig. 6), the MLE4 and AVG4 are practically identical.

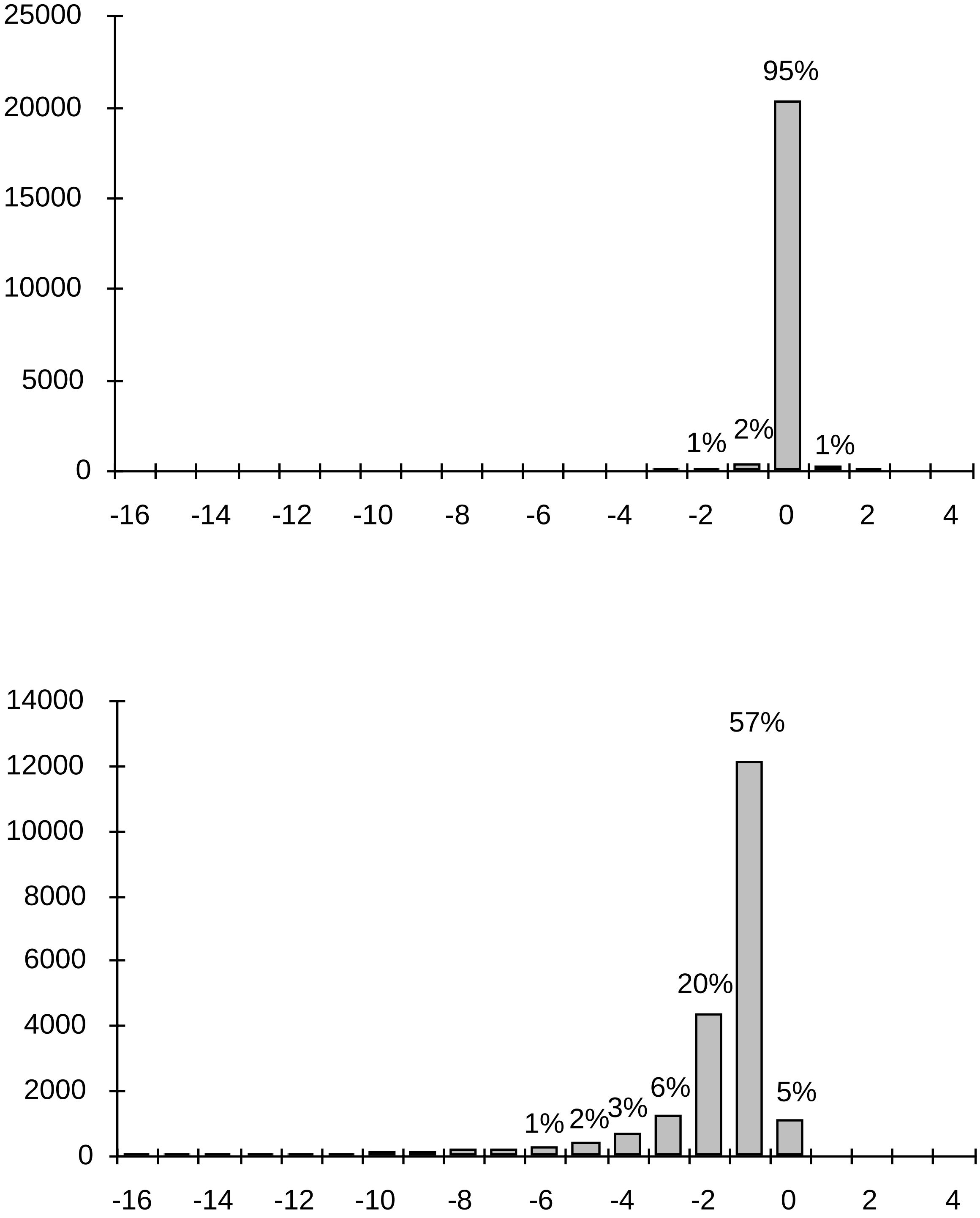


Fig. 8. Histograms of CHL estimation errors based on 21,290 bins analyzed and full-resolution (LAC) data. For each bin, the error is defined as the difference between the estimator and the arithmetic average (AVG) of all data in the bin expressed as a percentage of AVG. The top histogram shows the error calculated as (MLE – AVG)/ AVG) (%). The bottom histogram shows the error calculated as (MED – AVG)/ AVG (%).

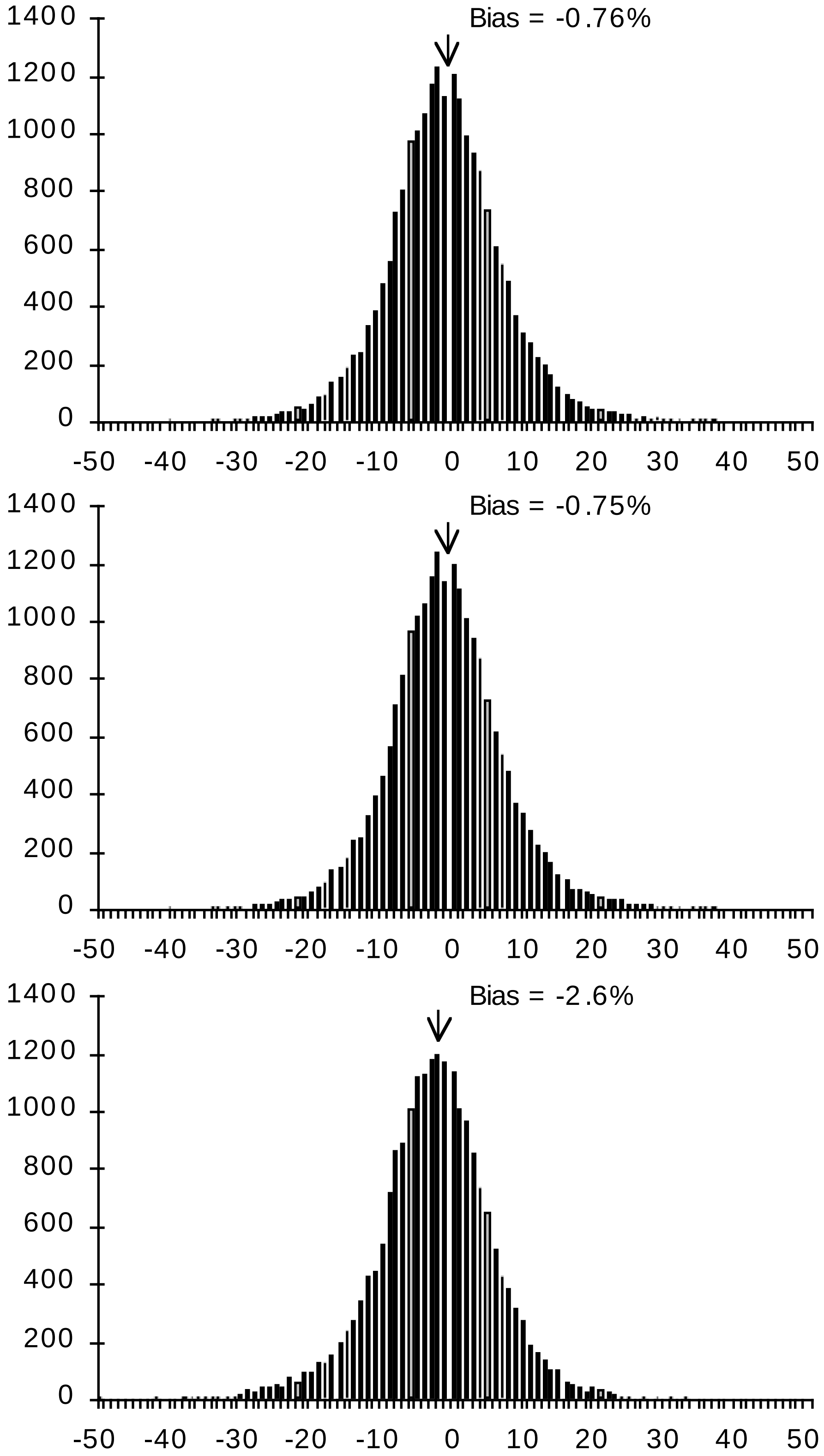
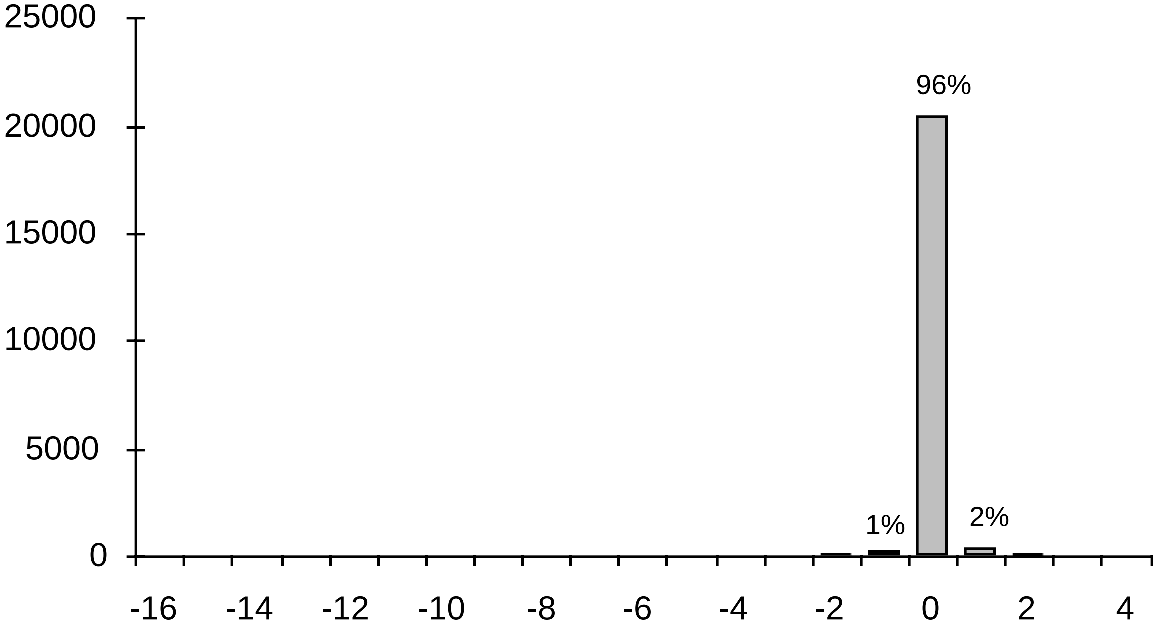


Fig. 9. Histograms of CHL estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data. The top histogram shows the error calculated by (AVG4 – AVG)/ AVG (%). The middle histogram shows the error calculated by (MLE4 – AVG)/ AVG (%). The bottom histogram shows the error calculated by (MED4 – AVG)/ AVG (%).



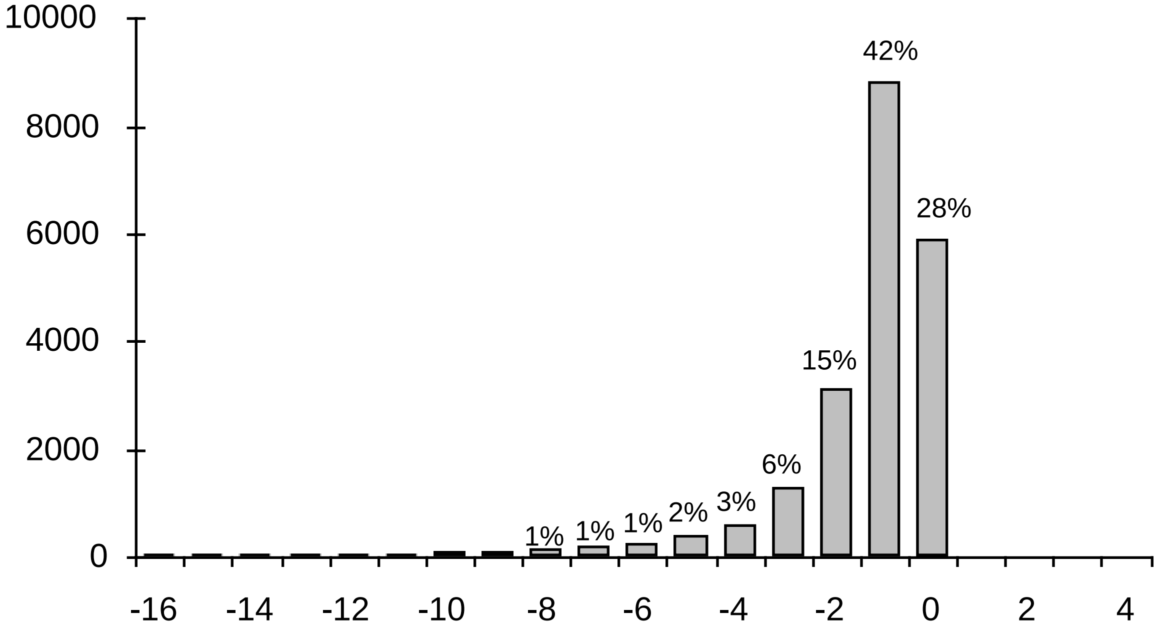
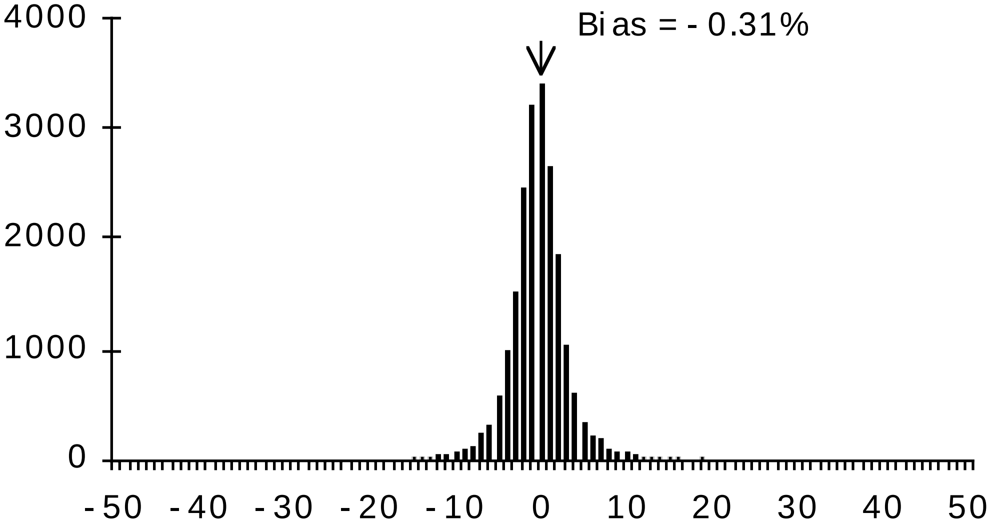


Fig. 10. Histograms of DIFF1 and DIFF2 based on 21,290 bins analyzed. The top histogram was calculated with (MLE4 – AVG4)/ AVG4 (%). The bottom histogram was calculated with (MED4 – AVG4)/ AVG4 (%).



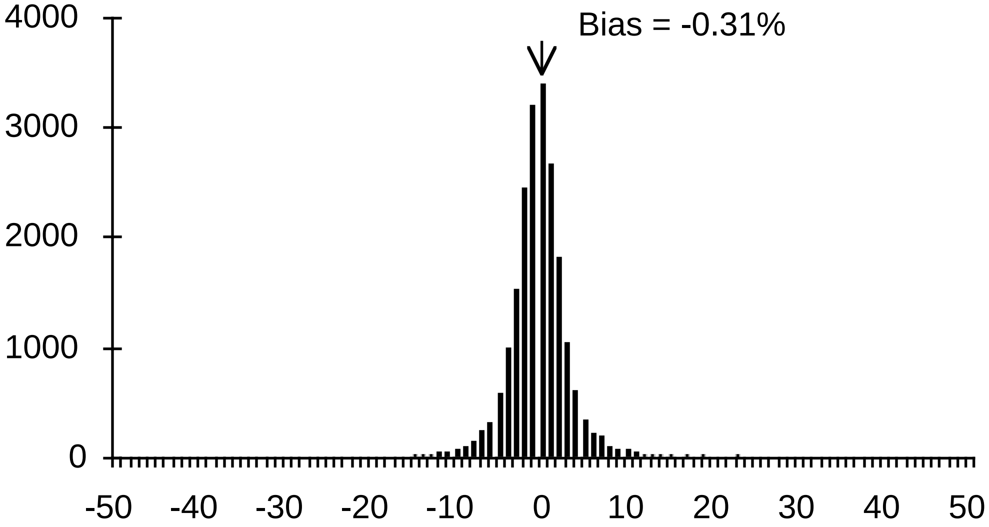


Fig. 11. Histograms of K_{490} estimation errors based on 20,373 bins analyzed and 4 km subsampled (GAC) data. The top histogram was calculated using (AVG4 – AVG)/ AVG (%). The bottom histogram was calculated for (MLE4 – AVG)/ AVG (%).

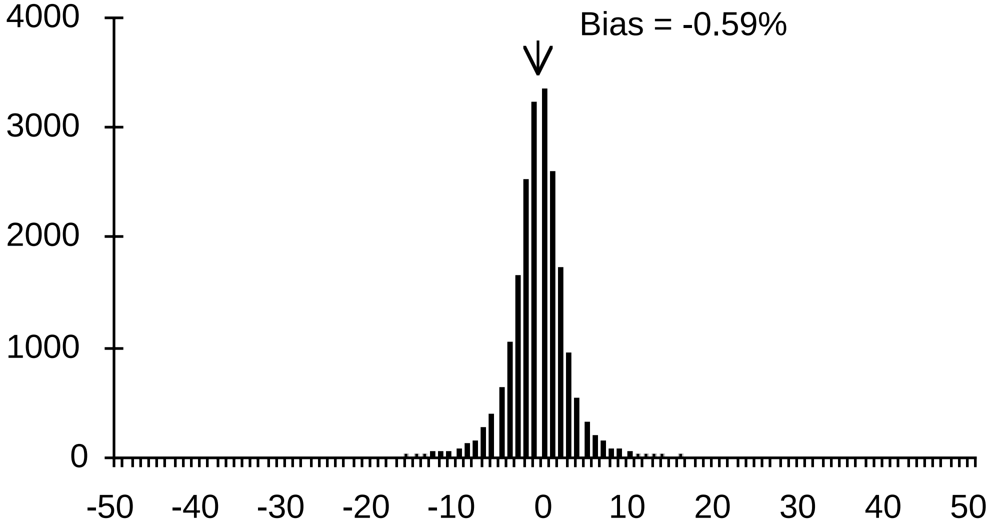


Fig. 11. (cont.) Histogram of K_{490} estimation errors based on 20,373 bins analyzed and 4 km subsampled (GAC) data was calculated using (MED4 − AVG)/ AVG (%).

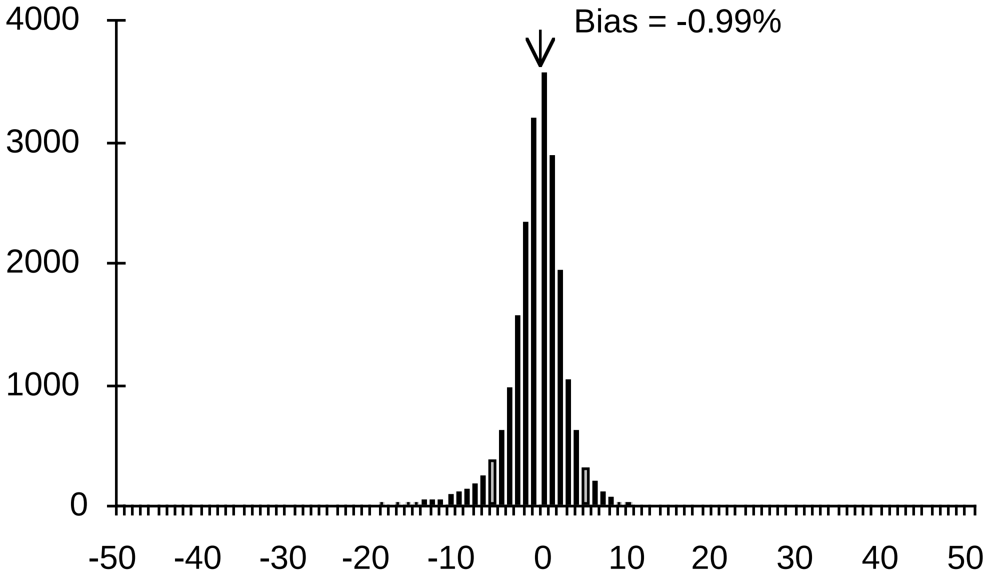
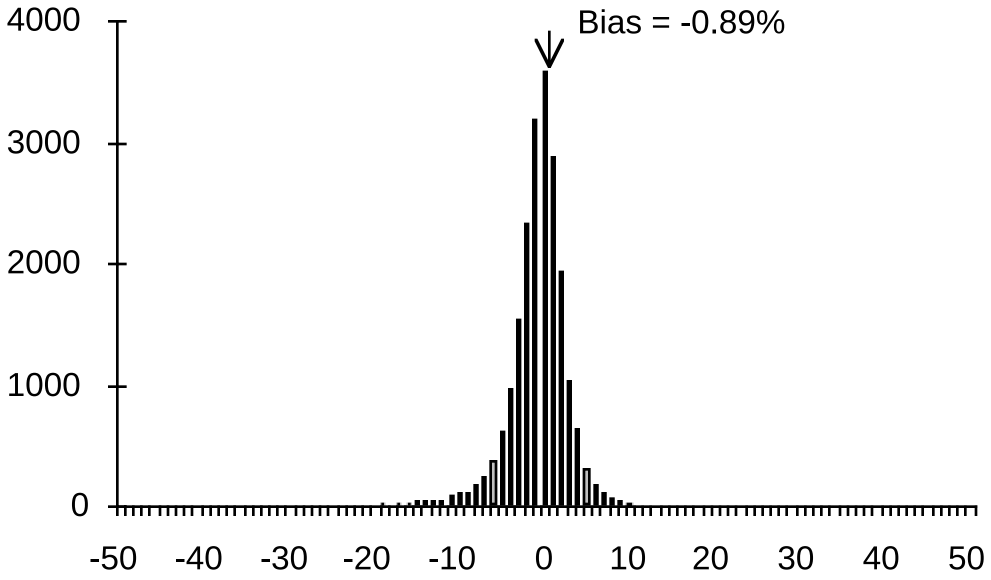


Fig. 12. Histogram of $L_{WN}(443)$ estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data was calculated for (AVG4 − AVG)/ AVG (%).



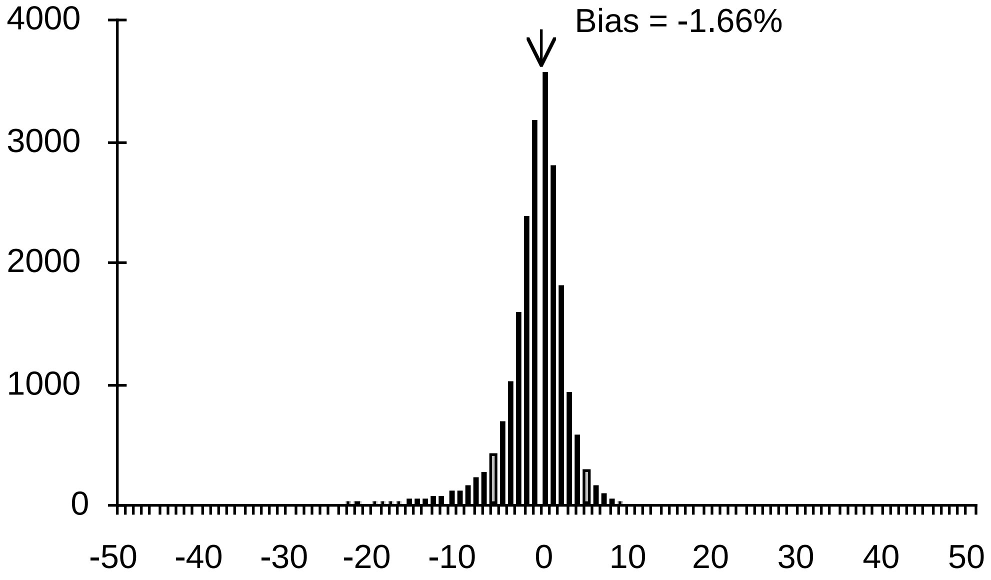
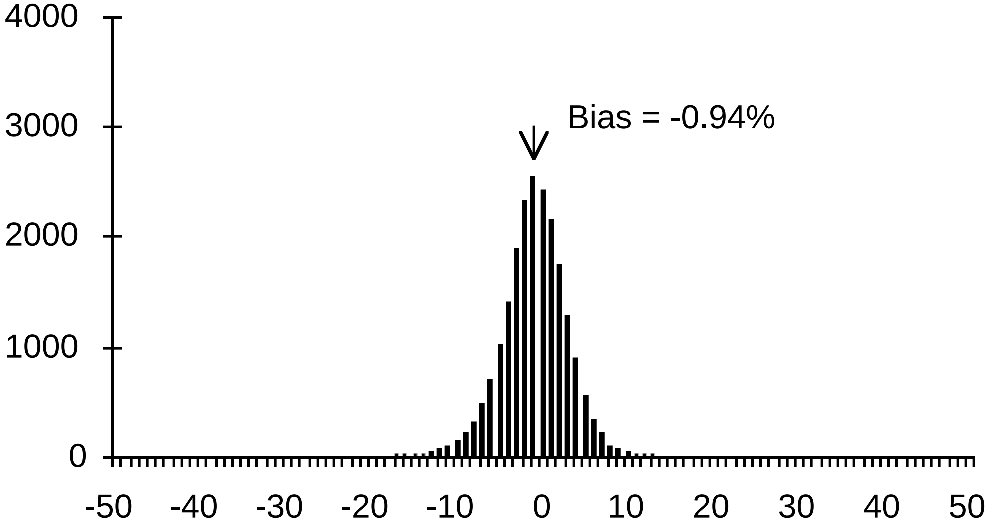


Fig. 12. (cont.) Histograms of $L_{WN}(443)$ estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data. The top histogram was calculated for (MLE4 − AVG)/ AVG (%). The bottom histogram was calculated for (MED4 − AVG)/ AVG (%).



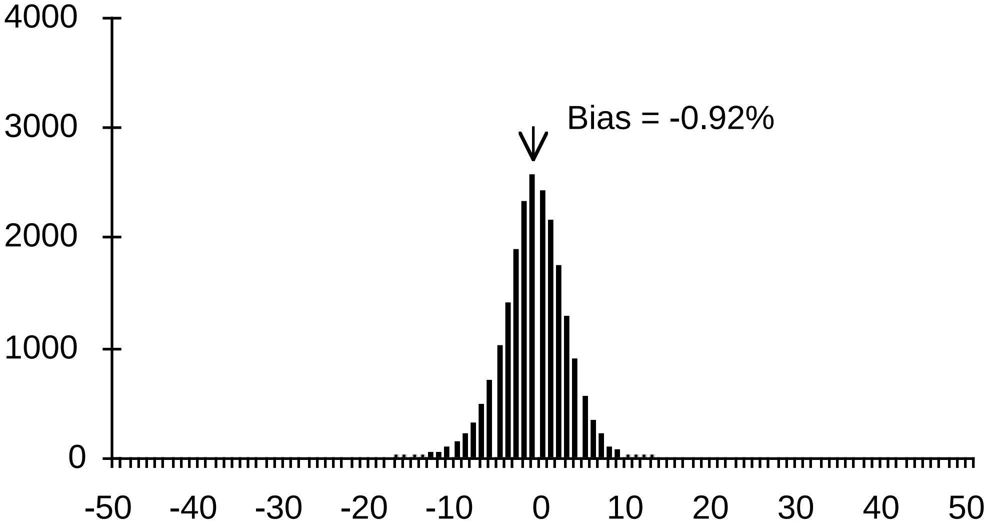


Fig. 13. Histograms of $L_{WN}(520)$ estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data. The top histogram was calculated for (AVG4 – AVG)/ AVG (%). The bottom histogram was calculated for (MLE4 – AVG)/ AVG (%).

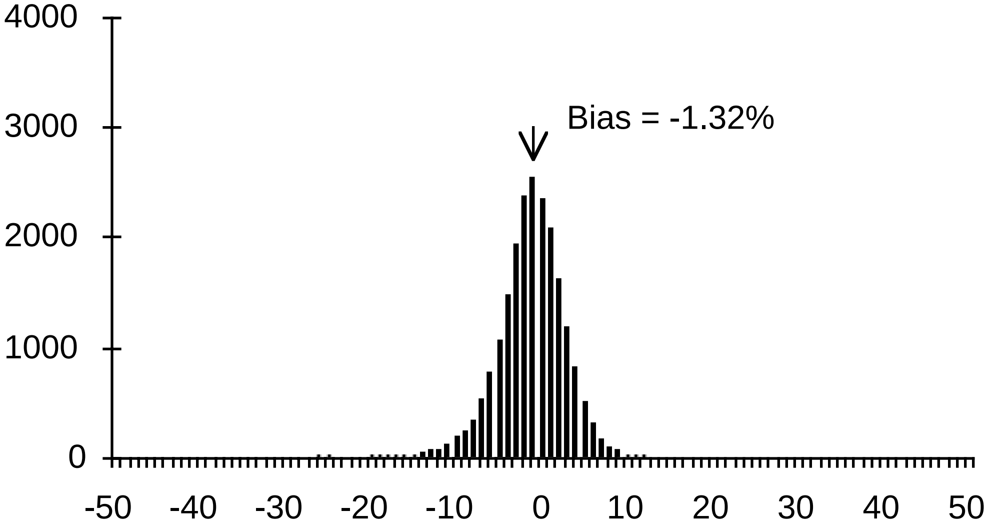


Fig. 13. (cont.) Histogram of $L_{WN}(520)$ estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data was calculated for (MED4 – AVG)/ AVG (%).

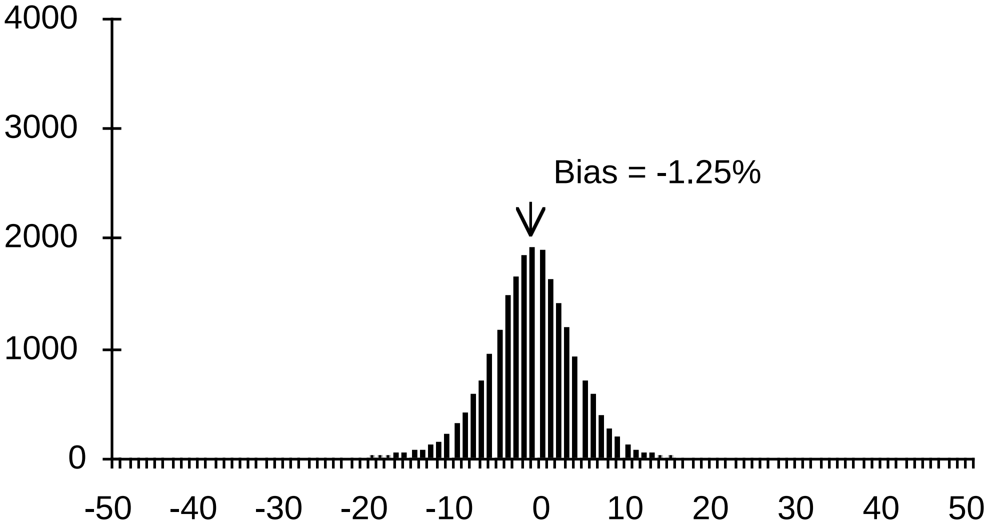
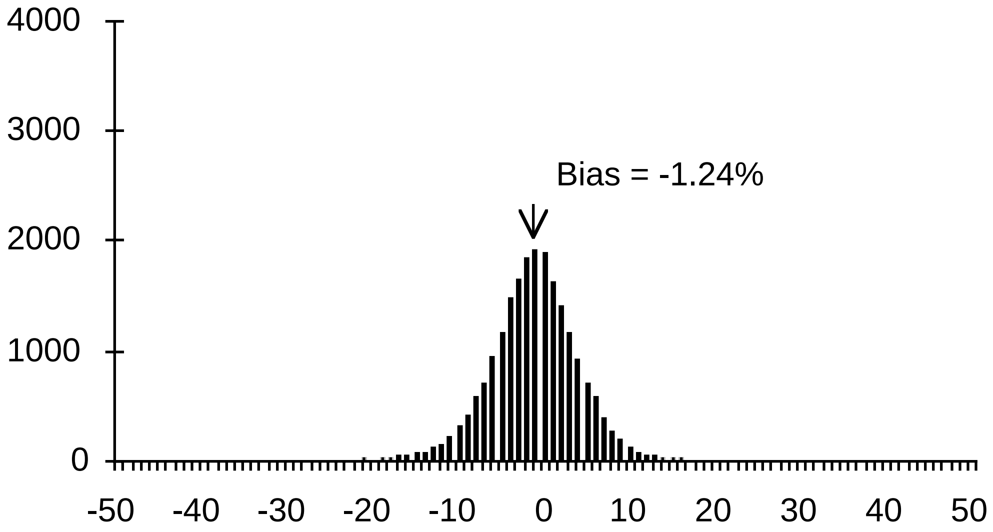


Fig. 14. Histogram of $L_{WN}(550)$ estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data was calculated for (AVG4 – AVG)/ AVG (%).



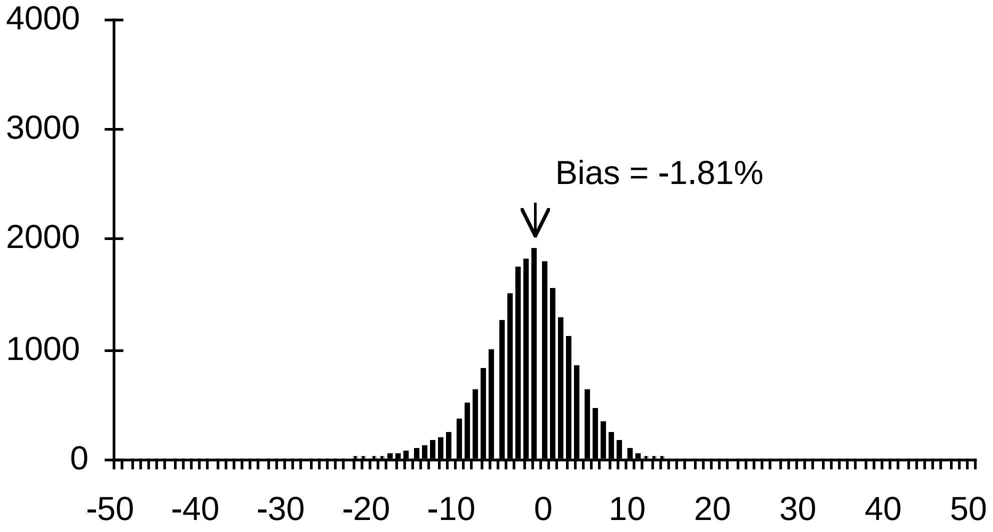


Fig. 14. (cont.) Histograms of $L_{WN}(550)$ estimation errors based on 21,290 bins analyzed and 4 km subsampled (GAC) data. The top histogram was calculated for (MLE4 – AVG)/ AVG (%). The bottom histogram was calculated for (MED4 – AVG)/ AVG (%).

of relative errors are listed in Tables 2a–e for CHL, K_{490}, $L_{WN}(\lambda_1)$, $L_{WN}(\lambda_2)$, and $L_{WN}(\lambda_3)$, respectively. The last two columns on the right in this table give the 95th percentile range for the relative errors. All GAC estimators had negative biases. AVG4 and MLE4 were nearly identical with average errors on the order of -1%; whereas, MED4 had average errors of approximately -2%. The CHL variable had the highest overall errors, with a 95th percentile range generally around ±20%; whereas, the other variables had errors that were generally within ±10%.

The three scenes that used both CHL_{13} and CHL_{23} had substantially higher errors than those of the other scenes. These scenes also had the highest variance in CHL and other variables. Although higher estimation errors would be expected when sampling from distributions with higher variance, there was the need to determine whether the CHL errors were anomalously large due to the bifurcated CHL algorithm. The higher variance in CHL might have been an artifact resulting from the discontinuous nature of the pigment distribution.

To determine whether this was true, the analysis was repeated for scenes 6 and 7 using CHL_{13} to derive CHL for all pixels (Fig. 5). The results were essentially the same. These images still had large intrabin variances in CHL (CHL_{13}), and their error distributions (not shown) were essentially unchanged.

Statistics pertaining to the $\mathrm{IC_K}$ function estimators are presented in Table 3, and error histograms for the MLE and FNC estimators are shown in Fig. 15 and for the AVG4, MLE4, and FNC4 estimators in Fig. 16. In contrast to the MED estimator, the FNC estimator tended to overestimate the true mean. In the case of the FNC4 estimator, this tendency (positive bias) was apparently offset by the underestimation tendency (negative bias) associated with small sample sizes. The result was that the bias of the FNC4 estimator was nearly zero.

In Section 2.4, a protocol was presented for estimating the mean of level-4 variables of the form $Y = AX^B$ based on saved statistics of the level-2 variable X. The accuracy of the prescribed protocol depends strictly on whether the MLE estimator is a good approximation to the AVG or true mean of these functions.

Results for the Z_e and $Y_{A,B}$ functions (not shown) established that the MLE estimator was essentially identical to the AVG estimator. Root-mean-square (rms) errors for the MLE4 and AVG4 estimators were within ±5% for Z_e. Errors for $Y_{A,B}$ increased as B changed from -1 to -3, with the highest rms errors being associated with scenes 6 and 7. MLE4 and AVG4 errors were within ±5% for $B = -1$, within ±15% for $B = -2$, and ±30% for $B = -3$. These ranges are consistent with the results for the CHL algorithm where $B = -1.7$ (CHL_{13}) and -2.4 (CHL_{23}). As in the case of the $\mathrm{IC_K}$ function, the FNC4 estimator was not significantly different from the MLE4 and AVG4 estimators. In all three cases, errors were dominated by the effects of reduced sample size.

3.2 Temporal Statistics

After the spatial statistics are derived from data on a single orbital pass, these statistics will be averaged over time to produce temporal statistics. No further reduction in spatial resolution takes place, but after being averaged over time, temporal statistics will have reduced temporal resolution.

Statistical questions regarding the use of weighted versus unweighted statistics have been discussed above. These questions were not addressed in this study. This phase of the study focused on questions concerning the performance of the estimators studied in the earlier (spatial statistics) phase of the study. Specifically, the questions addressed were:

1. Would the MLE estimator continue to be equivalent to the AVG estimator as variance increases due to temporal variability?
2. Would the MED and FNC estimators diverge further from the AVG?

In other words, the goal of this phase of the study was to determine whether the results obtained for spatial statistics would also pertain to temporal statistics. The greatest differences between the MLE and AVG estimators occurred in bins having the highest variance. Since temporal statistics, in general, will have increased variance due to temporal variability within bins, it was not known whether the MLE and AVG estimators would remain equivalent. Furthermore, it was predicted that the small but systematic errors in the MED and FNC estimators would increase with increases in variance.

3.2.1 Methods

Ideally, several time series of CZCS images from different geographic regions should be analyzed to address these questions. However, this approach was not considered feasible. Since CZCS was operated only 10% of the time, its sampling frequency for any bin was much lower than that expected for SeaWiFS, which will operate continuously.

To investigate how phytoplankton pigment distributions vary over time at a fixed location, and to answer the above questions, the Shelf Edge Exchange Program II (SEEP II) moored fluorometer data (Medeiros and Wirick 1992) was analyzed. These data consisted of temporal records of chlorophyll fluorescence from six moored fluorometer arrays located along the outer edge of the continental shelf off the Delmarva Peninsula. The mooring arrays were deployed between February 1988 and May 1989. Details of the SEEP II data are given in Table 4.

At each mooring, a time series of daily satellite-derived surface chlorophyll measurements was simulated by selecting the SEEP measurement closest to 10 AM from the shallowest fluorometer. The depths of these instruments ranged from 16–39 m (see Table 4).

Table 2a. Summary of relative errors for CHL estimators.

Estimator	*Scene*	*Number*	*Bias*	*Error*	*95% Range*	
Used	*Number*	*of Bins*	[%]	(rms) [%]	*Minimum*	*Maximum*
MLE	1	2,750	0.00	0.00	−1	0
	2	1,850	0.00	0.00	−1	0
	3	2,584	−0.03	1.26	−1	0
	4	5,773	−0.03	0.42	−1	0
	5	4,535	−0.05	1.42	−1	0
	6	513	0.07	1.19	−2	2
	7	3,285	−0.21	1.13	−3	1
	Combined	21,290	−0.05	0.95	−1	0
MED	1	2,750	−1.14	1.20	−3	−1
	2	1,850	−1.16	1.25	−3	−1
	3	2,584	−1.08	2.15	−6	0
	4	5,773	−1.98	2.95	−8	−1
	5	4,535	−1.37	2.27	−3	−1
	6	513	−4.67	7.32	−24	0
	7	3,285	−5.13	7.38	−21	−1
	Combined	21,290	−2.11	3.74	−11	−1
AVG4	1	2,750	1.12	7.28	−13	15
	2	1,850	0.68	7.07	−13	15
	3	2,584	−2.00	6.07	−14	8
	4	5,773	−3.12	9.35	−21	13
	5	4,535	0.00	7.31	−14	14
	6	513	0.71	11.79	−22	25
	7	3,285	0.71	16.72	−26	42
	Combined	21,290	−0.76	9.86	−19	18
MLE4	1	2,750	1.12	7.29	−13	15
	2	1,850	0.69	7.06	−13	15
	3	2,584	−1.99	6.06	−14	8
	4	5,773	−3.10	9.26	−21	13
	5	4,535	−0.01	7.25	−14	14
	6	513	0.88	11.86	−22	25
	7	3,285	0.69	16.51	−26	41
	Combined	21,290	−0.75	9.77	−19	18
MED4	1	2,750	0.17	7.14	−14	14
	2	1,850	−0.26	6.95	−14	14
	3	2,584	−2.96	6.50	−16	7
	4	5,773	−4.81	9.95	−24	11
	5	4,535	−1.04	7.21	−15	13
	6	513	−3.50	12.52	−30	18
	7	3,285	−4.07	14.02	−32	24
	Combined	21,290	−2.60	9.38	−22	14

Table 2b. Summary of relative errors for K_{490} estimators.

Estimator Used	*Scene Number*	*Number of Bins*	*Bias* [%]	*Error* (rms) [%]	*95% Range* *Minimum*	*Maximum*
MLE	1	2,757	0.00	0.00	−1	0
	2	1,875	0.00	0.00	−1	0
	3	2,584	0.00	0.02	−1	0
	4	5,773	0.00	0.03	−1	0
	5	4,562	0.00	0.02	−1	0
	6	424	0.00	0.16	−1	0
	7	2,398	0.00	0.05	−1	0
	Combined	20,373	0.00	0.04	−1	0
MED	1	2,757	0.00	0.00	−1	0
	2	1,875	0.00	0.00	−1	0
	3	2,584	−0.05	0.24	−2	0
	4	5,773	−0.16	0.55	−2	0
	5	4,562	−0.12	0.36	−2	0
	6	424	−1.09	2.08	−8	0
	7	2,398	−0.93	1.41	−4	0
	Combined	20,373	−0.21	0.67	−3	0
AVG4	1	2,757	0.26	1.63	−4	3
	2	1,875	0.20	1.71	−4	3
	3	2,584	−0.69	2.12	−5	3
	4	5,773	−1.02	3.45	−8	5
	5	4,562	0.01	3.46	−7	7
	6	424	0.37	5.39	−12	11
	7	2,398	0.06	6.54	−12	16
	Combined	20,373	−0.31	3.59	−8	7
MLE4	1	2,757	0.26	1.63	−4	3
	2	1,875	0.20	1.71	−4	3
	3	2,584	−0.68	2.12	−5	3
	4	5,773	−1.03	3.44	−8	5
	5	4,562	0.01	3.45	−7	7
	6	424	0.37	5.39	−12	11
	7	2,398	0.05	6.52	−12	16
	Combined	20,373	−0.31	3.58	−8	7
MED4	1	2,757	0.21	1.62	−4	3
	2	1,875	0.14	1.71	−4	3
	3	2,584	−0.80	2.16	−6	3
	4	5,773	−1.27	3.51	−9	4
	5	4,562	−0.24	3.41	−7	6
	6	424	−0.75	5.37	−14	8
	7	2,398	−0.81	6.09	−13	13
	Combined	20,373	−0.59	3.50	−8	6

Table 2c. Summary of relative errors for $L_{WN}(\lambda_1)$ estimators.

Estimator Used	*Scene Number*	*Number of Bins*	*Bias* [%]	*Error* (rms) [%]	*95% Range* *Minimum*	*Maximum*
MLE	1	2,750	0.00	0.00	−1	0
	2	1,850	0.00	0.00	−1	0
	3	2,584	0.00	0.00	−1	0
	4	5,773	0.01	0.31	−1	0
	5	4,535	0.01	0.32	−1	0
	6	513	0.00	0.04	−1	0
	7	3,285	0.75	3.39	−1	8
	Combined	21,290	0.12	1.35	−1	0
MED	1	2,750	0.00	0.00	−1	0
	2	1,850	0.00	0.02	−1	0
	3	2,584	−0.03	0.19	−1	0
	4	5,773	−0.20	0.74	−2	0
	5	4,535	−0.08	0.41	−2	0
	6	513	−0.60	1.13	−4	0
	7	3,285	−3.97	9.10	−31	0
	Combined	21,290	−0.70	3.60	−6	0
AVG4	1	2,750	0.35	1.57	−3	3
	2	1,850	0.46	1.73	−4	3
	3	2,584	−0.76	1.93	−5	2
	4	5,773	−1.58	3.53	−9	4
	5	4,535	0.73	3.05	−6	6
	6	513	0.08	4.18	−9	8
	7	3,285	−4.63	12.01	−36	11
	Combined	21,290	−0.99	5.39	−12	5
MLE4	1	2,750	0.35	1.57	−3	3
	2	1,850	0.46	1.73	−4	3
	3	2,584	−0.76	1.93	−5	2
	4	5,773	−1.58	3.53	−9	4
	5	4,535	0.73	3.04	−6	6
	6	513	0.09	4.18	−9	8
	7	3,285	−3.95	12.32	−33	15
	Combined	21,290	−0.89	5.50	−11	5
MED4	1	2,750	0.30	1.57	−3	3
	2	1,850	0.39	1.73	−4	3
	3	2,584	−0.88	2.01	−5	2
	4	5,773	−1.85	3.78	−10	3
	5	4,535	0.52	3.09	−6	6
	6	513	−0.50	4.36	−10	7
	7	3,285	−7.94	16.22	−54	6
	Combined	21,290	−1.66	6.93	−16	5

Table 2d. Summary of relative errors for $L_{WN}(\lambda_2)$ estimators.

Estimator Used	*Scene Number*	*Number of Bins*	*Bias* [%]	*Error* (rms) [%]	*95% Range* *Minimum*	*Maximum*
MLE	1	2,750	0.01	0.12	−1	0
	2	1,850	0.04	0.29	−1	0
	3	2,584	0.00	0.03	−1	0
	4	5,773	0.03	0.22	−1	0
	5	4,535	0.02	0.20	−1	0
	6	513	0.03	0.22	−1	0
	7	3,285	0.03	0.36	−1	0
	Combined	21,290	0.02	0.23	−1	0
MED	1	2,750	−0.04	0.24	−1	0
	2	1,850	−0.15	0.55	−2	0
	3	2,584	−0.09	0.33	−2	0
	4	5,773	−0.27	0.73	−3	0
	5	4,535	−0.11	0.47	−2	0
	6	513	−1.26	2.45	−9	0
	7	3,285	−0.90	2.23	−7	0
	Combined	21,290	−0.30	1.07	−3	0
AVG4	1	2,750	0.43	3.09	−6	6
	2	1,850	0.49	3.61	−7	7
	3	2,584	−1.29	2.80	−7	3
	4	5,773	−2.30	4.55	−11	5
	5	4,535	0.57	3.24	−6	6
	6	513	0.37	5.41	−13	11
	7	3,285	−2.50	6.08	−14	7
	Combined	21,290	−0.94	4.19	−10	6
MLE4	1	2,750	0.44	3.09	−6	6
	2	1,850	0.49	3.59	−7	7
	3	2,584	−1.29	2.79	−7	3
	4	5,773	−2.27	4.51	−11	5
	5	4,535	0.59	3.21	−6	6
	6	513	0.41	5.40	−13	11
	7	3,285	−2.48	6.02	−14	7
	Combined	21,290	−0.92	4.16	−10	6
MED4	1	2,750	0.25	3.13	−7	6
	2	1,850	0.22	3.71	−8	7
	3	2,584	−1.48	2.94	−7	3
	4	5,773	−2.68	5.00	−12	4
	5	4,535	0.34	3.40	−7	6
	6	513	−0.78	5.80	−16	9
	7	3,285	−3.38	6.71	−17	5
	Combined	21,290	−1.32	4.53	−11	6

Table 2e. Summary of relative errors for $L_{WN}(\lambda_3)$ estimators.

Estimator	*Scene*	*Number*	*Bias*	*Error*	*95% Range*	
Used	*Number*	*of Bins*	[%]	(rms) [%]	*Minimum*	*Maximum*
MLE	1	2,750	0.01	0.09	−1	0
	2	1,850	0.03	0.22	−1	0
	3	2,584	0.00	0.03	−1	0
	4	5,773	0.02	0.19	−1	0
	5	4,535	0.00	0.07	−1	0
	6	513	0.06	0.35	−1	1
	7	3,285	0.01	0.25	−1	0
	Combined	21,290	0.01	0.17	−1	0
MED	1	2,750	−0.60	0.79	−2	0
	2	1,850	−0.77	0.95	−2	0
	3	2,584	−0.19	0.50	−2	0
	4	5,773	−0.67	1.02	−3	0
	5	4,535	−0.13	0.38	−2	0
	6	513	−2.09	3.50	−11	0
	7	3,285	−1.02	2.01	−6	0
	Combined	21,290	−0.59	1.19	−3	0
AVG4	1	2,750	1.07	4.75	−9	10
	2	1,850	0.94	4.88	−9	10
	3	2,584	−1.92	3.75	−9	4
	4	5,773	−3.43	6.18	−15	6
	5	4,535	0.77	3.91	−8	8
	6	513	0.57	6.90	−15	14
	7	3,285	−3.16	6.37	−15	7
	Combined	21,290	−1.25	5.26	−13	8
MLE4	1	2,750	1.07	4.74	−9	10
	2	1,850	0.95	4.86	−9	10
	3	2,584	−1.92	3.74	−9	4
	4	5,773	−3.40	6.14	−15	6
	5	4,535	0.77	3.90	−8	8
	6	513	0.65	6.93	−14	14
	7	3,285	−3.14	6.36	−15	7
	Combined	21,290	−1.24	5.24	−13	8
MED4	1	2,750	0.67	4.73	−9	9
	2	1,850	0.49	4.98	−10	9
	3	2,584	−2.24	3.95	−10	3
	4	5,773	−4.07	6.78	−16	5
	5	4,535	0.47	3.92	−8	8
	6	513	−1.26	7.42	−20	11
	7	3,285	−4.08	7.05	−17	6
	Combined	21,290	−1.81	5.63	−14	8

Table 3. Summary of relative errors for $\mathrm{IC_K}$ estimators.

Estimator	Image	Number	Bias	RMS	95% Range	
Used	Number	of Bins	[%]	Error [%]	Minimum	Maximum
MLE	1	2,750	0.00	0.00	−1	0
	2	1,850	0.00	0.00	−1	0
	3	2,584	−0.01	0.45	−1	0
	4	5,752	−0.01	0.10	−1	0
	5	4,533	−0.02	0.48	−1	0
	6	423	−0.01	0.23	−1	0
	7	2,379	−0.04	0.20	−1	0
	Combined	20,271	−0.01	0.30	−1	0
FNC	1	2,750	0.14	0.38	−1	1
	2	1,850	0.23	0.48	−1	1
	3	2,584	0.39	2.61	−1	2
	4	5,752	0.86	1.45	−1	3
	5	4,533	0.80	1.32	−1	1
	6	423	2.57	4.52	−1	13
	7	2,379	1.84	2.53	0	6
	Combined	20,271	0.78	1.76	−1	3
AVG4	1	2,750	0.87	5.74	−11	12
	2	1,850	0.49	5.44	−11	11
	3	2,584	−1.34	3.98	−10	5
	4	5,752	−2.19	6.09	−15	8
	5	4,533	0.07	3.68	−8	7
	6	423	0.61	7.96	−15	16
	7	2,379	−0.30	6.96	−13	16
	Combined	20,271	−0.64	5.45	−12	10
MLE4	1	2,750	0.87	5.74	−11	12
	2	1,850	0.49	5.45	−11	11
	3	2,584	−1.34	3.98	−10	5
	4	5,752	−2.18	6.06	−15	8
	5	4,533	0.07	3.68	−8	7
	6	423	0.62	7.97	−16	17
	7	2,379	−0.33	6.87	−-13	15
	Combined	20,271	−0.64	5.43	−12	10
FNC4	1	2,750	1.19	5.83	−11	12
	2	1,850	0.82	5.52	−10	12
	3	2,584	−0.91	3.92	−9	6
	4	5,752	−1.45	6.10	−14	9
	5	4,533	0.59	3.83	−7	8
	6	423	2.86	10.05	−15	24
	7	2,379	1.28	8.66	−12	24
	Combined	20,271	0.05	5.84	−11	11

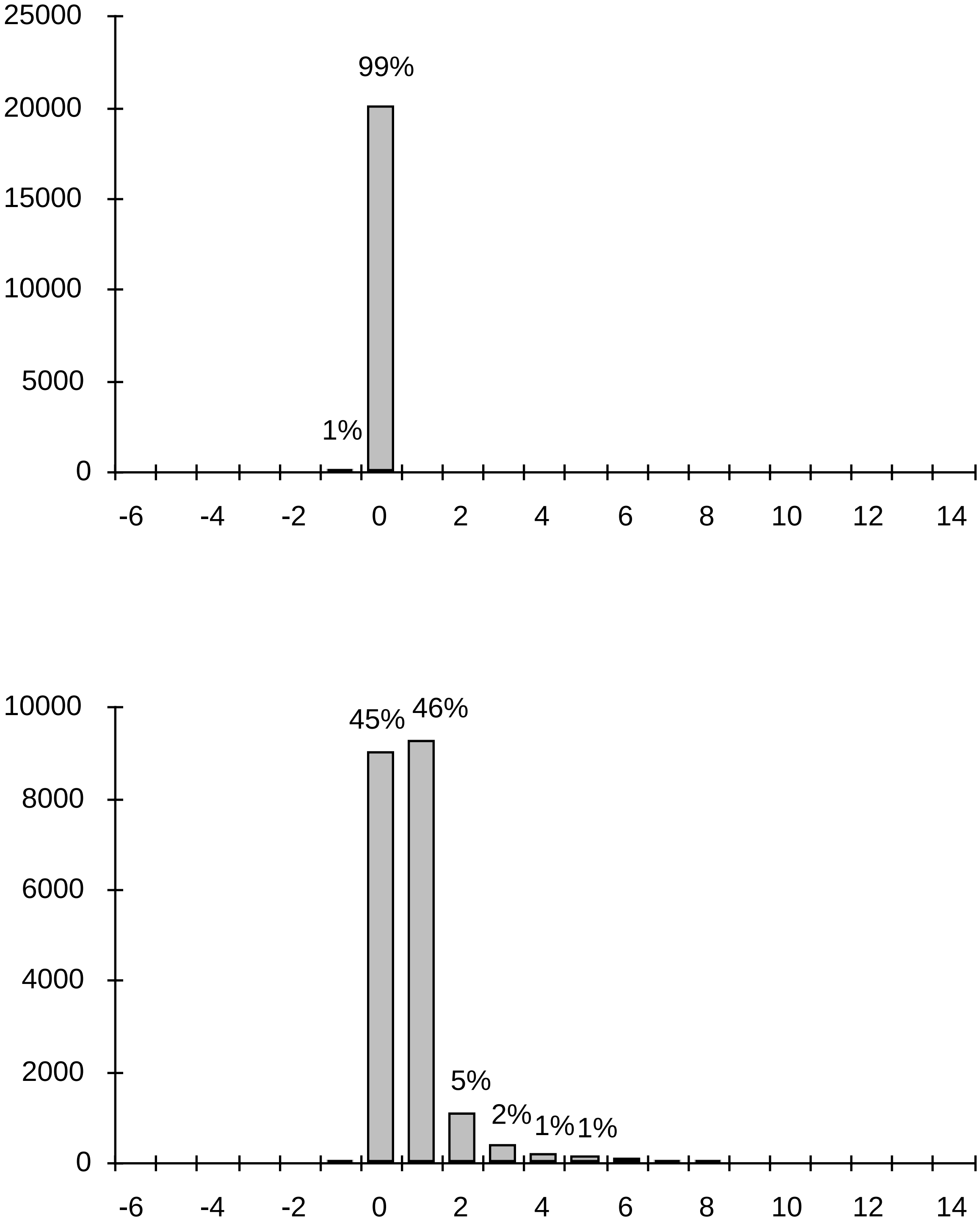


Fig. 15. Histograms of CHL/K_{490} estimation errors based on 20,271 bins analyzed and full resolution (LAC) data. For each bin, the error is defined as the difference between the estimator and the arithmetic average (AVG) of all data in the bin expressed as a percentage of AVG. The FNC estimator is the AVG estimator of CHL divided by the AVG estimator of K_{490}. The top histogram was calculated for (MLE − AVG)/ AVG (%). The bottom histogram was calculated for (FNC − AVG)/ AVG (%).

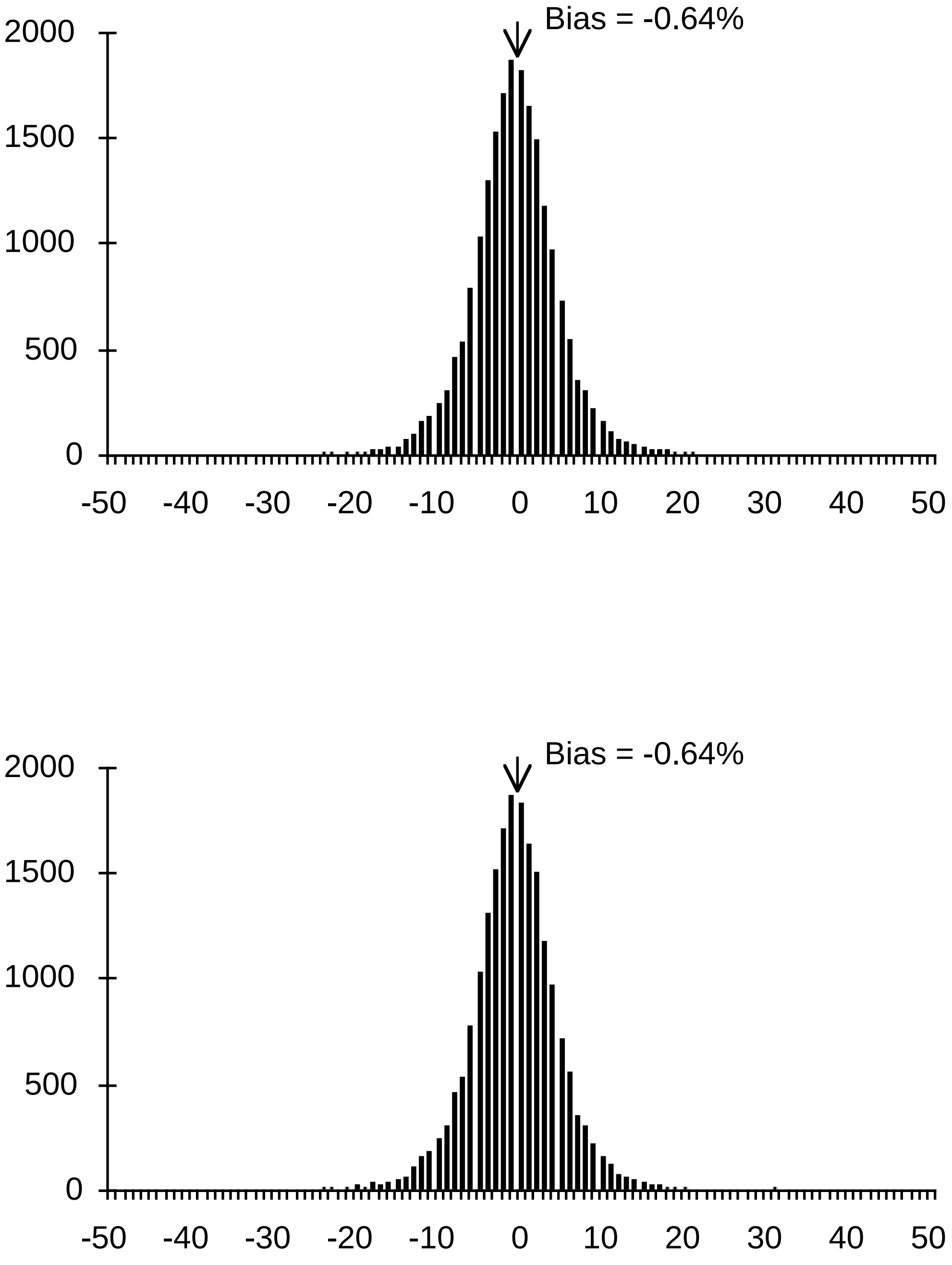


Fig. 16. Histograms of CHL/K_{490} estimation errors based on 20,271 bins analyzed and 4 km subsampled (GAC) data. The top histogram was calculated for (AVG4 − AVG)/ AVG (%). The bottom histogram was calculated for (FNC4 − AVG)/ AVG (%).

Table 4. Location and depth of SEEP II moored fluorometers and the period covered by the time series data used in the analysis of temporal statistics.

		Latitude	*Longitude*	*Depth*	*Time Series*	
ID	*Deployment*	[deg. min.]	[deg. min.]	[m]	*Start*	*Finish*
1	Spring	37 52.60	74 43.90	39	7 Feb 88	8 Apr 88
	Summer	37 52.49	74 43.90	18	25 Jun 88	19 Oct 88
	Winter	37 47.62	74 44.60	19	12 Nov 88	17 Mar 89
2	Spring	37 46.11	74 29.50	16	8 Feb 88	9 Apr 88
	Winter	37 34.69	74 35.13	24	15 Nov 88	28 Jan 89
3	Spring	37 41.99	74 20.35	19	8 Feb 88	9 Jun 88
	Summer	37 41.98	74 20.37	21	25 Jun 88	17 Oct 88
	Winter	37 41.96	74 20.27	19	11 Nov 88	8 May 89
5	Spring	37 39.80	74 15.85	21	8 Feb 88	7 Jun 88
	Summer	37 39.78	74 15.72	22	26 Jun 88	17 Oct 88
	Winter	37 39.73	74 15.78	21	15 Nov 88	2 May 89
6	Spring	37 37.91	74 12.86	20	12 Feb 88	7 Jun 88
	Summer	37 37.90	74 12.87	20	25 Jun 88	19 Oct 88
	Winter	37 37.95	74 12.77	35	13 Nov 88	6 May 89
8	Spring	36 52.63	74 39.04	22	13 Feb 88	8 Jun 88

K_{490} was derived from the chlorophyll measurement by the formula

$$K_{490} = 0.022 + 0.079\,\mathrm{CHL}^{0.875}. \qquad (43)$$

This is the relationship between K_{490} (34) and CHL_{13} (35). In the CZCS imagery analyzed, this relationship would hold for most of the data since CHL equals CHL_{13} in most pixels.

Weekly and monthly means of CHL and K_{490} were derived using the AVG, MLE, and MED estimators. When sample sizes are small (e.g., $n \leq 7$), the effect of sample size dominates the error statistics. To control for this effect in weekly means, only weeks having 7 days, i.e., no missing data, were analyzed. However, because there were fewer months, all months were analyzed, regardless of their sample size. The AVG estimator was regarded as the *true* mean. Errors for the MLE and MED estimators were expressed as a percentage of the AVG estimator.

Weekly and monthly means of the function $\mathrm{IC_K}$ (1) were also derived. Estimators compared with the AVG or *true* mean were the MLE and MED estimators, and an FNC estimator defined in two ways:

$$\mathrm{FNC(AVG)} = \frac{\text{AVG estimator of CHL}}{\text{AVG estimator of } K_{490}} \qquad (44)$$

and

$$\mathrm{FNC(MLE)} = \frac{\text{MLE estimator of CHL}}{\text{MLE estimator of } K_{490}}. \qquad (45).$$

The FNC(MLE) estimator would be applicable if, as recommended, spatial statistics are derived according to the MLE estimator.

To investigate the behavior of the AVG, MLE, MED, and FNC estimators as samples sizes increase over time, cumulative means were obtained as follows:

AVG(n) arithmetic average of all data from days 1 to n,

MLE(n) MLE estimate based on data from days 1 to n,

MED(n) MED estimate based on data from days 1 to n, and

FNC(n) FNC estimate based on data from days 1 to n.

The cumulative means began day 1 at the start of each deployment. Since each mooring had up to three separate deployments (see Table 4), there were 1–3 sets of cumulative means for each mooring. These were plotted against n to observe how the estimators behaved as a function of sample size.

In a similar manner, the behavior of the estimators as functions of area were investigated using CZCS data. Beginning at one or two selected locations in each CZCS scene (Table 1), the estimators were calculated for bins of increasing area $L \times L$, with L increasing from 9 km to the size of the image. The maximum value of L was 480 km. Increases in area may be regarded as analogous to increases in time. To the extent that this is true, these results would pertain to the estimation of temporal means.

3.2.2 Results

Histograms of log(CHL) from each mooring are shown in Fig. 17. Based on the normal (Gaussian) appearance of these histograms, the distribution of chlorophyll over time at a single location is approximately lognormal, or a mixture of lognormals.

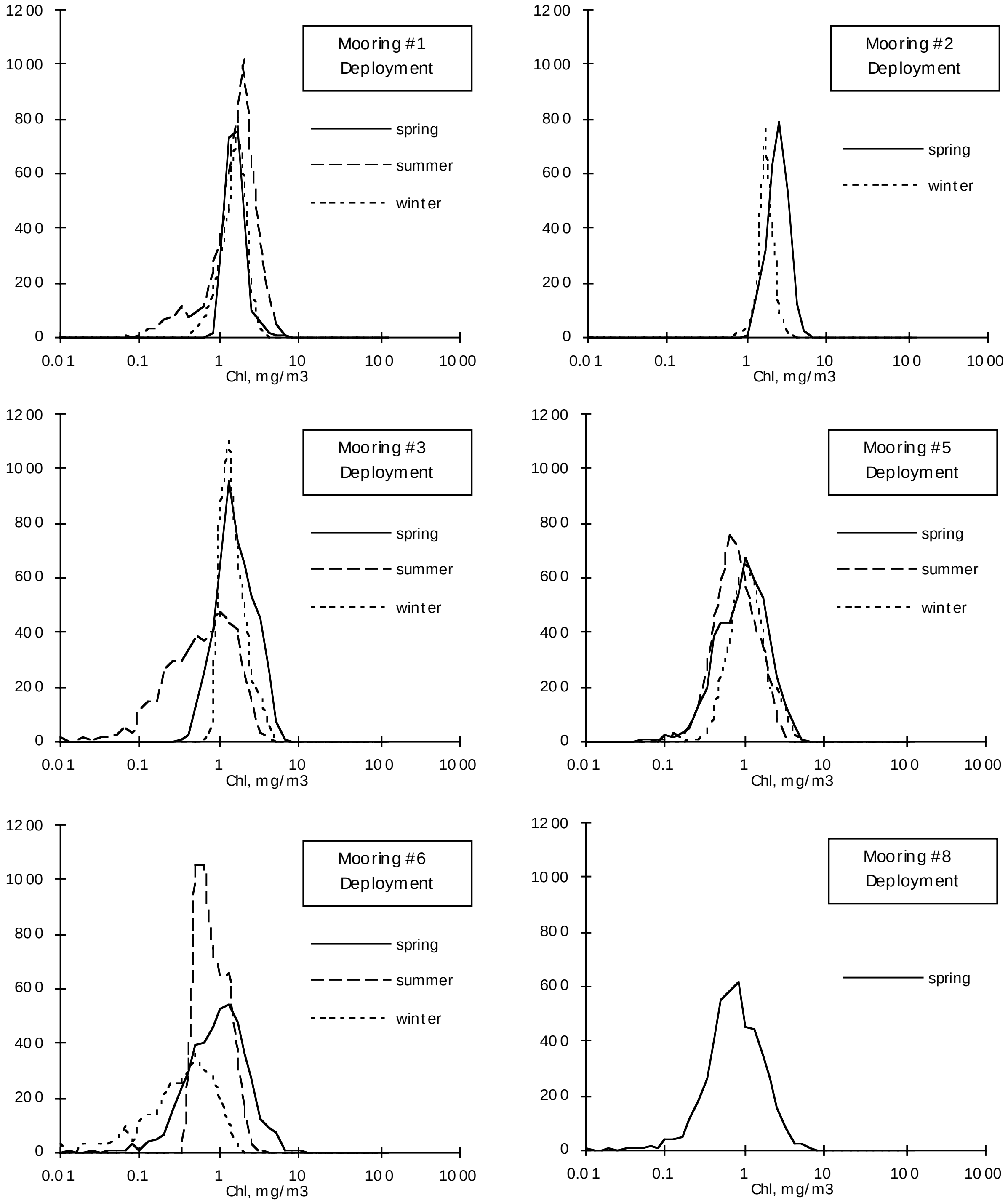


Fig. 17. Histograms of CHL from SEEP II moored fluorometer data. Data are from the shallowest fluorometer at each mooring. All data from moorings 1, 2, 3, 5, 6, and 8 are included in these histograms.

Histograms of relative errors in weekly and monthly means are shown in Figs. 18–25 and summarized in Table 5. The upper panel in each figure is the error histogram for weekly means, and the lower panel is for monthly means. The patterns seen in Fig. 18 for the MLE estimator of CHL were similar to those obtained for the MLE estimators of the other variables. The MLE and AVG estimates agreed within ±5% most of the time, with a slight tendency for MLE to exceed AVG, as indicated by the small positive biases (usually much less than 1%) in all cases. As in the case of the CZCS data, the K_{490} and $\mathrm{IC_K}$ variables had much smaller MLE errors than the CHL variable.

The MED estimator had relatively large negative errors for all three variables. That is, the MED estimator underestimated the arithmetic average by 40% or more in some cases, and monthly mean errors were about a factor of 2 greater than weekly mean errors.

Errors that are associated with the FNC(AVG) and FNC(MLE) estimators are shown in Figs. 24 and 25, respectively, and are summarized in Table 5d. Both distributions are positively skewed, with errors as high as 30% or more. The FNC(MLE) estimator had larger errors than the FNC(AVG) estimator.

Results for cumulative means provided important insight concerning the behavior of the estimators, in particular when the MLE and AVG estimators were substantially different. These insights will be illustrated here with results from moorings 3 and 6. Figure 26 shows cumulative mean CHL estimates from the spring deployments of mooring 3 (upper panel) and mooring 6 (lower panel). In the case of mooring 3, MLE(n) and AVG(n) remained approximately equal over the entire averaging period, whereas MED(n) was always less than the other two and gradually diverged as the averaging period increased. These results are typical of what was obtained for the majority of the cases.

The lower panel in Fig. 26 illustrates a case where MLE(n) and AVG(n) diverged. The two cumulative means showed an abrupt divergence at about day 70; prior to that day, they had been nearly equal. Inspection of the data (Fig. 27, upper panel) revealed that there were a number of anomalously low values beginning after day 60. The dark squares in Fig. 27 were data that were missing from the original records. These had been set to zero and were ignored when calculating cumulative means. However, the open squares lying near the horizontal axis were small positive values (e.g., 0.01, 0.02, etc.) which may have also been bad data. If these are eliminated from the record, then MLE(n) and AVG(n) agree (bottom panel of Fig. 27).

This suggests that the MLE estimator can be sensitive to outliers, particularly outliers that are close to zero. When a data value approximately equal to zero is included in the arithmetic average of n values, the effect is to reduce the AVG estimator by a factor of $(n-1)/n$. However, the logarithm of a number approximately equal to zero is a large negative number, and its effect on the statistics of the logarithm can be extreme. Including this value will reduce the mean of the logarithm but increase the variance, somewhat offseting effects on the MLE estimator. In general, however, the net effect will be to increase the MLE estimator since the variance of the logarithm is increased substantially by the inclusion of a large negative value.

Another case in which MLE(n) and AVG(n) diverged was the summer deployment of mooring 3. The simulated satellite data from this record are shown in the upper panel of Fig. 28, and the cumulative means in the lower panel. Like the previous example, there were a number of low values in the record. However, it is not obvious that these are bad data, and so there is no justification for removing them to make MLE(n) and AVG(n) agree. MLE(n) was approximately 10% higher than AVG(n) for $n \geq 35$ days. The cumulative means of K_{490} and $\mathrm{IC_K}$ for this mooring are shown in Fig. 29. Differences between MLE(n) and AVG(n) for these variables were much smaller than those for CHL. However, the two FNC estimates were consistently higher than AVG(n) and MLE(n), with differences approaching 30% by the end of the record.

Cumulative means starting at two locations in CZCS scene 4 are illustrated in Fig. 30 (LAC means) and Fig. 31 (GAC means). In these figures, the cumulative mean CHL within areas of size L^2 is plotted against L. In the northern portion of scene 4 (off the west coast of Mexico), the MLE and AVG cumulative means diverged at length scales larger than 50 km. However, in the southern region of this scene, the MLE and AVG means remained nearly equal for areas up to $460 \times 460\,\mathrm{km}^2$. Results for all the CZCS scenes are summarized in Table 6. Whenever the MLE and AVG estimators diverged for CZCS cumulative means, the AVG estimator was greater than the MLE estimator. This occurred in the scenes that had high chlorophyll levels and/or high variances. In contrast, when the MLE and AVG estimators in SEEP data diverged, the MLE estimator was usually greater than the AVG estimator.

3.3 Discussion

From the study of CZCS and SEEP II data, it was concluded that the AVG and MLE estimators are equivalent with respect to their accuracy as estimators of means within sampling domains. The MED and FNC estimators are not considered acceptable as estimators of the mean. The MED estimator systematically underestimated the mean, and the magnitude of its error increased with increasing intrabin variance. The FNC estimator, i.e., the result of substituting a mean into a function to derive a *level-4* variable, also had systematic errors that increased with increasing variance.

In the case of satellite data from the same scene (spatial statistics), the MLE estimator proved to be nearly identical to the AVG estimator when sample sizes were large ($n = 121$). The same was true for the MLE4 and

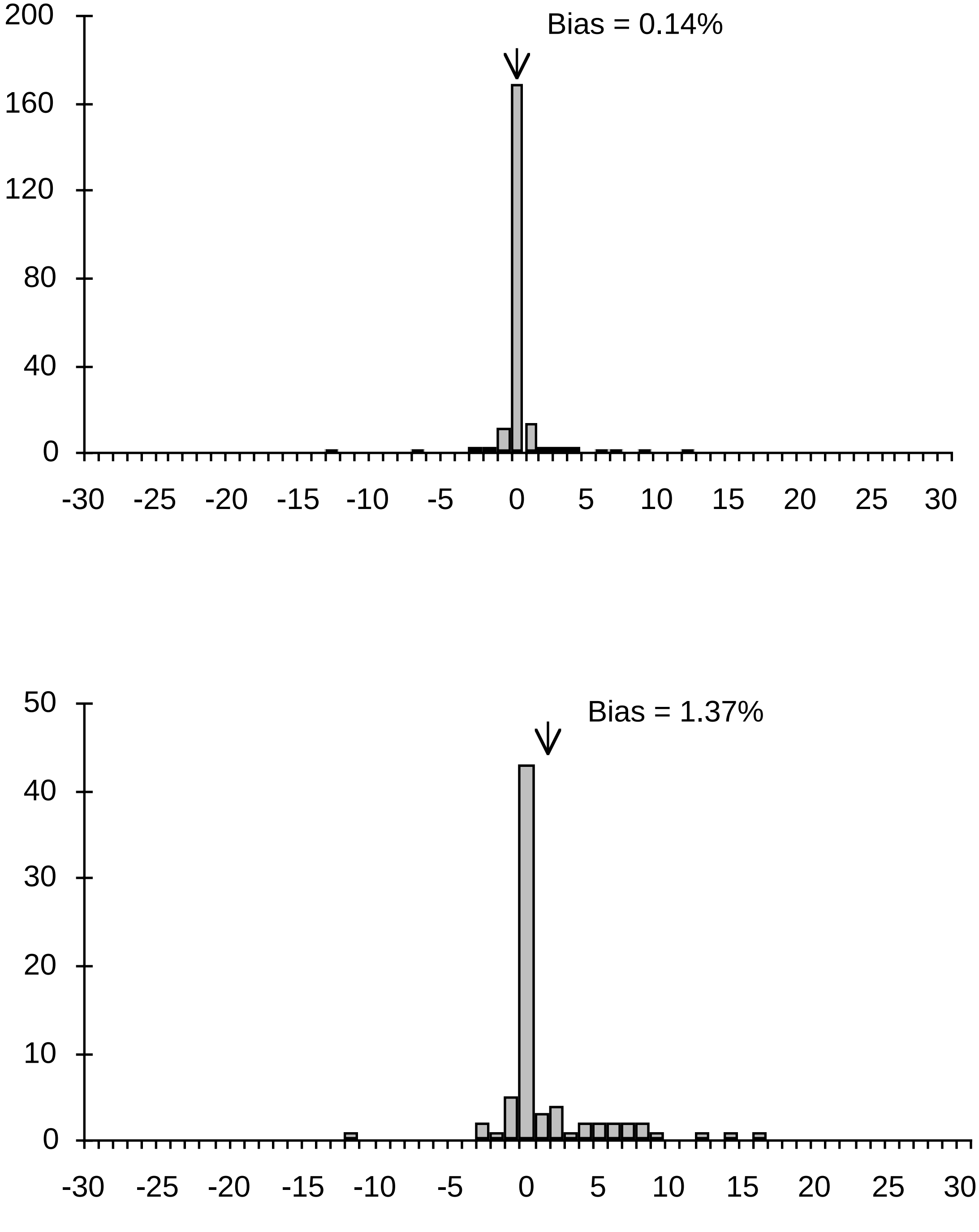


Fig. 18. Histograms of the relative error in MLE estimates of mean CHL at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% × (MLE − AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% × (MLE − AVG)/ AVG.

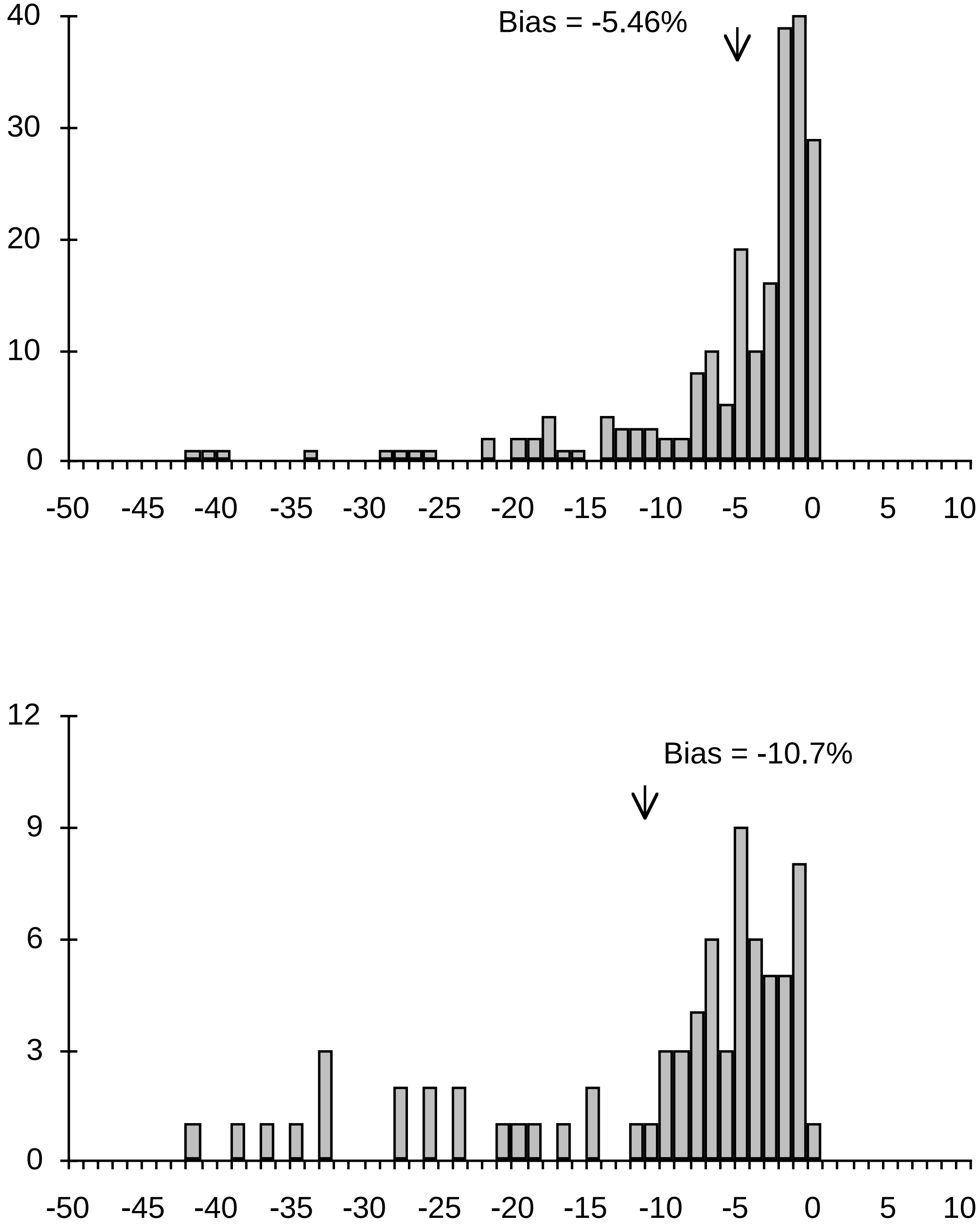


Fig. 19. Histograms of the relative error in MED estimates of mean CHL at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% × (MED − AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% × (MED − AVG)/ AVG.

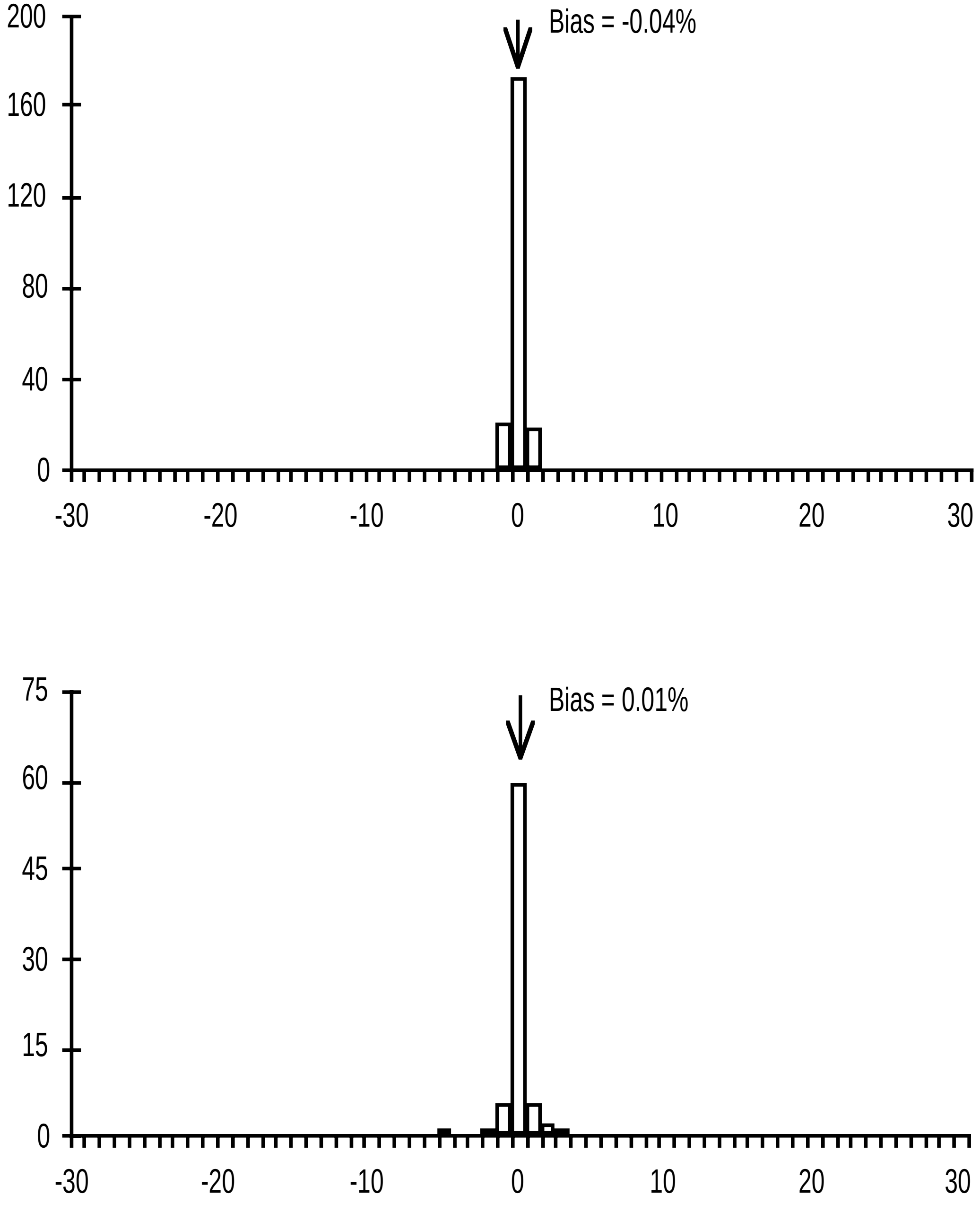


Fig. 20. Histograms of the relative error in MLE estimates of mean K_{490} at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% $\times$ (MLE $-$ AVG)/AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% $\times$ (MLE $-$ AVG)/ AVG.

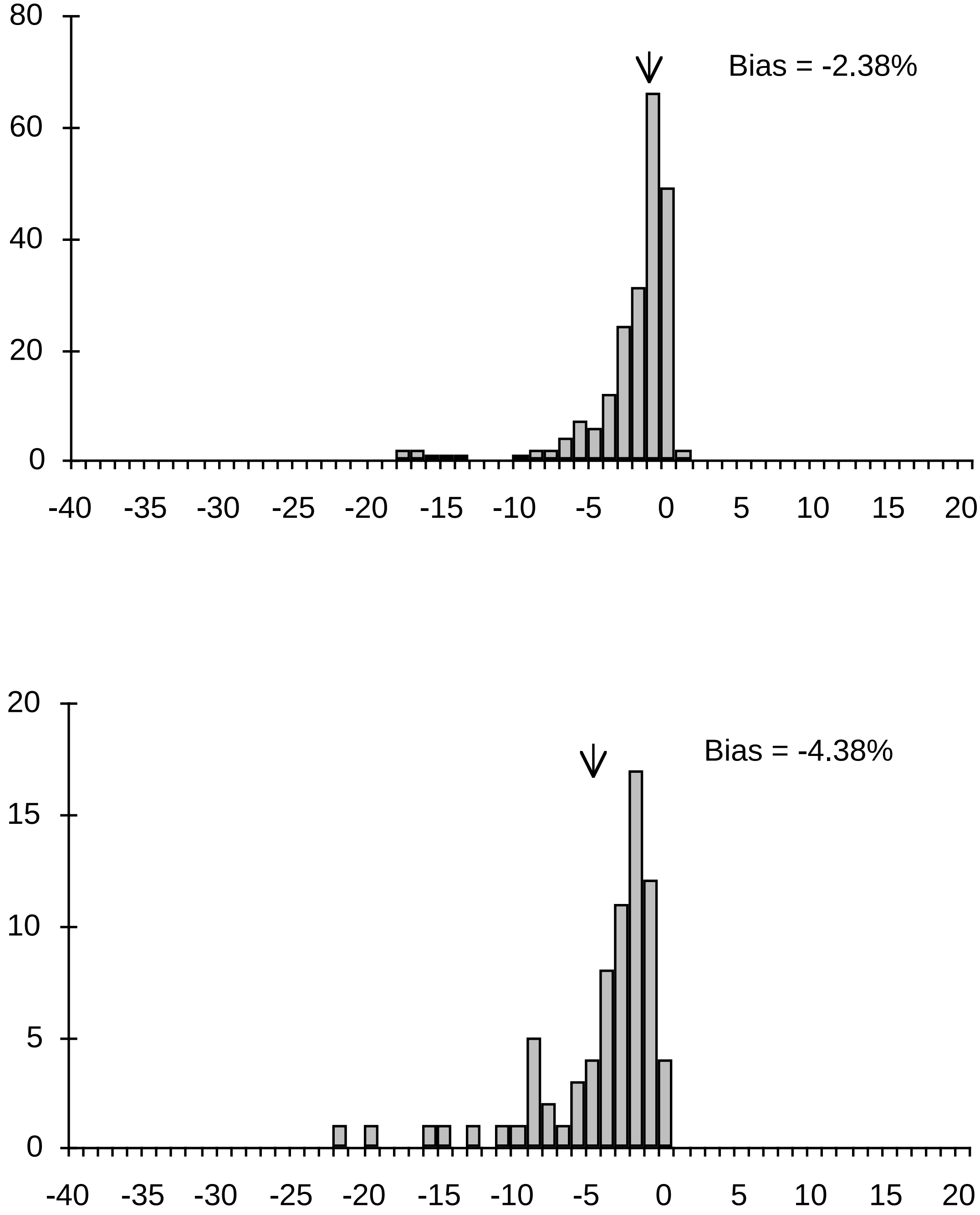


Fig. 21. Histograms of the relative error in MED estimates of mean K_{490} at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% × (MED − AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% × (MED − AVG)/ AVG.

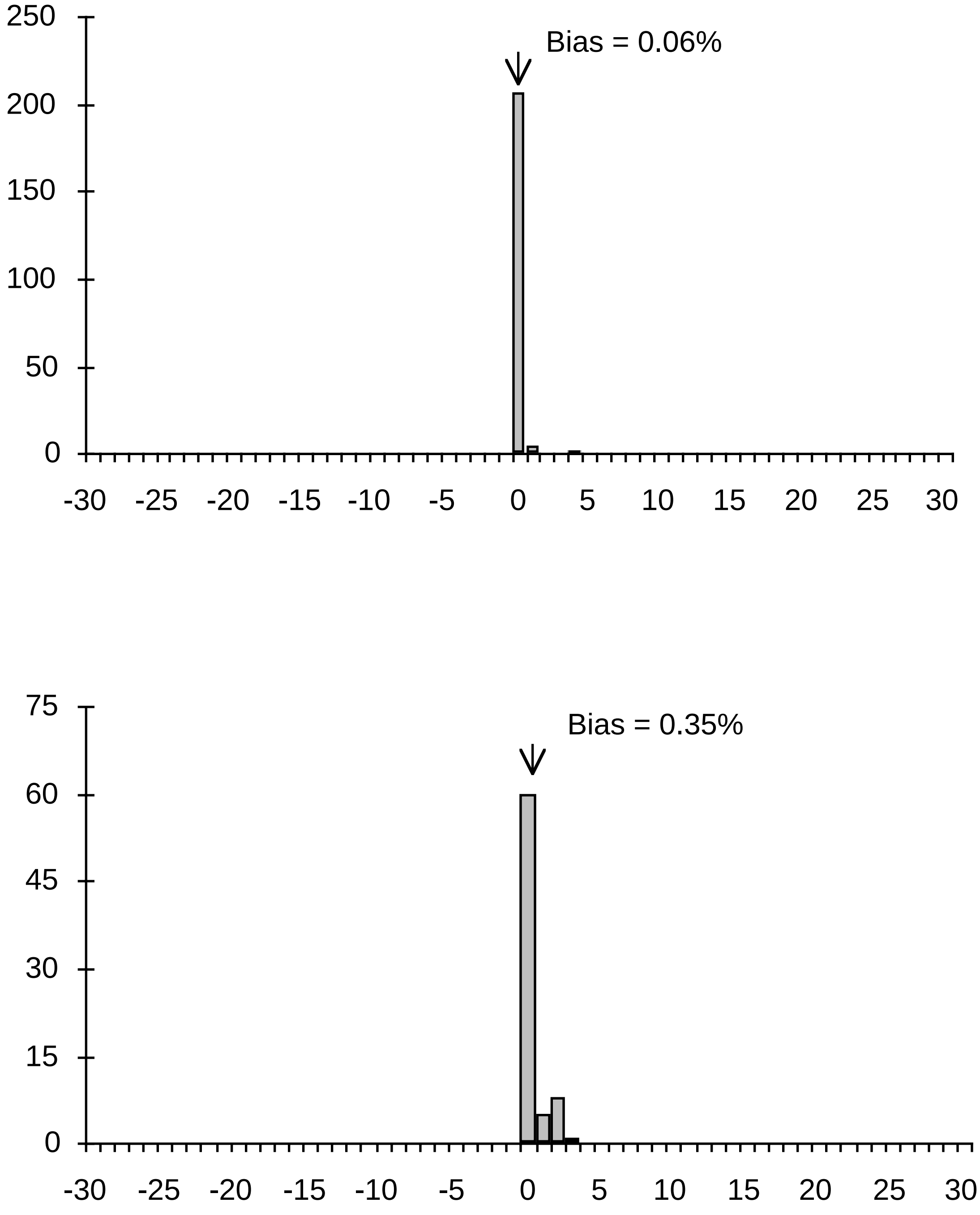


Fig. 22. Histograms of the relative error in MLE estimates of mean CHL/K_{490} at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% × (MLE − AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% × (MLE − AVG)/ AVG.

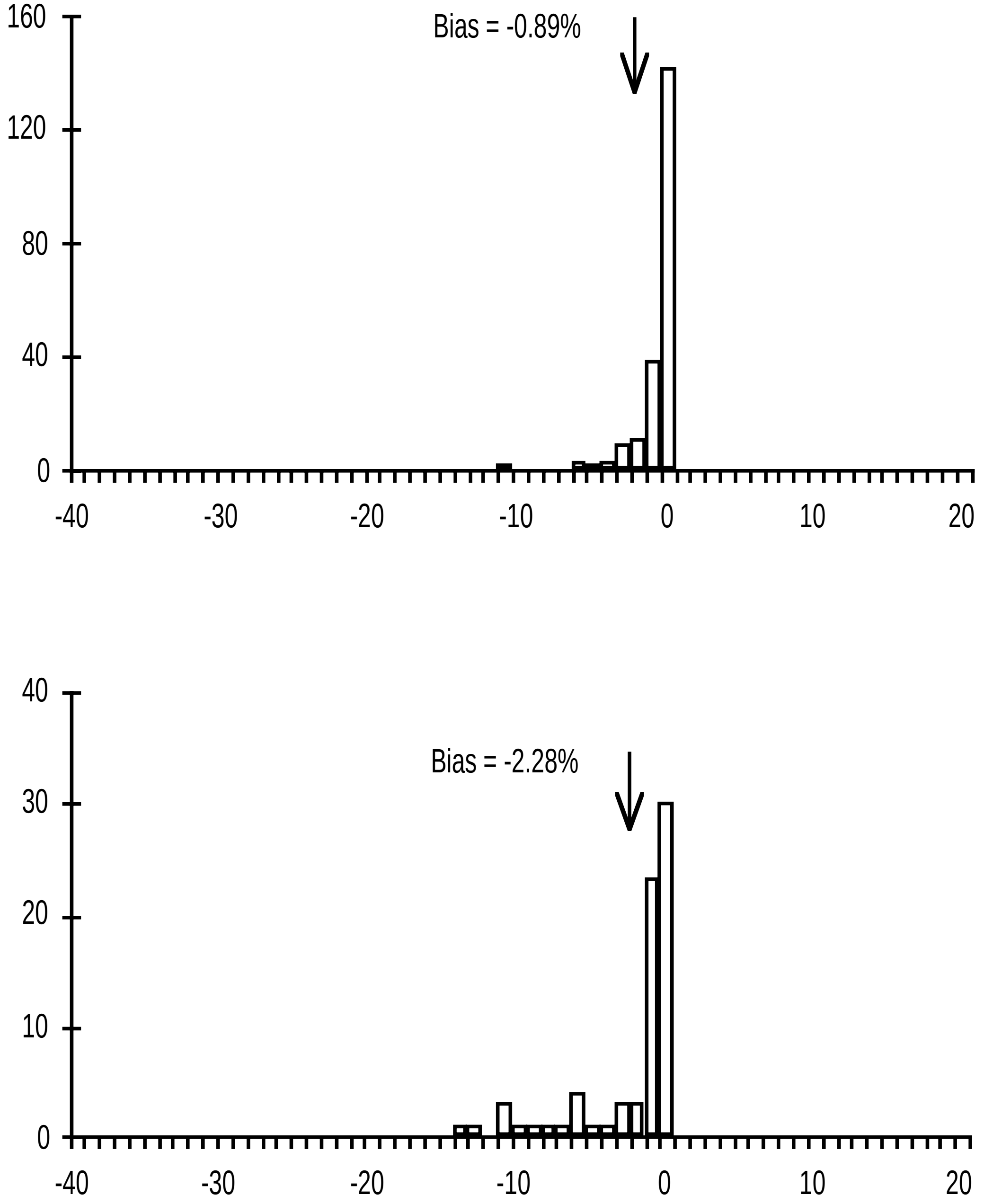


Fig. 23. Histograms of the relative error in MED estimates of mean CHL/K_{490} at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% $\times$ (MED $-$ AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% $\times$ (MED $-$ AVG)/ AVG.

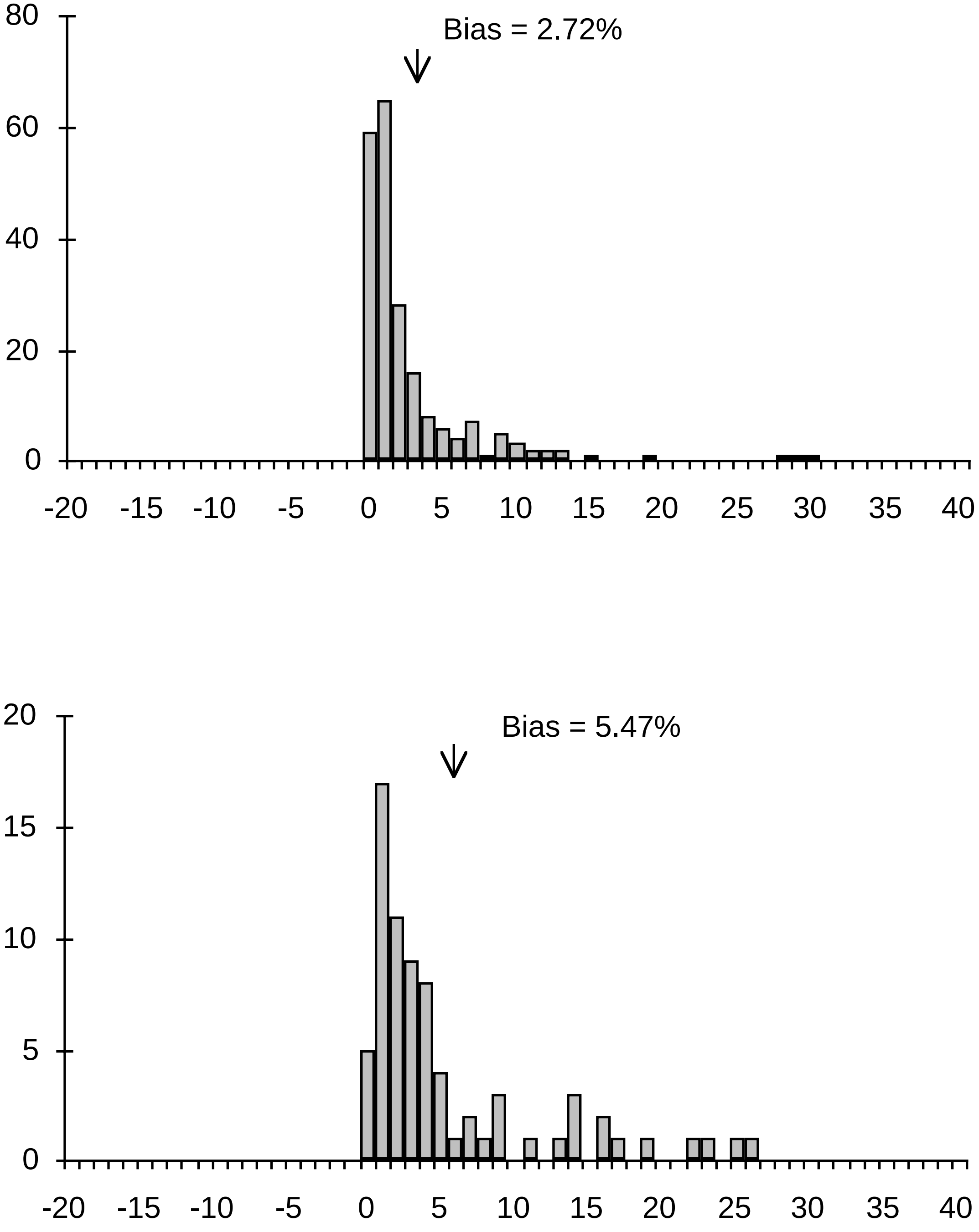


Fig. 24. Histograms of the relative error in FNC(AVG) estimates of mean CHL/K_{490} at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% $\times$(FNC $-$ AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% $\times$(FNC $-$ AVG)/ AVG.

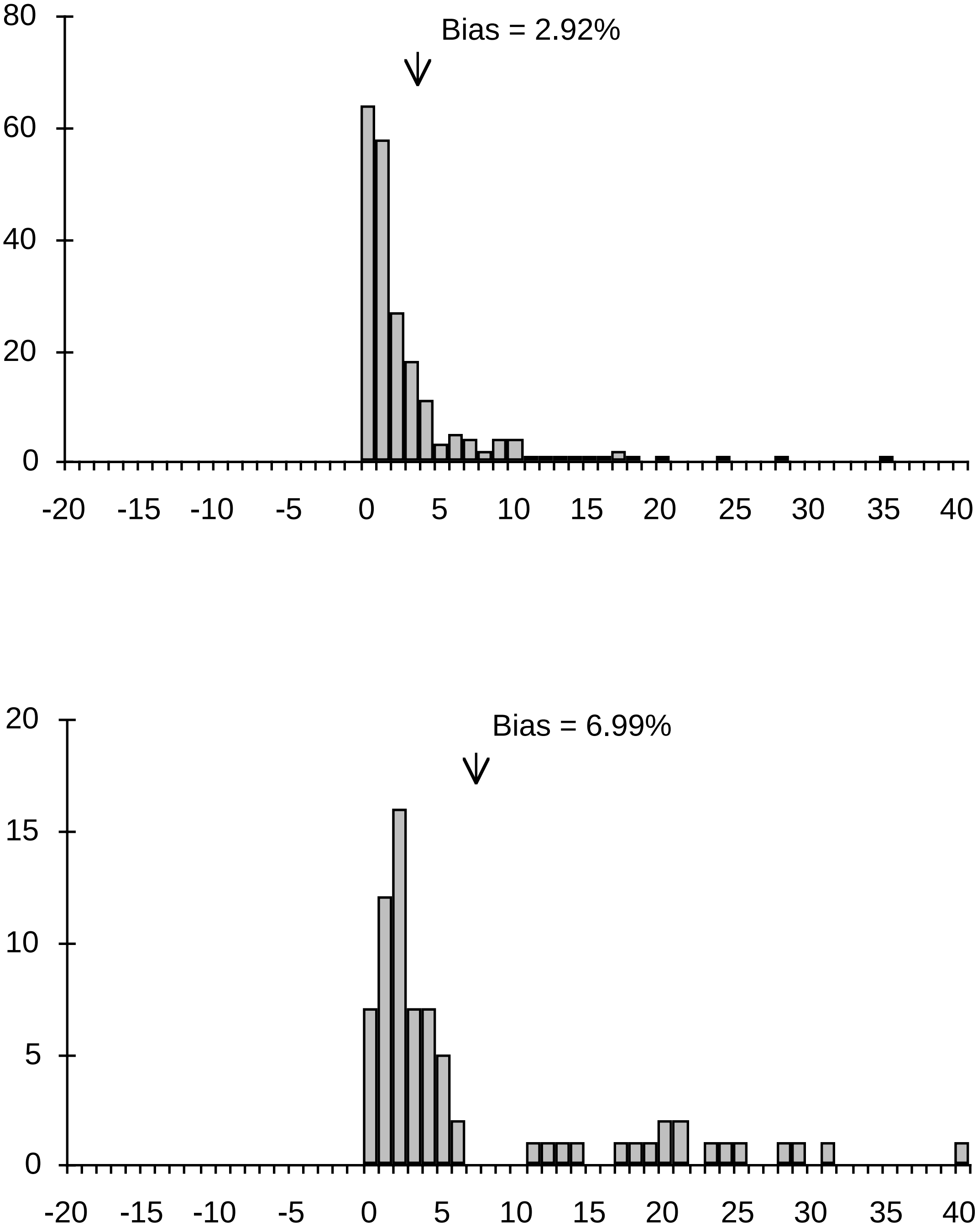


Fig. 25. Histograms of the relative error in FNC(MLE) estimates of mean CHL/K_{490} at SEEP moorings. The top histogram is for the weekly means ($n = 213$), calculated with 100% $\times$(FNC − AVG)/ AVG. The bottom panel is for the monthly means ($n = 74$), also calculated with 100% $\times$(FNC − AVG)/ AVG.

Table 5a. Summary of relative errors for weekly means derived from SEEP mooring data: results for CHL.

Estimator	*Mooring*	*Number*	*Bias*	*Error*	*Range*	
Used	*Reference*	*of Weeks*	[%]	(rms) [%]	*Minimum*	*Maximum*
MLE	1	39	−0.33	2.64	−13	4
	2	18	0.02	0.06	0	0
	3	53	0.18	1.28	−2	9
	5	53	0.24	1.21	−2	6
	6	37	0.47	2.27	−3	12
	8	13	0.26	2.33	−3	7
	Combined	213	0.14	1.79	−13	12
MED	1	39	−5.32	10.83	−41	0
	2	18	−1.17	1.44	−3	0
	3	53	−4.01	7.26	−27	0
	5	53	−6.07	9.06	−42	0
	6	37	−7.15	11.07	−34	0
	8	13	−10.51	12.81	−22	−1
	Combined	213	−5.46	9.24	−42	0

Table 5a. (cont.) Summary of relative errors for monthly means derived from SEEP mooring data: results for CHL.

Estimator	*Mooring*	*Number*	*Bias*	*Error*	*Range*	
Used	*Reference*	*of Months*	[%]	(rms) [%]	*Minimum*	*Maximum*
MLE	1	13	−1.69	4.90	−1	16
	2	6	0.06	0.18	0	0
	3	17	0.91	4.09	−3	14
	5	16	0.98	2.53	−2	8
	6	17	2.29	6.02	−12	12
	8	5	1.75	2.88	0	5
	Combined	74	1.37	4.18	−12	16
MED	1	13	−7.59	13.75	−42	−1
	2	6	−2.27	2.70	−5	0
	3	17	−10.07	15.02	−37	−1
	5	16	−9.24	11.10	−28	−1
	6	17	−17.55	21.92	−39	−1
	8	5	−11.92	15.18	−24	−3
	Combined	74	−10.66	15.24	−42	0

Table 5b. Summary of relative errors for weekly means derived from SEEP mooring data: results for K_{490}.

Estimator Used	*Mooring Reference*	*Number of Weeks*	*Bias* [%]	*Error* (rms) [%]	*Range* *Minimum*	*Maximum*
MLE	1	39	−0.18	0.89	−4	1
	2	18	−0.03	0.15	0	0
	3	53	0.03	0.29	−1	1
	5	53	−0.01	0.41	−1	1
	6	37	−0.11	0.57	−1	1
	8	13	0.06	0.70	−1	2
	Combined	213	−0.04	0.54	−4	2
MED	1	39	−2.52	4.72	−17	0
	2	18	−0.72	0.90	−2	0
	3	53	−1.65	3.05	−15	0
	5	53	−2.62	3.81	−17	0
	6	37	−2.92	4.95	−18	1
	8	13	−4.69	5.75	−10	0
	Combined	213	−2.38	4.02	−18	1

Table 5b. (cont.) Summary of relative errors for monthly means derived from SEEP mooring data: results for K_{490}.

Estimator Used	*Mooring Reference*	*Number of Months*	*Bias* [%]	*Error* (rms) [%]	*Range* *Minimum*	*Maximum*
MLE	1	13	0.27	0.75	0	3
	2	6	−0.09	0.26	0	0
	3	17	−0.14	0.70	−2	2
	5	16	0.20	0.47	0	1
	6	17	−0.20	1.36	−5	2
	8	5	0.15	0.56	−1	1
	Combined	74	0.01	0.82	−5	3
MED	1	13	−3.94	7.19	−22	0
	2	6	−1.42	1.68	−3	−1
	3	17	−4.49	6.01	−15	−1
	5	16	−4.06	5.06	−13	0
	6	17	−5.79	7.88	−20	0
	8	5	−4.98	6.43	−10	−2
	Combined	74	−4.38	6.25	−22	0

Table 5c. Summary of relative errors for weekly means derived from SEEP mooring data: results for $\mathrm{IC_K}$.

Estimator	*Mooring*	*Number*	*Bias*	*Error*	*Range*	
Used	*Reference*	*of Weeks*	[%]	(rms) [%]	*Minimum*	*Maximum*
MLE	1	39	0.00	0.13	−1	0
	2	18	−0.01	0.03	0	0
	3	53	0.09	0.55	0	4
	5	53	0.04	0.11	0	1
	6	37	0.16	0.69	0	4
	8	13	0.05	0.23	0	1
	Combined	213	0.06	0.40	−1	4
MED	1	39	−0.73	1.93	−9	0
	2	18	−0.08	0.10	0	0
	3	53	−0.74	2.09	−11	0
	5	53	−0.95	1.90	−11	0
	6	37	−1.40	2.71	−12	0
	8	13	−1.44	1.96	−5	0
	Combined	213	−0.89	2.03	−12	0

Table 5c. (cont.) Summary of relative errors for monthly means derived from SEEP mooring data: results for $\mathrm{IC_K}$.

Estimator	*Mooring*	*Number*	*Bias*	*Error*	*Range*	
Used	*Reference*	*of Months*	[%]	(rms) [%]	*Minimum*	*Maximum*
MLE	1	13	0.12	0.33	0	1
	2	6	−0.03	0.03	0	0
	3	17	0.25	0.70	0	2
	5	16	0.11	0.29	0	1
	6	17	0.95	1.51	0	3
	8	5	0.40	0.86	0	2
	Combined	74	0.35	0.84	0	3
MED	1	13	−1.15	2.66	−9	0
	2	6	−0.17	0.20	0	0
	3	17	−1.91	3.89	−11	0
	5	16	−1.40	1.91	−6	0
	6	17	−5.09	7.13	−14	0
	8	5	−2.26	3.12	−6	−1
	Combined	74	−2.28	4.17	−14	0

Table 5d. Summary of relative errors for weekly means derived from SEEP mooring data: results for FNC estimators of IC_{K}.

Estimator	*Mooring*	*Number*	*Bias*	*Error*	*Range*	
Used	*Reference*	*of Weeks*	[%]	(rms) [%]	*Minimum*	*Maximum*
FNC(AVG)	1	39	2.90	7.25	0	30
	2	18	0.40	0.52	0	1
	3	53	1.90	3.63	0	13
	5	53	3.01	5.16	0	28
	6	37	3.53	5.71	0	19
	8	13	5.32	6.62	0	11
	Combined	213	2.72	5.24	0	30
FNC(MLE)	1	39	2.65	5.85	0	24
	2	18	0.45	0.56	0	1
	3	53	2.07	4.32	0	18
	5	53	3.29	6.06	0	35
	6	37	4.16	7.28	0	28
	8	13	5.53	7.23	1	17
	Combined	213	2.92	5.66	0	35

Table 5d. (cont.) Summary of relative errors for monthly means derived from SEEP mooring data: results for FNC estimators of IC_{K}.

Estimator	*Mooring*	*Number*	*Bias*	*Error*	*Range*	
Used	*Reference*	*of Months*	[%]	(rms) [%]	*Minimum*	*Maximum*
FNC(AVG)	1	13	3.62	7.23	0	23
	2	6	0.70	0.90	0	1
	3	17	5.09	8.59	0	25
	5	16	4.46	5.43	1	14
	6	17	9.76	12.62	1	26
	8	5	5.96	7.85	1	13
	Combined	74	5.47	8.45	0	26
FNC(MLE)	1	13	5.28	12.14	0	40
	2	6	0.86	1.02	0	2
	3	17	6.28	11.26	0	31
	5	16	5.32	7.42	0	23
	6	17	12.52	16.30	0	29
	8	5	7.72	10.60	2	18
	Combined	74	6.99	11.46	0	40

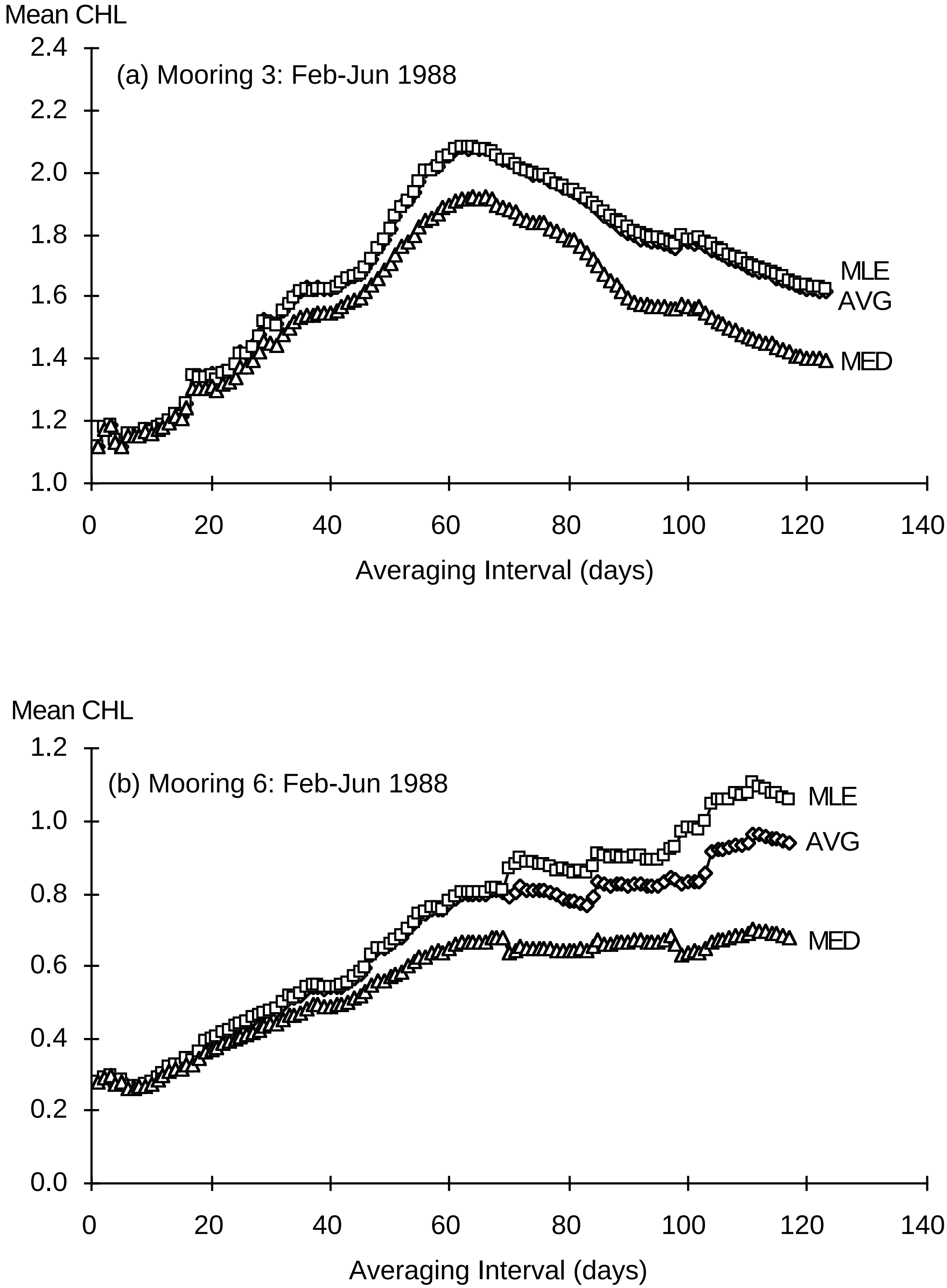


Fig. 26. Cumulative mean CHL estimates for data from the spring 1988 deployment of SEEP moorings 3 and 6.

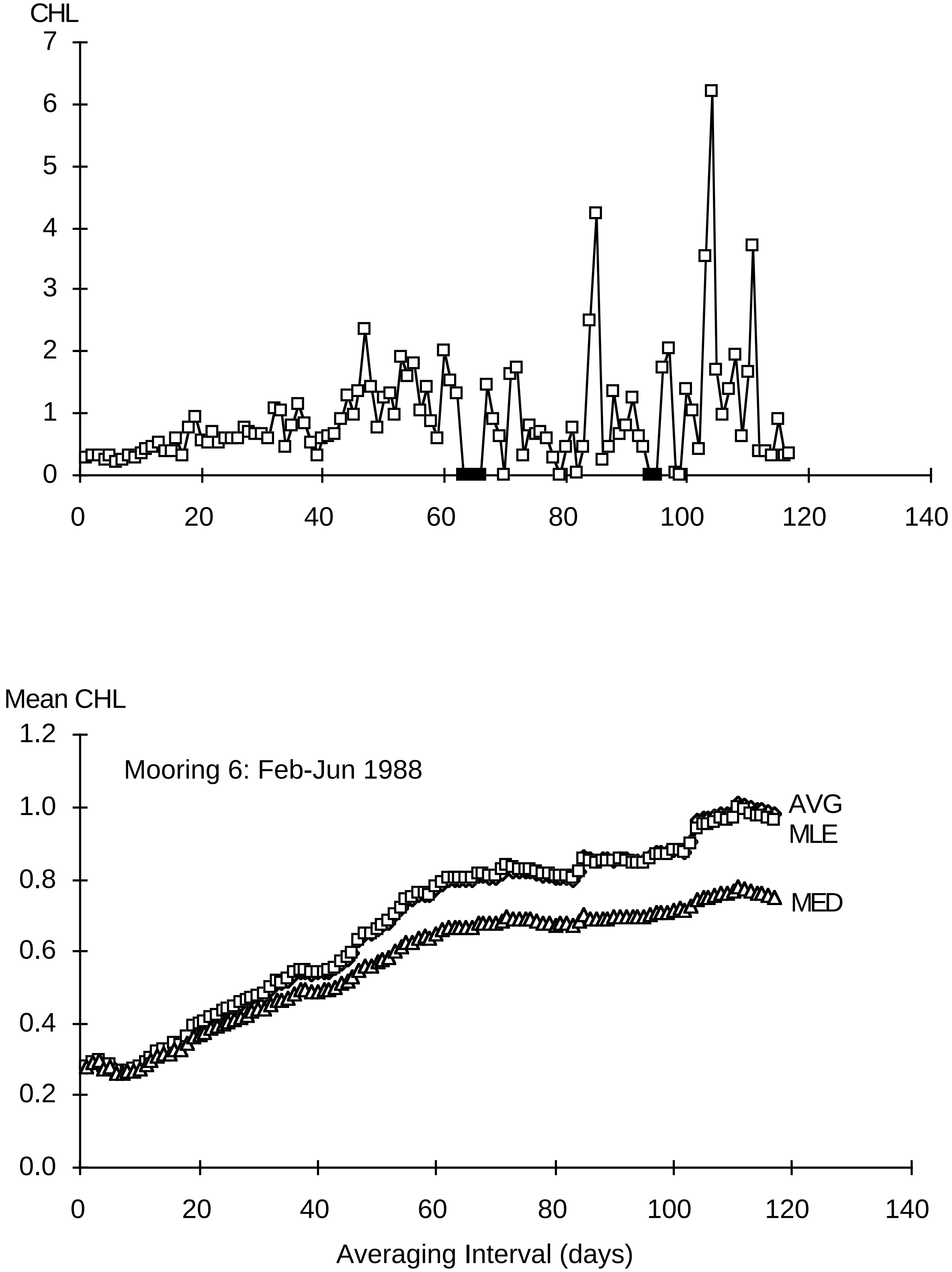


Fig. 27. Results of correcting for *bad data* in the time series at mooring 6 (Fig. 26). In the upper panel, which displays a record of the 10 AM CHL measurements versus time at mooring 6, the dark squares are data missing from the original record. The open squares near zero are probably bad data. Cumulative means derived after removing these low values are shown in bottom panel.

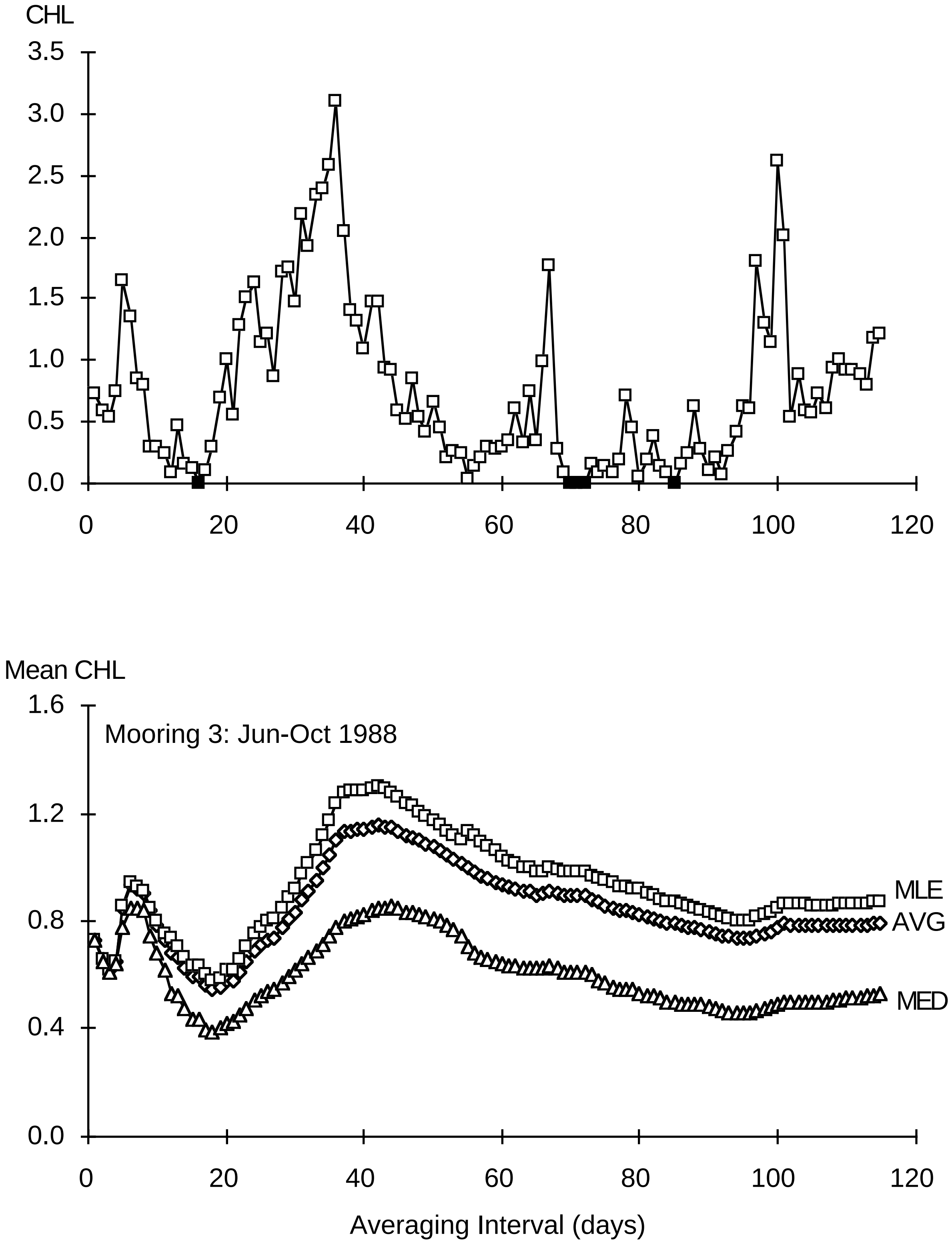


Fig. 28. Simulated satellite CHL time series and cumulative mean CHL for summer deployment of mooring 3 (Jun.–Oct. 1988). The upper panel displays 10 AM CHL measurements, and the lower panel displays cumulative means for the data shown in the upper panel.

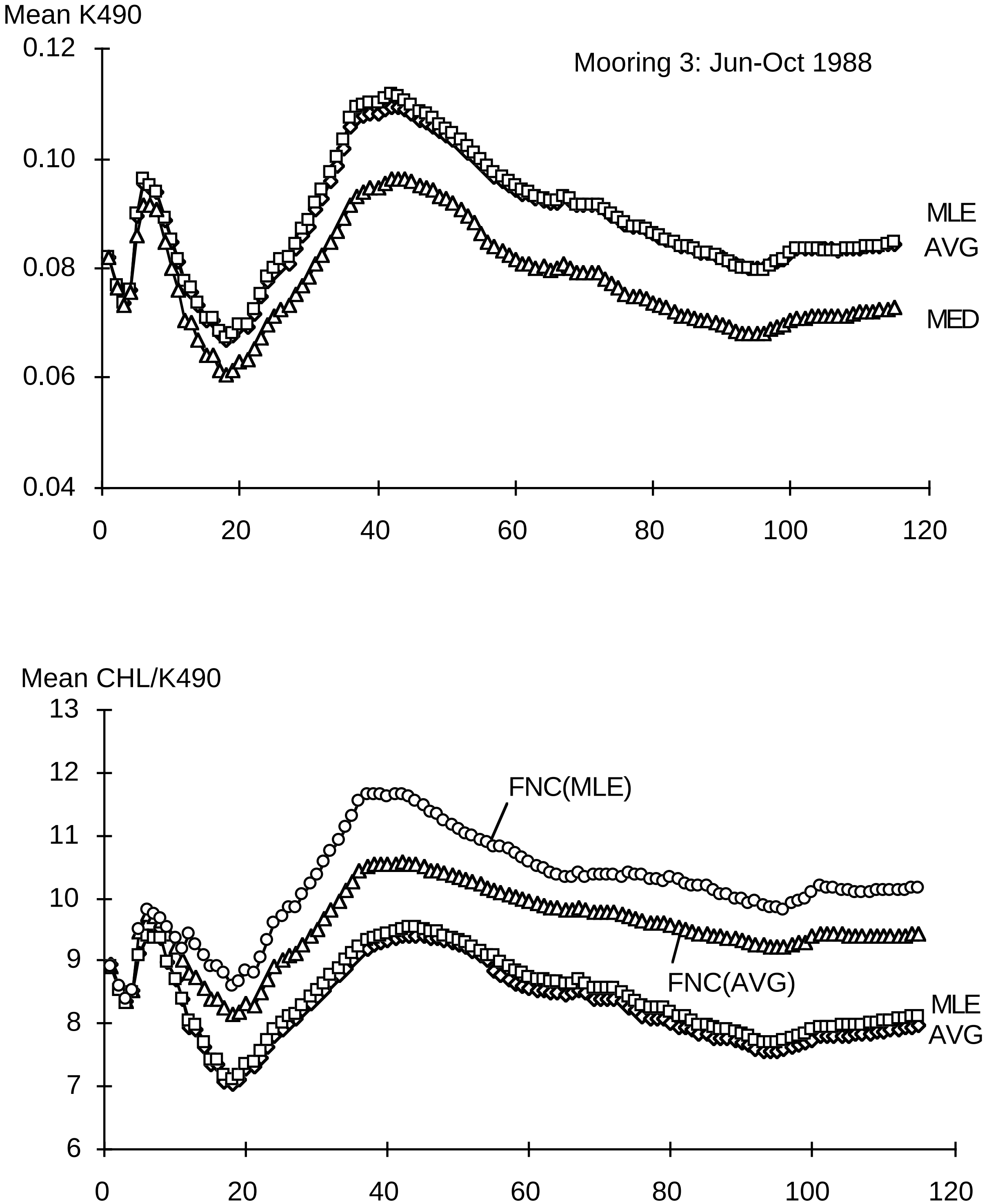
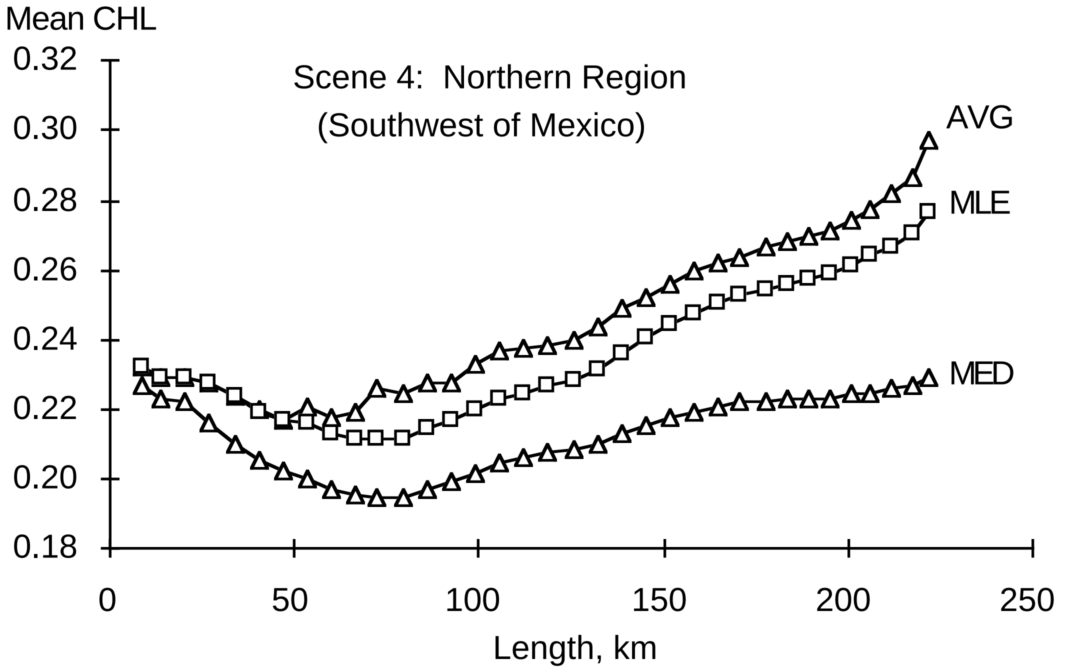


Fig. 29. Cumulative means of K_{490} (upper panel) and CHL/K_{490} (lower panel) for summer deployment of mooring 3 (Jun.–Oct. 1988).



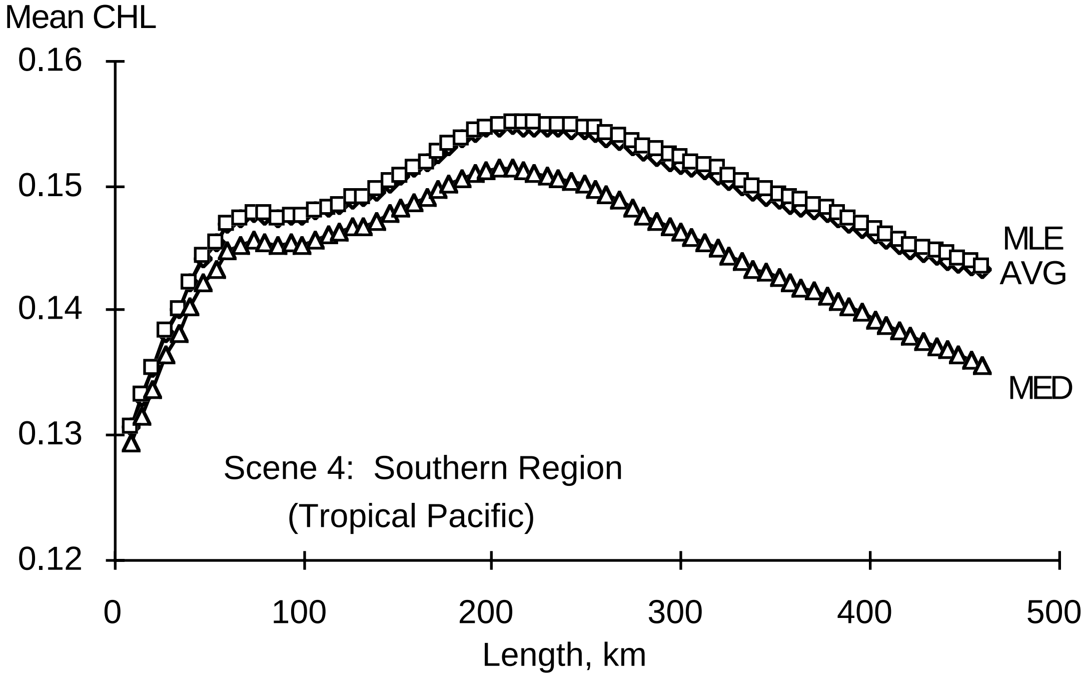


Fig. 30. Cumulative mean CHL in CZCS scene 4 based on LAC data within boxes of increasing area ($L \times L$) plotted against length, L. Results for boxes in the northern nearshore region (upper panel) and for the southern offshore region (lower panel).

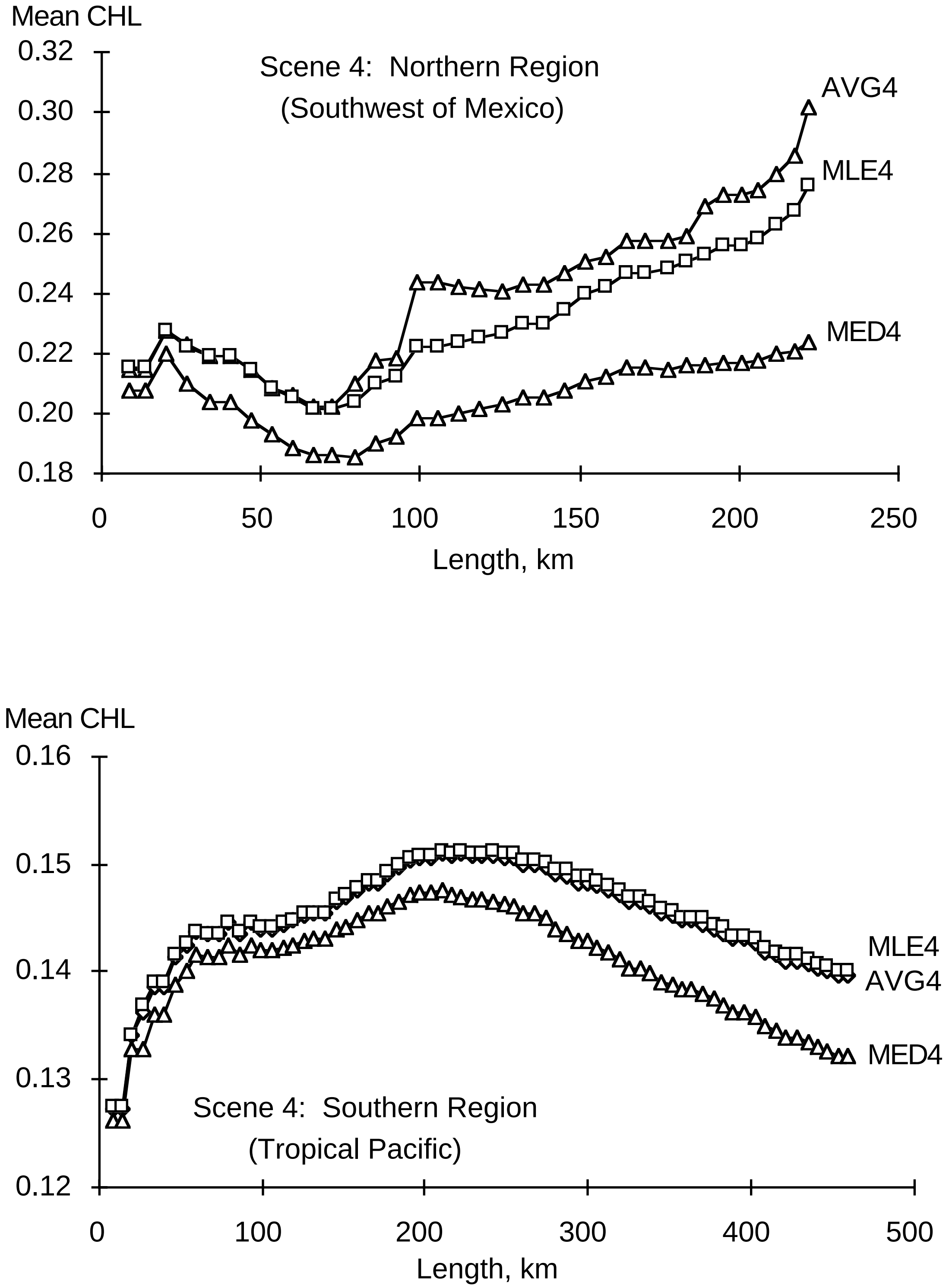


Fig. 31. Cumulative mean CHL in CZCS scene 4, in this case based on GAC data, within boxes of increasing area ($L \times L$) plotted against length, L. Results for boxes in the northern nearshore region (upper panel) and for the southern offshore region (lower panel).

Table 6. Comparison of cumulative means derived from CZCS data for largest areas ($L \times L$). The length (L) is shown in column 2.

CZCS	*Length*	*Number*	*Estimator*		*% Error*	
Scene	[km]	*of Pixels*	AVG	AVG4	AVG	AVG4
1	479	336,494	0.059	0.059	0.0	1.0
2	463	314,643	0.061	0.061	0.0	0.7
3	441	286,176	0.166	0.163	0.0	−1.5
4-N	222	72,381	0.297	0.302	0.0	1.4
4-S	460	310,518	0.143	0.140	0.0	−2.4
5-N	480	339,127	0.298	0.297	0.0	−0.4
5-S	291	124,475	0.384	0.382	0.0	−0.4
6	224	73,390	0.629	0.634	0.0	0.1
7-N	361	191,092	0.919	0.933	0.0	1.5
7-S	238	82,714	1.052	1.072	0.0	1.9
CZCS	*Length*	*Number*	*Estimator*		*% Error*	
Scene	[km]	*of Pixels*	MLE	MLE4	MLE	MLE4
1	479	336,494	0.058	0.059	−0.2	1.0
2	463	314,643	0.061	0.061	−0.2	0.5
3	441	286,176	0.166	0.164	0.4	−1.0
4-N	222	72,381	0.276	0.276	−7.1	−7.3
4-S	460	310,518	0.144	0.140	0.3	−2.2
5-N	480	339,127	0.297	0.297	−0.5	−0.4
5-S	291	124,475	0.386	0.385	0.5	0.2
6	224	73,390	0.531	0.537	−15.5	−14.7
7-N	361	191,092	0.883	0.908	−3.9	−1.2
7-S	238	82,714	0.973	0.992	−7.4	−5.7
CZCS	*Length*	*Number*	*Estimator*		*% Error*	
Scene	[km]	*of Pixels*	MED	MED4	MED	MED4
1	479	336,494	0.057	0.058	−2.05	−0.85
2	463	314,643	0.059	0.059	−3.93	−3.28
3	441	286,176	0.157	0.155	−5.01	−6.46
4-N	222	72,381	0.229	0.224	−22.97	−24.71
4-S	460	310,518	0.136	0.132	−5.31	−7.75
5-N	480	339,127	0.287	0.287	−3.99	−3.79
5-S	291	124,475	0.335	0.335	−12.76	−12.87
6	224	73,390	0.192	0.194	−69.47	−69.17
7-N	361	191,092	0.488	0.490	−46.86	−46.72
7-S	238	82,714	0.582	0.583	−44.70	−44.56

AVG4 estimators which were based on much smaller samples ($n \leq 9$). In both cases, differences were less than ±2% (Fig. 8 and Fig. 10). These results differ somewhat from those of Baker and Gibson (1987) who found that the arithmetic average underestimated the true mean of a lognormal variate, and that the maximum likelihood estimator was a better estimator of the mean when sample sizes were small. In the small samples that resulted from using GAC data, both the MLE4 and AVG4 estimators had a slight tendency to underestimate the *true* mean (AVG), as indicated by their small negative biases (Fig. 9, and Figs. 11–14), but no significant difference was found between the two estimators.

In the case of weekly and monthly means derived from the SEEP II data, the MLE and AVG estimators again proved to be nearly identical. The AVG estimator was nominally the *true* mean, but since it was based on small samples (7 days for weekly means and 31 or fewer days for monthly means), it is not necessarily better than other estimators of the mean.

Although the MLE and AVG estimators are equivalent with respect to accuracy, it was recommended that the MLE estimator be used because of its flexibility in allowing the estimation of level-4 variables from saved statistics of level-3 variables. In the remainder of this discussion, two questions are raised regarding the equivalence of the MLE and AVG estimators, and the answers discussed.

The first question is: *How important is the assumption* *that the variable is lognormally distributed?* If the variable being sampled is lognormally distributed, then the MLE estimator (5) is the maximum likelihood estimator of the mean, and the estimators for the standard deviation (23), median (24), and mode (25) are also maximum likelihood estimators of these parameters. But what if the underlying distribution is not lognormally distributed? How robust will the estimator be if the lognormal assumption is not valid?

The empirical evidence based on CZCS data and the SEEP II time series data supports the use of the MLE estimator. These data sets taken as a whole, i.e., a whole CZCS scene or a 16-month record from a single moored fluorometer, were approximately lognormal or mixtures of lognormal distributions. This was demonstrated for both satellite and *in situ* CHL distributions (Figs. 4 and 17), and observed for the other variables, but not shown. It is not surprising, therefore, that small subsets drawn from the whole data set behave as random samples drawn from a lognormal distribution.

However, the binned data were not random samples. Instead, they consisted of measurements made close together in space or time, and thus, they were correlated. To the extent that the binned data are positively correlated, the intrabin variance will be less than the variance of a random sample of the same size drawn from the whole data set.

It is possible to show that the MLE estimator will be a good approximation to the mean of any distribution with $s^2 \leq 0.5$, where s^2 is the variance of the natural logarithm of the variable. This result is derived from the series expansion for the exponential function

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots \tag{46}$$

Let X be any random variable whose distribution is unknown. Define $\mathrm{x} = \ln(\mathrm{X})$, and let m and s^2 be the mean and variance of x. Then

$$\mathrm{X} = e^{\mathrm{x}} = e^m e^{(\mathrm{x}-m)}, \tag{47}$$

and the expected value of X is

$$\begin{aligned} \mathrm{E[X]} &= e^m \,\mathrm{E}\Big[1 + (\mathrm{x}-m) + \frac{(\mathrm{x}-m)^2}{2!} \\ &\qquad + \frac{(\mathrm{x}-m)^3}{3!} + \ldots\Big] \\ &= e^m \,\mathrm{E}\Big[1 + \frac{m_2}{2!} + \frac{m_3}{3!} + \ldots\Big] \end{aligned} \tag{48}$$

where m_i denotes the ith central moment of x, defined by

$$m_i \equiv \mathrm{E}\big[(\mathrm{x}-m)^i\big]. \tag{49}$$

It is also noted that $m_1 = 0$, and $m_2 = s^2$.

If the variance is less than or equal to 0.5, then the terms involving higher order central moments will be a rapidly decreasing series. In fact, the series in brackets in (48) can be approximated by its first two terms

$$\begin{aligned} \mathrm{E[X]} &\approx e^m \left[1 + \frac{s^2}{2}\right]. \\ &\approx e^m e^{\frac{1}{2}s^2} \end{aligned} \tag{50}$$

The term on the right is the MLE estimator of the mean. Thus, there are two situations when the MLE estimator is valid: 1) when the underlying distribution is lognormal, or 2) when the variance of the natural logarithm of the variable is less than or equal to 0.5 (or the standard deviation of the base-10 logarithm is less than or equal to 0.3).

Figure 32 is a plot of the average variance of the logarithm of CHL within bins of size $L \times L$ plotted as a function of L for the CZCS scenes 1–5. It is noted that the variance within $9 \times 9\,\mathrm{km}^2$ bins was less than 0.5 for all five scenes, and the variance remained less than 0.5 as L increased up to the maximum length of 480 km. In scene 4, variances exceeded 0.5 at L greater than about 100 km.

The second question is: *Under what circumstances do the MLE and AVG estimators disagree, and is it possible to predict the nature and magnitude of their differences?*

In the study of cumulative means (Figs. 26–31), there were examples shown where the MLE and AVG estimators began to diverge as the size of the sampling domain increased. In one example (Figs. 26–27), the divergence could be associated with bad data, and the conclusion was that the MLE estimator was sensitive to anomalously low values. The possibility that similar errors might affect level-3 SeaWiFS data should be considered.

The discussion related to the first question suggests another circumstance in which the MLE and AVG estimators might disagree: the situation where the variance of the logarithm is large and the variable is not lognormally distributed. A situation such as this would occur when the sampling domain contains a mixture of lognormal distributions. In the case of spatial statistics, this would occur in frontal areas between sharply contrasting water types, e.g., high-chlorophyll waters mixing with low-chlorophyll waters. It is likely to be more common in sampling domains covering longer time periods.

Most of the CZCS scenes can be modeled as mixtures of lognormal distributions. Table 7 lists the means and variances of lognormal distributions that were fit to modes of the histograms shown in Fig. 4. Values of CHL derived according to the CHL_{23} formula were excluded from the fits. Note that within all modes, the variance was less than 0.5. However, when two or more modes are mixed, the variance of the mixture distribution will be increased due to differences between modes.

It is possible to quantify errors associated with the MLE estimator in the case of mixture distributions. An

Table 7. Results of fitting normal distributions to the modes of the histograms of log(CHL) in the CZCS scenes analyzed.

Scene	*Mode 1*		*Mode 2*		*Mode 3*	
Number	m	s^2	m	s^2	m	s^2
1	−2.82	0.04				
2	−2.98	0.02	−2.68	0.06		
3	−2.41	0.04	−1.70	0.04		
4	−2.35	0.06	−1.81	0.06	−1.06	0.16
5	−2.38	0.07	−1.25	0.08	−0.63	0.04
6	−2.65	0.09	−1.58	0.45	0.22	0.09
7	−1.78	0.15	−1.00	0.18	−0.07	0.15

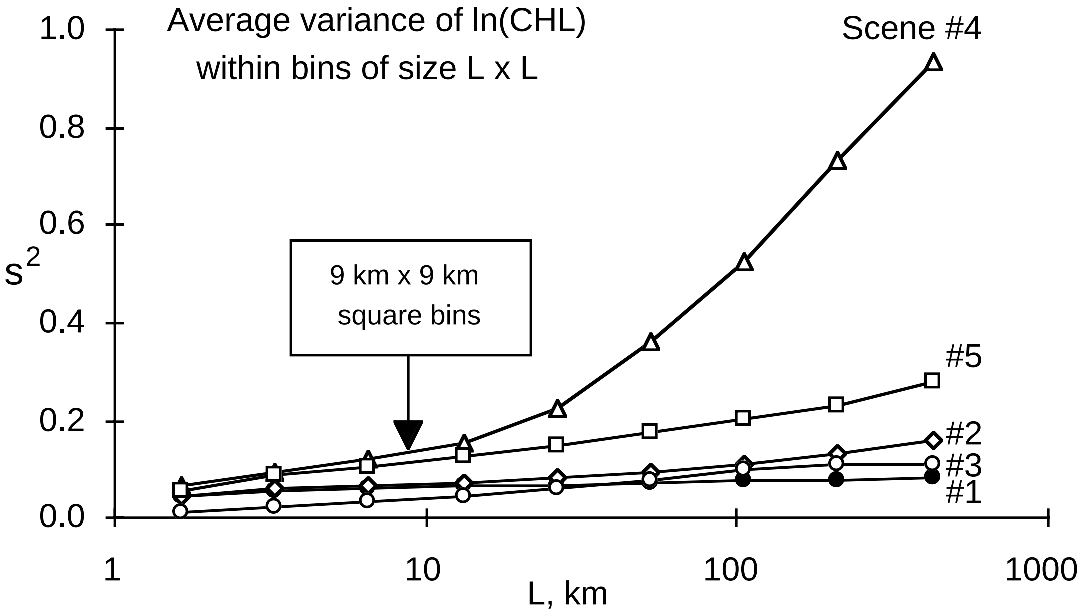


Fig. 32. Average variance of ln(CHL) within areas of size $L \times L$, as a function of L for CZCS scenes 1–5. Results for scenes 6 and 7 were not obtained because of the discontinuity in the CHL distributions in these scenes, which is an artifact of the bifurcated CZCS algorithm.

example is the case where there are two modes mixing in a sampling domain. Let each mode be a lognormal distribution with parameters m_i and s_i^2 where $i = 1$ or 2. If P is the proportion of the distribution that is mode 1, then the mean of the distribution is

$$\overline{X}_{\mathrm{avg}} = \mathrm{P}e^{\left(m_1+\frac{1}{2}s_1^2\right)} + (1-\mathrm{P})\mathrm{e}^{\left(\mathrm{m}_2+\frac{1}{2}\mathrm{s}_2^2\right)}, \qquad (51)$$

and the MLE estimator is

$$\overline{X}_{\mathrm{mle}} = \exp\left[\mathrm{P}\left(\mathrm{m}_1 + \frac{\mathrm{s}_1^2}{2}\right) + (1-\mathrm{P})\left(\mathrm{m}_2 + \frac{\mathrm{s}_2^2}{2}\right) + \mathrm{P}(1-\mathrm{P})\frac{(\mathrm{m}_1-\mathrm{m}_2)^2}{2}\right]. \qquad (52)$$

Relative errors for pair-wise mixtures of the modes listed in Table 7 are plotted against P in Fig. 33, where $m_1 < m_2$, and P is the proportion of the lower-chlorophyll mode. There are 14 curves shown in this figure, but only 5 have errors that are significantly different from zero. The largest errors (differences between MLE and AVG) occurred when modes from scene 6 were mixed, and especially when mode 1 (mean CHL $= 0.07\,\mathrm{mg\,m^{-3}}$) was mixed with mode 3 (mean CHL $= 1.3\,\mathrm{mg\,m^{-3}}$). Of all the cases considered here, the highest positive error (40%) occurred when 30% of mode 1 was mixed with 70% of mode 3 in scene 6, and the highest negative error (−30%) occurred when 90% of mode 1 was mixed with 10% of mode 3.

The patterns shown here indicate that the MLE can either under or over estimate the true mean when there are mixtures of lognormal distributions within the sampling domain. The MLE estimator tended to exceed AVG for low values of P, whereas AVG exceeded MLE for high val-

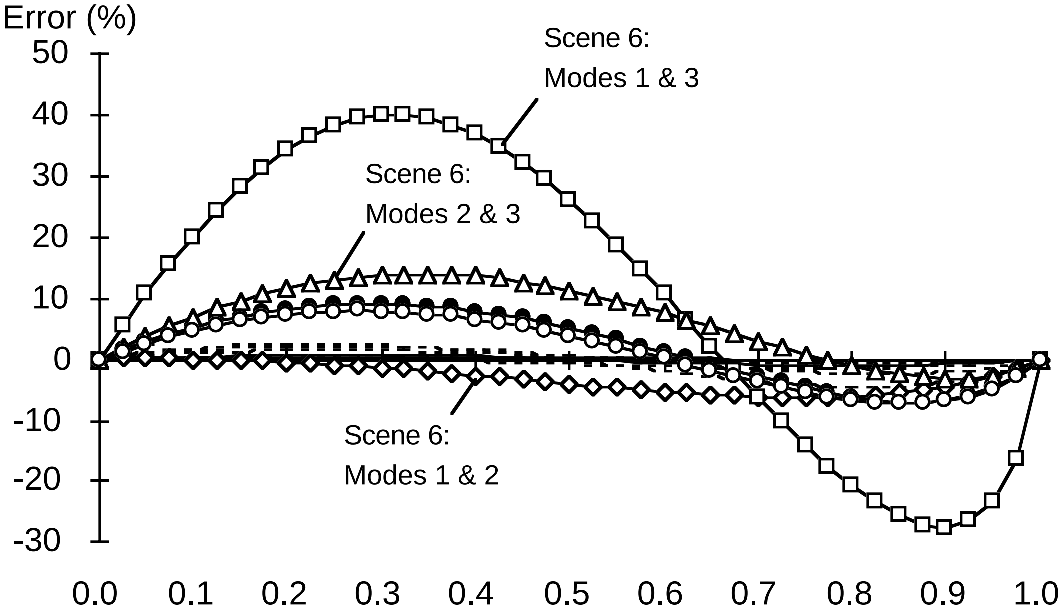


Fig. 33. Relative errors (%) in the MLE estimator resulting from mixtures of two lognormal distributions, plotted against P, the proportion of the mixture derived from the lower-chlorophyll mode. The 14 cases depicted in this figure represent all pair-wise combinations of modes within the seven CZCS scenes (see Table 7).

ues of P. The former was the situation with the SEEP data where there were a few low values in the data record that caused MLE(n) > AVG(n) to diverge. Likewise, the opposite seemed to be the case with CZCS data, where there were relatively few high values, e.g., values derived using the CHL_{23} algorithm, that brought about divergences between $\mathrm{MLE}(L) < \mathrm{AVG}(L)$.

The situations depicted in Fig. 33 may not be inclusive of all possible mixtures that would occur in nature, but they do span the range in the seven scenes analyzed. It is clear that patterns are complex, and yet, reassuring that with very few exceptions, errors were within ±10%.

3.4 Conclusions

The MLE estimator is a reasonably accurate estimator for the mean of CHL and other satellite-derived variables within sampling domains. It behaves as well as the arithmetic average, and yet it has an advantage over the AVG estimator in that it can be used to estimate means of a large class of *level-4* variables derived from the level-3 data. There were two situations that assure agreement between the AVG and MLE estimators: 1) if the variable is lognormally distributed within the sampling domain, or 2) if its variance is low. If the variance of its natural logarithm is less than 0.5, then AVG and MLE should agree regardless of the underlying distribution.

Two circumstances were identified where the MLE and AVG estimators are expected to disagree. One is the case where there are anomalously low values in the data (presumably bad data), and the other is where the sampling domain contains a mixture of lognormal distributions. Based on mixtures found in seven CZCS scenes spanning a wide variety of ocean environments, relative errors would typically be within ±10%.

Acknowledgments

The editors wish to acknowledge the editorial assistance of Barbara A. Leroux, who worked on this document during her participation in the NASA/Goddard Prince George's County Teacher Intern Program.

Editorial Note

This document is presented *as submitted* with minor modifications to correct typographical or obvious clerical errors and to maintain the established style of the *SeaWiFS Technical Report Series.*

Appendices

Appendix A

Equal-area Gridding Scheme for SeaWiFS Binned Data

Introduction: This appendix† describes the equal-area gridding scheme developed by the RSMAS Remote Sensing Group for binned ocean fields. The same approach has been adopted for AVHRR Ocean Pathfinder SST products and is proposed for MODIS. The gridding scheme is based on that adopted by the International Satellite Cloud Climatology Project (ISSCP).

This document does not motivate the need for an equal area grid for SeaWiFS or other oceanographic products. Such motivation can be found in a paper by W. Rossow and L. Gardner (1984). Furthermore, this document describes only the design of the proposed equal-area grid, and does not discuss other related topics such as rules for spatially or temporally combining observations into the equal-area bins.

Overview: The gridding scheme proposed consists of rectangular bins or tiles, arranged in zonal rows. A compromise between data processing and storage capabilities, on one hand, and the potential geophysical applications of satellite data, on the other hand, suggest that a suitable minimum bin size would be approximately 8–10 km on a side.

In the scheme proposed here, the tiles are approximately 9.28 km on a side. This size (9.28 km) was chosen because (a) it has approximately the desired minimum resolution, and (b) it results in 2,160 zonal rows of tiles from pole to pole, i.e., 1,080 in each hemisphere. This particular number of rows (2,160) has some advantages which will be discussed in more detail below. Because the total number of rows is even, the bins will never straddle the equator, i.e., there will be an equal number of rows above and below the equator. This avoids possible situations where the Coriolis factor is zero, a characteristic that numerical modellers expect from any gridding scheme adopted.

The total number of approximately 9 km bins is 5,940,422. The bins or tiles are arranged in a series of zonal rows; the number of tiles per row varies. The rows immediately above and below the equator have 4,320 tiles. This number is derived by dividing the perimeter of the Earth at the equator by the standard tile size, i.e., $2\pi R_e/9.28$, where R_e is the equatorial radius of the Earth ($R_e = 6378.145$ km). The number of tiles per row decreases approximately as a cosine function as the rows get closer to each pole (rigorously, there should be an adjustment for ellipticity of the Earth, as the equatorial radius decreases progressively to the smaller polar radius; this adjustment is not applied in the current implementation). At the poles, the number of tiles is always three. This special situation will be discussed in detail below. The number of tiles per row as a function of latitude is shown on Fig. A-1.

The number of bins in each zonal row is always an integer. To ensure an integer number of bins, the width of each bin (the size of a bin along a parallel, or x-length) must vary slightly from row to row. The bins, however, are always 9.28 km long along the meridians. That is, only one of the bin dimensions changes. The size of the bins at each zonal row is established in the following manner. First, a preliminary value for the number of tiles (N_p) at a given latitude (L) is computed as

$$N_p = 2\pi r/X,$$

where X is the x-size of a bin at the equator (9.28 km) and r is the radius of the circle produced by slicing the Earth with a plane parallel to the equator at latitude L. The radius r can be calculated as

$$r = R_e \cos(L),$$

where R_e is the equatorial radius of the Earth. If the fractional part of N_p is greater than or equal to 0.5, then N_p is rounded up to the nearest integer, i.e., the final number of tiles will be the integer portion of N_p plus one; otherwise, N_p is rounded down. The final number of tiles is the integer portion of N_p. Once the final integer number of tiles along a row is calculated, the x-size of the tiles must be adjusted. This is done by dividing the perimeter of the row ($2\pi r$) by the integer number of tiles. The result is the x-length (width) of a tile for a given row.

Because the x-length of the tiles is adjusted to ensure an integer number at each row, the *equal area* characteristics of this binning scheme are not rigorously preserved. However, variations in tile size are negligible throughout most of the globe and only become relevant at very high latitudes, where there are fewer tiles per row, and any adjustments are more noticeable. As the number of tiles increases with distance from the poles, the difference between tile sizes rapidly becomes practically unnoticeable. To provide an idea of the magnitude of the fluctuations in tile size, the worst possible case occurs when half a tile remains *uncovered* after filling a zonal row with an integer number of tiles. Once a row has 100 bins (approximately 16 rows, or 148 km from the poles), the worst possible difference between the actual tile x-length and the standard x-length is of the order of 0.5%, i.e., half a tile's length redistributed among about 100 tiles. For a tile of about 9 km a side, this represents a difference in the x-length of about 45 m. Through a similar calculation, a row with 50 bins (about 80 km away from the poles) has a 1% variation with respect to the standard bin size.

The gridding scheme described here has an extremely useful feature. The number of 9.28 km tiles in each hemisphere (1,080) is divisible by many numbers (e.g., 2,3,4,5,6); and therefore, it is extremely easy to generate an integer number of rows at many useful spatial resolutions. For instance, 12 rows of approximately 9.28 km tiles can be combined to generate zonal bands of 1° (1° of latitude is equal to 111.12 km; 12 bins would form a band 111.20 km wide). Another example is the use of 30 rows to generate zonal bands of 2.5°, a typical output resolution of atmospheric circulation models.

The poles: Both the North and South Poles are special cases in the gridding scheme presented here. The pole areas are always covered by three tiles shaped like pie sectors. While the meridional size of the polar bins (the y-length) will be the usual 9.28 km, the length of the bins along the arc of the sectors will be slightly larger. Neglecting sphericity, the area encompassed by the last row of tiles is πX^2, where $X = 9.28$ km. If the area of the circle is expressed as a rectangle of height X, the remaining dimension is πX. If the perimeter is divided by three (to yield three tiles), each tile will have dimensions X by $\pi X/3$ (approximately $1.05X$). Thus, the bases of the triangular polar

† This text is courtesy of the Remote Sensing Group, Rosenstiel School of Marine and Atmospheric Science, University of Miami.

tiles are about 5% larger than the x-length of the equatorial tiles.

Binning software: Several routines have been developed to perform the principal transformations required for binning and mapping data, such as converting latitudes and longitudes into bin numbers. Other routines perform the inverse transformation, i.e., given a bin number they return a latitude and longitude corresponding to the centroid of that bin. These routines use a common initialization routine that must be executed prior to calling the conversion routines.

Two numbering schemes are used internally, corresponding to one- and two-dimensional (1-D and 2-D, respectively) accessing schemes. The 1-D scheme numbers all bins consecutively, beginning with 1 at the southernmost row and working eastward from −180° around each circle of latitude. The 2-D scheme uses a row number, from 1 to 2,160, and a number to indicate its location within the row, beginning at 1 for each row.

Variable Dictionary: The variables and their definitions for the pseudocode are presented below.

`NUMROWS` The (integer) number of rows in the grid (equal to 2,160 for SeaWiFS).

`BASEBIN` An `integer*4` array of size `NUMROWS` that contains the index number of the first bin in each row.

`NUMBIN` An integer array of size `NUMROWS` containing the total number of possible bins in each row.

`LATBIN` A `real*4` array of size `NUMROWS` that contains the center latitudes (decimal degrees) of the corresponding `BASEBIN`s.

`TOTBINS` The (integer*4) number of possible bins in the grid (equal to 5,940,422 for `NUMROWS`=2,160).

`ROW` The row number (integer); range is 1 to `NUMROWS`.

`COL` The bin number (integer); the range is from 1 to `NUMBIN(ROW)`.

`IDX` The bin index number (`integer*4`); range is 1 to `TOTBINS`.

`LAT` The input latitude (`real*4`) for obtaining the corresponding bin's `ROW` and `COL`, or `IDX`; or the output latitude for a bin specified by `ROW` and `COL`, or `IDX`. (The range for `LAT` is −90 to +90 decimal degrees.)

`LON` The input longitude (`real*4`) for obtaining the corresponding bin's `ROW` and `COL`, or `IDX`; or the output longitude for a bin specified by `ROW` and `COL`, or `IDX`. (The range for `LAT` is −180 to 180 decimal degrees.)

Pseudocode: The following pseudocode demonstrates the generation of the grid and the calculations for determining the center latitude and longitude for a given bin and for identifying a bin given a latitude and a longitude. The algorithms are illustrative in purpose and do not necessarily represent an optimal implementation. They are based on software developed by J. Brown, University of Miami.

```
#
# Set up NUMBIN and BASEBIN arrays
#
BASEBIN(1) = 1
do from ROW = 1 to NUMROWS
  LATBIN(ROW) = ((ROW-0.5)*180.0/NUMROWS) - 90.0
  NUMBIN(ROW) =
     int((2*NUMROWS*cos_dbl_deg(LATBIN(ROW)))+0.5)
  if ROW>1 then BASEBIN(ROW) = BASEBIN(ROW-1) + NUMBIN(ROW-1)
end do
TOTBINS = BASEBIN(NUMROWS) + NUMBIN(NUMROWS) - 1
#
# Identify bin from lat (-90 to +90) and lon (-180 to 180)
#
ROW = integer((90.0+LAT)*(NUMROWS/180.0)) + 1
ROW = minimum(ROW,NUMROWS)
LON = LON + 180.0
COL = integer(LON*NUMBIN(ROW)/360.0) + 1
COL = minimum(COL,NUMBIN(ROW))
IDX = BASEBIN(ROW) + COL - 1
#
# Get bin center lat/lon for given bin index
#
ROW = NUMROWS
IDX = maximum(IDX,1)
do while IDX<BASEBIN(ROW)
   ROW = ROW - 1
end do
LAT = LATBIN(ROW)
LON = 360.0*(IDX-BASEBIN(ROW)+0.5)/NUMBIN(ROW)
LON = LON - 180.0
#
# Get bin center lat/lon for given bin row/column
#
LAT = LATBIN(ROW)
LON = 360.0*(COL-0.5)/NUMBIN(ROW)
LON = LON - 180.0
```

Appendix B

Scheme for Weighting Data

This appendix describes the scheme used to weight data from different times (orbits) in producing temporal means and variances. The level-2 SeaWiFS data will be log-transformed before the following schemes are applied. Note that the lower case letter x is used to denote the natural logarithm of the variable X, that is, x = ln(X). The MLE estimator for the mean of a lognormal variable X requires that the maximum likelihood estimators of the mean and variance of x be obtained first.

The Textbook Case for Unweighted Data: If the data within a sampling domain, $\mathrm{x_i}, \mathrm{i} = 1, \ldots, \mathrm{n}$, are independent and identically distributed normal random variables with a true mean μ and variance σ^2, then the sample mean

$$\overline{\mathrm{x}} = \frac{1}{\mathrm{n}} \sum_{i=1}^{\mathrm{n}} \mathrm{x}_i \tag{B1}$$

is the maximum likelihood estimator of μ. The sample variance is defined as

$$\mathrm{s}^2 = \frac{1}{\mathrm{n}} \sum_{i=1}^{\mathrm{n}} \left(\mathrm{x}_i - \overline{\mathrm{x}}\right)^2 \tag{B2}$$

and computed as

$$\mathrm{s}^2 = \frac{1}{\mathrm{n}} \sum_{i=1}^{\mathrm{n}} \mathrm{x}_i{}^2 - \overline{\mathrm{x}}^2 \tag{B3}$$

is the maximum likelihood estimator of the variance, σ^2. Note that s^2 is not the more common unbiased estimator of the variance which is obtained by multiplying (B3) by n/(n − 1). For the specific case of SeaWiFS spatial statistics, i.e., for data within a bin obtained during the same orbital pass, equations (B1) and (B3) will be used to compute the mean and variance of $\mathrm{x} = \ln(\mathrm{X})$, for each variable X.

The General Case for Weighted Data: Let w_i be the weight given to the ith observation (data point). The weighted mean and variance analogous to (B1) and (B3) are

$$\overline{\mathrm{x}} = \frac{1}{\mathrm{W}} \sum_{i=1}^{\mathrm{n}} \mathrm{w}_i \mathrm{x}_i \qquad (B4)$$

$$\mathrm{s}^2 = \frac{1}{\mathrm{W}} \sum_{i=1}^{\mathrm{n}} \mathrm{w}_i \mathrm{x}_i^2 - \overline{\mathrm{x}}^2 \qquad (B5)$$

where W is the sum of the weights

$$\mathrm{W} = \sum_{i=1}^{\mathrm{n}} \mathrm{w}_i. \qquad (B6)$$

The Specific Case for Weighted Data: How this applies to the weighting of spatial statistics as they are binned over time are considered here. In general, there will be N sets of spatial statistics, each corresponding to a time $\mathrm{t_i}, \mathrm{i} = 1, \ldots, \mathrm{N}$, and each set of spatial statistics will be based on $\mathrm{n_i}$ observations from the same orbital pass. To be obtained is a weighted mean and variance of the data over observation $\mathrm{x_{ij}}$ where j refers to the jth observation at time $\mathrm{t_i}$ and $i = 1, \ldots, \mathrm{N}$, and $j = 1, \ldots, \mathrm{n_i}$. (Recall that $\mathrm{x_{ij}} = \ln(\mathrm{X_{ij}})$.

One approach would be to compute a mean, $\overline{\mathrm{x}}_i$, and variance, s_i^2, for each set of spatial statistics, and then simply average the means and variances over all times, $\mathrm{t_i} = 1, \ldots, \mathrm{N}$. If this approach is used, the weights applied to each observation would be

$$\mathrm{w}_{ij} = \frac{1}{\mathrm{n}_i}. \qquad (B7)$$

It was decided that this gave too much weight to data sets having few observations. The alternative is to weight all data equally, but this gives too much weight to the data sets with numerous observations. The compromise was to use

$$\mathrm{w}_{ij} = \frac{1}{\sqrt{\mathrm{n_i}}}. \qquad (B8)$$

This is equivalent to weighting $\overline{\mathrm{x}}_i$ and $\mathrm{s_i^2}$ by $\sqrt{\mathrm{n_i}}$.

The weighted mean and variance are

$$\overline{\mathrm{x}} = \frac{1}{\mathrm{W}} \sum_{i=1}^{\mathrm{N}} \frac{1}{\sqrt{\mathrm{n_i}}} \sum_{j=1}^{\mathrm{n_i}} \mathrm{x}_{ij} \qquad (B9)$$

$$\mathrm{s}^2 = \frac{1}{\mathrm{W}} \sum_{i=1}^{\mathrm{N}} \frac{1}{\sqrt{\mathrm{n_i}}} \sum_{j=1}^{\mathrm{n_i}} \mathrm{x}_{ij}^2 - \overline{\mathrm{x}}^2 \qquad (B10)$$

where the sum of the weights is

$$\mathrm{W} = \sum_{i=1}^{\mathrm{N}} \sum_{j=1}^{\mathrm{n_i}} \frac{1}{\sqrt{\mathrm{n_i}}} = \sum_{i=1}^{\mathrm{N}} \sqrt{\mathrm{n}_i}. \qquad (B11)$$

Equations (B9) and (B10) will be used to obtain the temporal statistics of ln(X) in each sampling domain.

Appendix C

Algorithms for Binning and Interpreting SeaWiFS Binned Data

Three algorithms are described and their pseudocodes presented in this appendix: the Space Binner algorithm bins data from a single scene; the Time Binner algorithm bins output from the Space Binner (or from the Time Binner) to accumulate sums over binning periods; and the Bin Data Interpreter is used to interpret binned data products to derive the mean, standard deviation, median and mode of level-3 data. Only GAC data will be binned operationally to generate archived level-3 products.

Spatial Binning Algorithm: The spatial binning algorithm is applied to the level-2 GAC scenes. In general, there will be one set of spatial statistics created for each scene. The only exception will be when an orbit crosses 180° longitude, in which case there will be two sets of spatial statistics corresponding to different days.

Let $\mathrm{X_{ji}}$ be an acceptable observation of the variable $\mathrm{X_j}$ in pixel i, and let `LON`(i) and `LAT`(i) be the longitude and latitude at the center of pixel i. (A pixel will be considered to have acceptable level-2 data if it passes screens for sun glint, clouds and other masks, in which case all of the variables will be considered acceptable.) From these coordinates, the bin index number b will be determined according to known relationships (see Appendix A).

Then for each variable j, the natural logarithm $\mathrm{LOGX} = \ln(\mathrm{X_{ji}})$ is obtained, and the following sums incremented

$$\mathrm{SUMX}(b,j) = \mathrm{SUMX}(b,j) + \mathrm{LOGX} \qquad (C1)$$

and

$$\mathrm{SUMXX}(b,j) = \mathrm{SUMXX}(b,j) + \mathrm{LOGX} \times \mathrm{LOGX}\,. \qquad (C2)$$

In addition, the number of pixels contributing to the sums in bin b is incremented

$$\mathrm{N}(b) = \mathrm{N}(b) + 1, \qquad (C3)$$

and a binary-valued variable is set to 1 to indicate bin b contains data

$$\mathrm{NSEG}(b) = 1. \qquad (C4)$$

After processing all valid data from this scene, the total weight for each bin is computed

$$\mathrm{W(b)} = \sqrt{\mathrm{N(b)}}, \qquad (C5)$$

and the variable sums are weighted as per (B9) and (B10)

$$\mathrm{SUMX}(b,j) = \frac{\mathrm{SUMX}(b,j)}{\mathrm{W(b)}} \qquad (C6)$$

and

$$\mathrm{SUMXX}(b,j) = \frac{\mathrm{SUMXX}(b,j)}{\mathrm{W(b)}}. \qquad (C7)$$

Finally, a 16-bit number $\mathrm{TT}(b)$ is defined for each bin. This number will be used in subsequent stages of temporal binning to indicate the temporal distribution of the data. In the spatial binning algorithm, all bits in TT(b) will be 0 except the lowest bit which will be set to 1 if there are data in bin b.

The output from the spatial binning algorithm consists of the spatial statistics for each bin: b, N(b), NSEG(b), W(b), TT(b), and a pair of weighted sums, SUMX(b,j) and SUMXX(b,j), for each variable j.

Space Binner Code: This program takes one level-2 scene as input and bins it into one (or two, if the level-2 scene crosses 180° longitude) level-3 binned data product as output. This is called *spatial* binning since the bins are of lower resolution and the level-2 product is considered to represent a snapshot, i.e., no time averaging occurs, of the Earth's surface. Products generated by this program are not archived but are used as input to the time binner.

Variable and Constant Dictionary: The variables and their definitions for the pseudocode are presented below.

Constants

`MAXBINS` The maximum number of bins (5,940,422).

`NVARS` The number of derived level-3 geophysical variables whose observational values are stored in the associated `SUMX` and `SUMXX` pairs.

Level-2 Variables

`NPIXELS` The number of pixels in a scan line of the input level-2 product.

`NSCANS` The number of scan lines in the input level-2 product.

`PXLAT` A `real*4` 1-D array of size `NPIXELS`; represents the latitude for a given pixel `I` of a given scan line `L`.

`PXLON` A `real*4` 1-D array of size `NPIXELS`; represents the longitude for a given pixel `I` of a given scan line `L`.

`OBS` A `real*4` 2-D array of size `NPIXELS`×`NVARS`; represents the derived level-2 values that are to be binned into the level-3 product for a given pixel `I` of a given scan line `L`.

Output Variables

`SUMX` A `real*4` 2-D array of size `MAXBINS`×`NVARS`; represents the sum of the natural logarithm of the level-3 geophysical variable's values divided by the square root of the number of those values for a given bin `IDX`; saved in the output product if, and only if, `N(IDX)` is greater than zero.

`SUMXX` A `real*4` 2-D array of size `MAXBINS`×`NVARS`; represents the sum of squares of the natural logarithm of the level-3 geophysical variable's values divided by the square root of the number of those values for a given bin `IDX`; saved in the output product if, and only if, `N(IDX)` is greater than zero.

`N` An `integer*2` 1-D array of size `MAXBINS`; represents the number of values summed into `SUMX` and `SUMXX` for all variables (Js) and for a given bin `IDX`; saved in the output product if, and only if, `N(IDX)` is greater than zero.

`NSEG` An `integer*2` 1-D array of size `MAXBINS`; represents the number of level-2 scenes which contribu ted to `SUMX` and `SUMXX` for all `J`s for a given bin `IDX`; saved in the output product if, and only if, `N(IDX)` is greater than zero. For this program, since only one scene is input, all saved values of `NSEG` will be 1.

`W` A `real*4` 1-D array of size `MAXBINS`; represents the weight factor for all `J`s for a given bin `IDX`; calculated as the square root of `N(IDX)`; saved in the output product if, and only if, `N(IDX)` is greater than zero.

`TT` An `integer*2` 1-D array of size `MAXBINS`; the bit values of `TT` represent the time trend of the values summed into `SUMX` and `SUMXX` for all `J`s for a given bin `IDX`; saved in the output product if, and only if, `N(IDX)` is greater than zero. For this program, since only one scene is input, all saved values of `TT` will have the lowest bit only set to 1.

`IDX` An `integer*4` word representing the index number of each bin with a value ranging from 1 to `MAXBINS`; saved in the output product if, and only if, `N(IDX)` is greater than 0.

Note: For each `N(IDX)`> 0, 8×`NVARS`+14 bytes of information will be output.

Other Variables

`I` Counter index of pixels on a scan line. Range is from 1 to `NPIXELS`.

`J` Counter index of geophysical variables to be binned. Range is from 1 to `NVARS`.

`L` Counter index of scan lines. Range is from 1 to `NSCANS`.

`XLOG` Natural logarithm (`real*4`) of `OBS` for a given `I` and `J`.

```
#
# Initialize
#
do from IDX=1 to MAXBINS
   do from J=1 to NVARS
      SUMX(IDX,J) = 0.0
      SUMXX(IDX,J) = 0.0
   end do
   N(IDX) = 0
   NSEG(IDX) = 0
   TT(IDX) = 0
end do
read from level-2 scene: NPIXELS, NSCANS
#
# Input level-2 scene and accumulate stats for each bin
#
#
do from L=1 to NSCANS
   read arrays PXLAT, PXLON, OBS for scan line L
   do from I=1 to NPIXELS
      if sample I passes screen flags then
        IDX = get_bin_index(PXLAT(I),PXLON(I))
        do from J=1 to NVARS
           XLOG = natural_log(OBS(I,J))
           SUMX(IDX,J)  = SUMX(IDX,J)  + XLOG
           SUMXX(IDX,J) = SUMXX(IDX,J) + XLOG*XLOG
        end do
        N(IDX) = N(IDX) + 1
        NSEG(IDX) = 1
     end if
   end do
end do
#
```

```
# Divide sums by weight and output space binned product
#
do from IDX=1 to MAXBINS
   if N(IDX) > 0 then
      set lowest bit of TT(IDX)
      W(IDX) = square_root(N(IDX))
      do from J=1 to NVARS
         SUMX(IDX,J)  = SUMX(IDX,J)/W(IDX)
         SUMXX(IDX,J) = SUMXX(IDX,J)/W(IDX)
      end do
      write to space binned level-3 product:
         IDX, N(IDX), NSEG(IDX), W(IDX), TT(IDX)
         SUMX(IDX,J), SUMXX(IDX,J), for J=1 to NVARS
       end write
   end if
end do
```

Temporal Binning Algorithm: The temporal binning algorithm combines the appropriate spatial statistics within each sampling domain. The sampling domain for a particular bin will be either a day, *week*, month, or year.

For each set of spatial statistics there is an associated time t. The output from the spatial algorithm at time t will be the input for the temporal binning algorithm. Let this input be indexed by the time t: $\mathrm{N(b)_t}$, $\mathrm{NSEG}(b)_\mathrm{t}$, $\mathrm{W(b)_t}$, $\mathrm{TT}(b)_\mathrm{t}$, and the pairs of weighted sums, $\mathrm{SUMX}(b{,}j)_\mathrm{t}$ and $\mathrm{SUMXX}(b{,}j)_\mathrm{t}$, for each variable $\mathrm{X_j}$.

If $N(b)_\mathrm{t} > 0$, then the temporal sums

$$\mathrm{SUMX}(b{,}j) \;=\; \mathrm{SUMX}(b{,}j) \;+\; \mathrm{SUMX}(b{,}j)_\mathrm{t} \tag{C8}$$

and

$$\mathrm{SUMXX}(b{,}j) \;=\; \mathrm{SUMXX}(b{,}j) \;+\; \mathrm{SUMXX}(b{,}j)_\mathrm{t} \tag{C9}$$

are incremented for each variable j. In addition, the number of pixels contributing to the sums is counted

$$\mathrm{N(b)} \;=\; \mathrm{N(b)} \;+\; \mathrm{N(b)_t} \tag{C10}$$

and the number of spatial data sets (orbits) contributing to the sums

$$\mathrm{NSEG}(b) \;=\; \mathrm{NSEG}(b) \;+\; \mathrm{NSEG}(b)_\mathrm{t}. \tag{C11}$$

The sum of weights is computed

$$\mathrm{W(b)} \;=\; \mathrm{W(b)} \;+\; \mathrm{W(b)_t}. \tag{C12}$$

and the appropriate bit of the time distribution variable $\mathrm{TT}(b)$ is set to 1 to reflect that data were present at time t in bin b.

Output from the temporal binning algorithm consists of the level-3 data for each bin: b, N(b), NSEG(b), W(b), TT(b), and a pair of weighted sums, SUMX(b,j) and SUMXX(b,j), for each variable j. Note that the output from the temporal binning algorithm is in the same form as its input. In fact, daily binned products can serve as input to the temporal binning algorithm to produce weekly, monthly, or longer-term products.

Time Binner Code: This program takes as input level-3 binned segment products produced by the space binner and combines them into a binned product representing one day or takes binned products produced by the time binner (this program) and combines them into longer-term binned products. This process is called *temporal* binning since it combines data over a certain time period while not changing their spatial resolution.

Variable and Constant Dictionary: The variables and their definitions for the pseudocode are presented below.

Constants

MAXBINS The maximum number of bins (5,940,422).

NVARS The number of derived level-3 geophysical variables whose observational values are stored in the associated SUMX and SUMXX pairs.

Input Variables

NBINS The number of bins to read from an input level-3 product.

SUMX_INPUT A real*4 1-D array of size NVARS; represents SUMX as output by the space or time binner for all level-3 geophysical variables (Js) of a given bin IDX being read.

SUMXX_INPUT A real*4 1-D array of size NVARS; represents SUMXX as output by the space or time binner for all level-3 geophysical variables (Js) of a given bin IDX being read.

N_INPUT An integer*2 word; represents N as output by the space or time binner for a given bin IDX being read.

NSEG_INPUT An integer*2 word; represents NSEG as output by the space or time binner for a given bin IDX being read.

W_INPUT A real*4 word; represents W as output by the space or time binner for a given bin IDX being read.

TT_INPUT An integer*2 word; represents TT as output by the space or time binner for a given bin IDX being read.

IDX An integer*4 word representing the index number of the bin being read from an input level-3 product.

Output Variables

SUMX A real*4 2-D array of size MAXBINS×NVARS; represents the sum of the SUMX_INPUT for the level-3 geophysical variables (Js) from all input products for a given bin IDX; saved in the output product if, and only if, N(IDX) is greater than zero.

SUMXX A real*4 2-D array of size MAXBINS×NVARS; represents the sum of the SUMXX_INPUT for the level-3 geophysical variables (Js) from all input products for a given bin IDX; saved in the output product if, and only if, N(IDX) is greater than zero.

N An integer*2 1-D array of size MAXBINS; represents the sum of the N_INPUT from all input products for a given bin IDX; saved in the output product if, and only if, N(IDX) is greater than zero.

NSEG An integer*2 1-D array of size MAXBINS; represents the sum of the NSEG_INPUT from all input products for a given bin IDX; saved in the output product if, and only if, N(IDX) is greater than zero.

W A real*4 1-D array of size MAXBINS; represents the sum of the W_INPUT from all input products for a given bin IDX; saved in the output product if, and only if, N(IDX) is greater than zero.

TT An integer*2 1-D array of size MAXBINS; the bit sequence of TT represent the time trend of the values summed into SUMX and SUMXX for all Js for a given bin IDX; saved in the output product if, and only if, N(IDX) is greater than zero. The bits represent consecutive time in the binning period, the lowest bit being the earliest time. For daily binned products, the bits correspond to the relative sequence of orbits binned. For 8-day products, each bit represents one day; for monthly products, each bit represents two days; and for yearly products, each bit represents one month. A TT(IDX) bit will be set to 1 only if data, for the time corresponding to that bit, were binned in bin IDX.

IDX An integer*4 word representing the index number of each bin with a value ranging from 1 to MAXBINS; saved in the output product if, and only if, N(IDX) is greater than 0.

Note: For each N(IDX)>0, $8\times$NVARS$+14$ bytes of information will be output.

Other Variables

J Counter index of geophysical variable to be binned. Range is from 1 to NVARS.

B Counter index of bins read from input product. Range is from 1 to NBINS.

```
#
# Initialize
#
do from IDX=1 to MAXBINS
   do from J=1 to NVARS
      SUMX(IDX,J) = 0.0
      SUMXX(IDX,J) = 0.0
   end do
   N(IDX) = 0
   NSEG(IDX) = 0
   W(IDX) = 0.0
end do
#
# Input space or time binned products and accumulate
#    statistics for each bin
#
do for each binned input product
   read from metadata of binned input products: NBINS
   do from B=1 to NBINS
     read from bin B:
         IDX, N_INPUT, NSEG_INPUT, W_INPUT, TT_INPUT
         SUMX_INPUT(J), SUMXX_INPUT(J), for J=1 to NVARS
     end read
     do from J=1 to NVARS
        SUMX(IDX,J)  = SUMX(IDX,J)  + SUMX_INPUT(J)
        SUMXX(IDX,J) = SUMXX(IDX,J) + SUMXX_INPUT(J)
     end do
     N(IDX) = N(IDX) + N_INPUT
     NSEG(IDX) = NSEG(IDX) + NSEG_INPUT
     W(IDX) = W(IDX) + W_INPUT
     use TT_INPUT, date, or orbit of input to set TT(IDX)
   end do
end do
#
# Output time binned product
#
do from IDX=1 to MAXBINS
   if N(IDX) > 0 then
      write to time binned level-3 product
         IDX, N(IDX), NSEG(IDX), W(IDX), TT(IDX)
         SUMX(IDX,J), SUMXX(IDX,J), for J=1 to NVARS
      end write
   end if
end do
```

Algorithms for Calculating Statistics of Level-3 Variables: The means, variances, medians, and modes can be estimated using the level-3 data as described in Section 2.3. Here the same equations are described in terms of the pseudocode logic used in this Appendix. The level-3 data provided for each bin are: b, N(b), NSEG(b), W(b), TT(b), and a pair of weighted sums, SUMX(b,j) and SUMXX(b,j), for each level-3 variable j.

For each variable $\mathrm{X_j}$, the mean and variance of its natural logarithm are calculated

$$\mathrm{m_x} = \frac{\mathrm{SUMX}(b,j)}{\mathrm{W(b)}} \qquad (C13)$$

and

$$\mathrm{s_x^2} = \frac{\mathrm{SUMXX}(b,j)}{\mathrm{W(b)}} - \mathrm{m_x^2}. \qquad (C14)$$

The MLE estimator for the mean of $\mathrm{X_j}$ in bin b is

$$\overline{\mathrm{X}}(b,j) = e^{\left(\mathrm{m_x}+\frac{1}{2}\mathrm{s_x^2}\right)} \qquad (C15)$$

and the standard deviation of $\mathrm{X_j}$ is estimated by

$$\mathrm{SD}(b,j) = \overline{\mathrm{X}}(b,j)\left[e^{\mathrm{s_x^2}} - 1\right]^{\frac{1}{2}}, \qquad (C16)$$

and $\left[\mathrm{SD}(b,j)\right]^2$ is the estimated variance.

Assuming the distribution of $\mathrm{X_j}$ is approximately lognormal, then the median can be estimated by

$$\overline{\mathrm{X}}_{\mathrm{med}}(b,j) = e^{\mathrm{m_x}}, \qquad (C17)$$

and the mode (most common value) by

$$\overline{\mathrm{X}}_{\mathrm{mod}}(b,j) = e^{\left(\mathrm{m_x}-\mathrm{s_x^2}\right)}. \qquad (C18)$$

Bin Data Interpreter Code: This program interprets the geophysical data from binned products created by the space binner or the time binner. It will calculate the maximum likelihood estimate (MLE) of the mean, standard deviation, median, and mode for each level-3 binned geophysical variable.

Variable and Constant Dictionary: The variables and their definitions for the pseudocode presented are below.

Constants

NVARS The number of derived level-3 geophysical variables whose observational values are stored in the associated SUM_INPUT and SUMXX_INPUT pairs.

Level-3 Input Variables

NBINS The number of bins to read from an input level-3 product.

SUMX_INPUT A real*4 1-D array of size NVARS; represents SUMX as output by the space or time binner for all level-3 geophysical variables (Js) of a given bin IDX being read.

SUMXX_INPUT A real*4 1-D array of size NVARS; represents SUMXX as output by the space or time binner for all level-3 geophysical variables (Js) of a given bin IDX being read.

N_INPUT An integer*2 word; represents N as output by the space or time binner for a given bin IDX being read.

NSEG_INPUT An integer*2 word; represents NSEG as output by the space or time binner for a given bin IDX being read.

W_INPUT A real word; represents W as output by the space or time binner for a given bin IDX being read.

TT_INPUT An integer*2 word; represents TT as output by the space or time binner for a given bin IDX being read.

IDX An integer*4 word representing the index number of the bin being read from the input level-3 product.

Output Variables

XMEAN A real*4 1-D array of size NVARS; represents the mean of the weighted cumulative values of the level-3 geophysical variables (Js).

SIGMA A real*4 1-D array of size NVARS; represents the standard deviation for the weighted cumulative values of the level-3 geophysical variables (Js).

XMEDN A real*4 1-D array of size NVARS; represents the median of the weighted cumulative values of the level-3 geophysical variables (Js).

XMODE A real*4 1-D array of size NVARS; represents the mode of the weighted cumulative values of the level-3 geophysical variables (Js).

N_INPUT An integer*2 word; represents N as output by the space or time binner for a given bin IDX being output.

NSEG_INPUT An integer*2 word; represents NSEG as output by the space or time binner for a given bin IDX being output.

TT_INPUT An integer*2 word; represents TT as output by the space or time binner for a given bin IDX being output.

IDX An integer*4 word representing the index number of the bin being output.

Other Variables

B Counter index of bins read from input product. Range is from 1 to NBINS.

J Counter index of geophysical variables that have been binned. Range is from 1 to NVARS.

AVLOGS A real*4 word that represents the mean of the weighted logs for a geophysical variable J of bin B being processed. Used to calculate XMEAN, SIGMA, XMEDN, and XMODE.

VRLOGS A real*4 word that represents the variance of the weighted logs for a geophysical variable J of bin B being processed. Used to calculate XMEAN, SIGMA, and XMODE.

```
#
# Input information for each bin
#
read from metadata of binned input products: NBINS
do from B=1 to NBINS
   read from bin B:
      IDX, N_INPUT, NSEG_INPUT, W_INPUT, TT_INPUT
      SUMX_INPUT(J), SUMXX_INPUT(J), for J=1 to NVARS
   end read
#
#  Calc. mean, std.dev., median and mode, and then output
#
   do from J=1 to NVARS
      AVLOGS = SUMX_INPUT(J)/W_INPUT
      VRLOGS = (SUMXX_INPUT(J)/W_INPUT) - (AVLOGS*AVLOGS)
      XMEAN(J) = exponential(AVLOGS + (VRLOGS/2.))
      SIGMA(J) = XMEAN(J)*sqroot(exponential(VRLOGS) - 1)
      XMEDN(J) = exponential(AVLOGS)
      XMODE(J) = exponential(AVLOGS - VRLOGS)
   end do
   write to screen or file useful info for bin IDX
     IDX, N_INPUT, NSEG_INPUT, TT_INPUT
    XMEAN(J), SIGMA(J), XMEDN(J), XMODE(J), for J=1 to NVARS
   end write
end do
```

Glossary

AVHRR	Advanced Very High Resolution Radiometer
CZCS	Coastal Zone Color Scanner
DSP	Not an acronym; the name of a software package developed at RSMAS.
GAC	Global Area Coverage
GMT	Greenwich Mean Time
HRPT	High Resolution Picture Transmission
IFOV	Instantaneous Field-of-View
ISCCP	International Satellite Cloud Climatology Project
LAC	Local Area Coverage
MARMAP	Marine Resources Monitoring, Assessment, and Prediction
MODIS	Moderate Resolution Imaging Spectroradiometer
RSMAS	Rosenstiel School of Marine and Atmospheric Science
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SEEP	Shelf Edge Exchange Program
SST	Sea Surface Temperature
TDI	Time Delay and Integration

Symbols

A_g	CZCS pigment algorithm constant (global).
A_r	CZCS pigment algorithm constant (regional).
AVG	Arithmetic average based on LAC data.
AVG4	Arithmetic average based on GAC data.
b	Bin index number.
B_g	CZCS pigment algorithm constant (global).
B_r	CZCS pigment algorithm constant (regional).
$\langle \mathrm{Chl} \rangle_{\mathrm{tot}}$	Integral euphotic chlorophyll.
CHL	Chlorophyll concentration.
CHL_{13}	Pigment concentration calculated from CZCS bands 1 and 3.
CHL_{23}	Pigment concentration calculated from CZCS bands 2 and 3.

DIFF1	Relative difference between MLE4 and AVG4.
DIFF2	Relative difference between MED4 and AVG4.
E[X]	Expected value of x.
ERROR	Relative error, in percent, of the estimated mean from the arithmetic mean.
FNC	Function of vector variable X using LAC data.
FNC4	Function of vector variable X using GAC data.
$\mathrm{IC_K}$	Integrated chlorophyll concentration over the first optical depth.
ID	Mooring identification number.
K_{490}	The diffuse attenuation coefficient at 490 nm.
L	Bin dimension in kilometers.
$L_{WN}(\lambda_i)$	Normalized water-leaving radiances in i bands (1–5).
$L_a(\lambda_i)$	Atmospheric aerosol radiances in i bands (6–8).
L_W	Water-leaving radiance.
m_i	Central moment of x.
m_x	Sample mean of the natural logarithm of x.
m_y	Sample mean of the natural logarithm of y.
m_r	Sample mean of regional ln (pigment).
MLE	Maximum likelihood estimator of LAC data.
MLE4	Maximum likelihood estimator of GAC data.
MED	Geometric mean or median of LAC data.
MED4	Geometric mean or median of GAC data.
n	Sample size.
n_i	The number of pixels per bin on orbit i.
n	The number of days used for temporal averaging.
N	The number of orbits contributing to the temporal mean.
P	The proportion of the distribution that is mode 1.
PIG	CZCS pigment-like concentration.
$\mathrm{PIG_r}$	Pigment calculated with regionally-derived parameters.
s_r^2	The sample variance of regional ln (pigment).
s_x^2	The sample variance of the natural logarithm of x.
s_y^2	The sample variance of the natural logarithm of y.
S_1	The weighted sum of variable x.
S_2	The weighted sum of variable y.
SD_x	The standard deviation of x.
SD_y	The standard deviation of y.
t_i	The time at which orbit i was acquired.
T	A 16-bit time distribution variable.
V	The 8-bit image value of a pixel, i.e., gray level.
w_i	The weight factor for orbit i.
W	The sum of the weighting factors.
x	The natural logarithm of X.
X	Any random variable whose distribution is unknown.
X	A level-2 variable.
$\overline{X}$	The true mean of a level-2 variable.
$\overline{X}_{\mathrm{avg}}$	The arithmetic average of X.
$\overline{X}_{\mathrm{geom}}$	The geometric mean of X.
$\overline{X}_{\mathrm{mle}}$	The maximum likelihood estimator of $\overline{X}$.
$\overline{X}_{\mathrm{med}}$	The median of X.
$\overline{X}_{\mathrm{mod}}$	The mode of X.
$\overline{X}_{\mathrm{est}}$	The estimated mean of X.
X	The vector of standard level-2 variables.
Y	Any function of a level-2 variable X.
$\overline{Y}$	The true mean of Y.
$\overline{Y}_{\mathrm{avg}}$	The arithmetic average of Y.
$\overline{Y}_{\mathrm{mle}}$	The maximum likelihood estimator of Y.
$\overline{Y}_{\mathrm{fnc}}$	The arithmetic mean of FNC.
Z_e	The euphotic depth (depth to 1% light level).
λ_1	Wavelength of 440 nm.
λ_2	Wavelength of 520 nm.
λ_3	Wavelength of 550 nm.
$\tau_a(865)$	The aerosol optical thickness at 865 nm.

References

Aitchison, J., and J.A.C. Brown, 1957: *The Lognormal Distribution.* Cambridge University Press, 176 pp.

Baker, M.A., and C.H. Gibson, 1987: Sampling turbulence in the stratified ocean: statistical consequences of strong intermittency. *J. of Phys. Oceanogr.,* **17,** 1,817-1,836.

Balch, W., R. Evans, J. Brown, G. Feldman, C. McClain, and W. Esaias, 1992: The remote sensing of ocean primary productivity: use of a new data compilation to test satellite algorithms. *J. Geophys. Res.,* **97,** 2,279-2,293.

Bannister, T.T., 1974: Production equations in terms of chlorophyll concentration, quantum yield, and upper limit to production. *Limnol. Oceanogr.,* **19,** 1–12.

Campbell, J.W., and J.E. O'Reilly, 1988: Role of satellites in estimating primary productivity on the northwest Atlantic continental shelf. *Cont. Shelf Res.,* **8,** 179–204.

——, and T. Aarup, 1992: New production in the North Atlantic derived from seasonal patterns of surface chlorophyll. *Deep-Sea Res.,* **39,** 1,669–1,694.

Chelton, D.B., and M.G. Schlax, 1991: Estimation of time averages from irregularly spaced observations: with application to Coastal Zone Color Scanner estimates of chlorophyll concentration. *J. Geophys. Res.,* **96,** 14,669–14,692.

Clark, D.K., 1981: Phytoplankton algorithms for the Nimbus-7 CZCS. *Oceanography from Space,* J.F.R. Gower, Ed., Plenum Press, 227–238.

Crow, E.L., and K. Shimizu, editors, 1988: *Lognormal Distributions: Theory and Applications,* Marcel Dekker, Inc., New York, 387 pp.

Eppley, R.W., E. Stewart. M.R. Abbott, and U. Heyman, 1985: Estimating ocean primary production from satellite chlorophyll. Introduction to regional differences and statistics for the Southern California Bight. *J. Plankton Res.,* **7,** 57–70.

Esaias W.E., G.C. Feldman, C.R. McClain, and J.A. Elrod, 1986: Monthly satellite-derived phytoplankton pigment distribution for the North Atlantic Ocean Basin. *Eos, Trans. AGU,* **67,** 835–837.

Feldman G., N. Kuring, C. Ng, W.E. Esaias, C. McClain, J. Elrod, N. Maynard, D. Endres, R. Evans, J. Brown, S. Walsh, M. Carle, and G. Podesta, 1989: Ocean color: Availability of the global data set. *Eos, Trans. AGU,* **70,** 634–641.

Gordon, H.R., and A.Y. Morel, 1983: *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery.* Springer-Verlag, New York, 114 pp.

——, D.K. Clark, J.W. Brown, O.B. Brown, R.H. Evans, and W.W. Broenkow, 1983: Phytoplankton pigment concentrations in the Middle Atlantic Bight: Comparison of ship determinations and CZCS estimates. *Appl. Opt.,* **22,** 20–36.

——, and D.J. Castaño, 1987: Coastal Zone Color Scanner atmospheric correction algorithm: multiple scattering effects. *Appl. Opt.,* **26,** 2,111–2,122.

——, O.B. Brown, R.H. Evans, J.W. Brown, R.C. Smith, K.S. Baker, and D.K. Clark, 1988: A semianalytic radiance model of ocean color. *J. Geophys. Res.,* **93,** 10,909–10,924.

Journal, A.G., 1989: Fundamentals of Geostatistics in Five Lessons, Short Course. *Geology: Vol. 8,* American Geophysical Union, Washington, D.C., 40 pp.

Kirk, J.T.O., 1983: *Light and Photosynthesis in Aquatic Ecosystems.* Cambridge University Press, Cambridge, 401 pp.

Lewis M.R., N. Kuring, and C.S. Yentsch, 1988: Global patterns of ocean transparency: Implications for the new production of the open ocean. *J. Geophys. Res.,* **93,** 6,847–6,856.

McClain, C.R., E-n. Yeh, and G. Fu, 1992: An Analysis of GAC Sampling Algorithms: A Case Study. *NASA Technical Memorandum 104566, Vol. 4,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 21 pp.

Medeiros, W.H., and C.D. Wirick, 1992: SEEP II: Shelf Edge Exchange Processes II. Chlorophyll *a* Fluorescence, Temperature, and Beam Attenuation Measurements from Moored Fluorometers. *BNL 47211 Informal Report,* Brookhaven National Laboratory, Upton, New York, 205 pp.

Morel, A., and J.-F. Berthon, 1989: Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. *Limnol. Oceanogr.,* **34,** 1,545–1,562.

Platt, T., 1986: Primary production of the ocean water column as a function of surface light intensity: Algorithms for remote sensing. *Deep-Sea Res.,* **33**, 1–15.

——, and S. Sathyendranath, 1988: Oceanic primary production: estimation by remote sensing at local and regional scales. *Science,* **241,** 1,613–1,620.

RSMAS, 1990: *DSP User's Manual.* Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida, 255 pp.

Smith, R.C., R.W. Eppley, and K.S. Baker, 1982: Correlation of primary production as measured aboard ship in southern California coastal waters and as estimated from satellite chlorophyll images. *Mar. Biol.,* **66,** 281–288.

——, and K.S. Baker, 1978: The bio-optical state of ocean waters and remote sensing. *Limnol. Oceanogr.,* **23,** 247–259.

Yentsch, C.S., 1990: Estimates of 'new production' in the Mid-North Atlantic. *J. Plankton Res.,* **12,** 717–734.

The SeaWiFS Technical Report Series

Vol. 1

Hooker, S.B., W.E. Esaias, G.C. Feldman, W.W. Gregg, and C.R. McClain, 1992: An Overview of SeaWiFS and Ocean Color. *NASA Tech. Memo. 104566, Vol. 1,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 24 pp., plus color plates.

Vol. 2

Gregg, W.W., 1992: Analysis of Orbit Selection for SeaWiFS: Ascending vs. Descending Node. *NASA Tech. Memo. 104566, Vol. 2,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 16 pp.

Vol. 3

McClain, C.R., W.E. Esaias, W. Barnes, B. Guenther, D. Endres, S.B. Hooker, G. Mitchell, and R. Barnes, 1992: Calibration and Validation Plan for SeaWiFS. *NASA Tech. Memo. 104566, Vol. 3,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 41 pp.

Vol. 4

McClain, C.R., E. Yeh, and G. Fu, 1992: An Analysis of GAC Sampling Algorithms: A Case Study. *NASA Tech. Memo. 104566, Vol. 4,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 22 pp., plus color plates.

Vol. 5

Mueller, J.L., and R.W. Austin, 1992: Ocean Optics Protocols for SeaWiFS Validation. *NASA Tech. Memo. 104566, Vol. 5,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 43 pp.

Vol. 6

Firestone, E.R., and S.B. Hooker, 1992: SeaWiFS Technical Report Series Summary Index: Volumes 1–5. *NASA Tech. Memo. 104566, Vol. 6,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 9 pp.

Vol. 7

Darzi, M., 1992: Cloud Screening for Polar Orbiting Visible and IR Satellite Sensors. *NASA Tech. Memo. 104566, Vol. 7,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 7 pp.

Vol. 8

Hooker, S.B., W.E. Esaias, and L.A. Rexrode, 1993: Proceedings of the First SeaWiFS Science Team Meeting. *NASA Tech. Memo. 104566, Vol. 8,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 61 pp.

Vol. 9

Gregg, W.W., F.C. Chen, A.L. Mezaache, J.D. Chen, J.A. Whiting, 1993: The Simulated SeaWiFS Data Set, Version 1. *NASA Tech. Memo. 104566, Vol. 9,* S.B. Hooker, E.R. Firestone, and A.W. Indest, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 17 pp.

Vol. 10

Woodward, R.H., R.A. Barnes, C.R. McClain, W.E. Esaias, W.L. Barnes, and A.T. Mecherikunnel, 1993: Modeling of the SeaWiFS Solar and Lunar Observations. *NASA Tech. Memo. 104566, Vol. 10,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 26 pp.

Vol. 11

Patt, F.S., C.M. Hoisington, W.W. Gregg, and P.L. Coronado, 1993: Analysis of Selected Orbit Propagation Models for the SeaWiFS Mission. *NASA Tech. Memo. 104566, Vol. 11,* S.B. Hooker, E.R. Firestone, and A.W. Indest, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 16 pp.

Vol. 12

Firestone, E.R., and S.B. Hooker, 1993: SeaWiFS Technical Report Series Summary Index: Volumes 1–11. *NASA Tech. Memo. 104566, Vol. 12,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 28 pp.

Vol. 13

McClain, C.R., K.R. Arrigo, J. Comiso, R. Fraser, M. Darzi, J.K. Firestone, B. Schieber, E-n. Yeh, and C.W. Sullivan, 1994: Case Studies for SeaWiFS Calibration and Validation, Part 1. *NASA Tech. Memo. 104566, Vol. 13,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 52 pp., plus color plates.

Vol. 14

Mueller, J.L., 1993: The First SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-1, July 1992. *NASA Tech. Memo. 104566, Vol. 14,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 60 pp.

Vol. 15

Gregg, W.W., F.S. Patt, and R.H. Woodward, 1994: The Simulated SeaWiFS Data Set, Version 2. *NASA Tech. Memo. 104566, Vol. 15,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 42 pp., plus color plates.

Vol. 16

Mueller, J.L., B.C. Johnson, C.L. Cromer, J.W. Cooper, J.T. McLean, S.B. Hooker, and T.L. Westphal, 1994: The Second SeaWiFS Intercalibration Round-Robin Experiment, SIRREX-2, June 1993. *NASA Tech. Memo. 104566, Vol. 16,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 121 pp.

Vol. 17

Abbott, M.R., O.B. Brown, H.R. Gordon, K.L. Carder, R.E. Evans, F.E. Muller-Karger, and W.E. Esaias, 1994: Ocean Color in the 21st Century: A Strategy for a 20-Year Time Series. *NASA Tech. Memo. 104566, Vol. 17,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 20 pp.

Vol. 18

Firestone, E.R., and S.B. Hooker, 1995: SeaWiFS Technical Report Series Summary Index: Volumes 1–17. *NASA Tech. Memo. 104566, Vol. 18,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 47 pp.

Vol. 19

McClain, C.R., R.S. Fraser, J.T. McLean, M. Darzi, J.K. Firestone, F.S. Patt, B.D. Schieber, R.H. Woodward, E-n. Yeh, S. Mattoo, S.F. Biggar, P.N. Slater, K.J. Thome, A.W. Holmes, R.A. Barnes, and K.J. Voss, 1994: Case Studies for SeaWiFS Calibration and Validation, Part 2. *NASA Tech. Memo. 104566, Vol. 19,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 73 pp.

Vol. 20

Hooker, S.B., C.R. McClain, J.K. Firestone, T.L. Westphal, E-n. Yeh, and Y. Ge, 1994: The SeaWiFS Bio-Optical Archive and Storage System (SeaBASS), Part 1. *NASA Tech. Memo. 104566, Vol. 20,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 40 pp.

Vol. 21

Acker, J.G., 1994: The Heritage of SeaWiFS: A Retrospective on the CZCS NIMBUS Experiment Team (NET) Program. *NASA Tech. Memo. 104566, Vol. 21,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 43 pp.

Vol. 22

Barnes, R.A., W.L. Barnes, W.E. Esaias, and C.R. McClain, 1994: Prelaunch Acceptance Report for the SeaWiFS Radiometer. *NASA Tech. Memo. 104566, Vol. 22,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 32 pp.

Vol. 23

Barnes, R.A., A.W. Holmes, W.L. Barnes, W.E. Esaias, C.R. McClain, and T. Svitek, 1994: SeaWiFS Prelaunch Radiometric Calibration and Spectral Characterization. *NASA Tech. Memo. 104566, Vol. 23,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 55 pp.

Vol. 24

Firestone, E.R., and S.B. Hooker, 1995: SeaWiFS Technical Report Series Summary Index: Volumes 1–23. *NASA Tech. Memo. 104566, Vol. 24,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 36 pp.

Vol. 25

Mueller, J.L., and R.W. Austin, 1995: Ocean Optics Protocols for SeaWiFS Validation, Revision 1. *NASA Tech. Memo. 104566, Vol. 25,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 66 pp.

Vol. 26

Siegel, D.A., M.C. O'Brien, J.C. Sorensen, D.A. Konnoff, E.A. Brody, J.L. Mueller, C.O. Davis, W.J. Rhea, and S.B. Hooker, 1995: Results of the SeaWiFS Data Analysis Round-Robin (DARR-94), July 1994. *NASA Tech. Memo. 104566, Vol. 26,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 58 pp.

Vol. 27

Mueller, J.L., R.S. Fraser, S.F. Biggar, K.J. Thome, P.N. Slater, A.W. Holmes, R.A. Barnes, C.T. Weir, D.A. Siegel, D.W. Menzies, A.F. Michaels, and G. Podesta, 1995: Case Studies for SeaWiFS Calibration and Validation, Part 3. *NASA Tech. Memo. 104566, Vol. 27,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 46 pp.

Vol. 28

McClain, C.R., K.R. Arrigo, W.E. Esaias, M. Darzi, F.S. Patt, R.H. Evans, J.W. Brown, C.W. Brown, R.A. Barnes, and L. Kumar, 1995: SeaWiFS Algorithms, Part 1. *NASA Tech. Memo. 104566, Vol. 28,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 38 pp., plus color plates.

Vol. 29

Aiken, J., G.F. Moore, C.C. Trees, S.B. Hooker, and D.K. Clark, 1995: The SeaWiFS CZCS-Type Pigment Algorithm. *NASA Tech. Memo. 104566, Vol. 29,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 34 pp.

Vol. 30

Firestone, E.R., and S.B. Hooker, 1995: SeaWiFS Technical Report Series Summary Index: Volumes 1–29. *NASA Tech. Memo. 104566, Vol. 30,* S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, (in production).

Vol. 31

Barnes, R.A., A.W. Holmes, and W.E. Esaias, 1995: Stray Light in the SeaWiFS Radiometer. *NASA Tech. Memo. 104566, Vol. 31,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 76 pp.

Vol. 32

Campbell, J.W., J.M. Blaisdell, and M. Darzi, 1995: Level-3 SeaWiFS Data Products: Spatial and Temporal Binning Algorithms. *NASA Tech. Memo. 104566, Vol. 32,* S.B. Hooker, E.R. Firestone, and J.G. Acker, Eds., NASA Goddard Space Flight Center, Greenbelt, Maryland, 73 pp., plus color plates.

COLOR PLATES

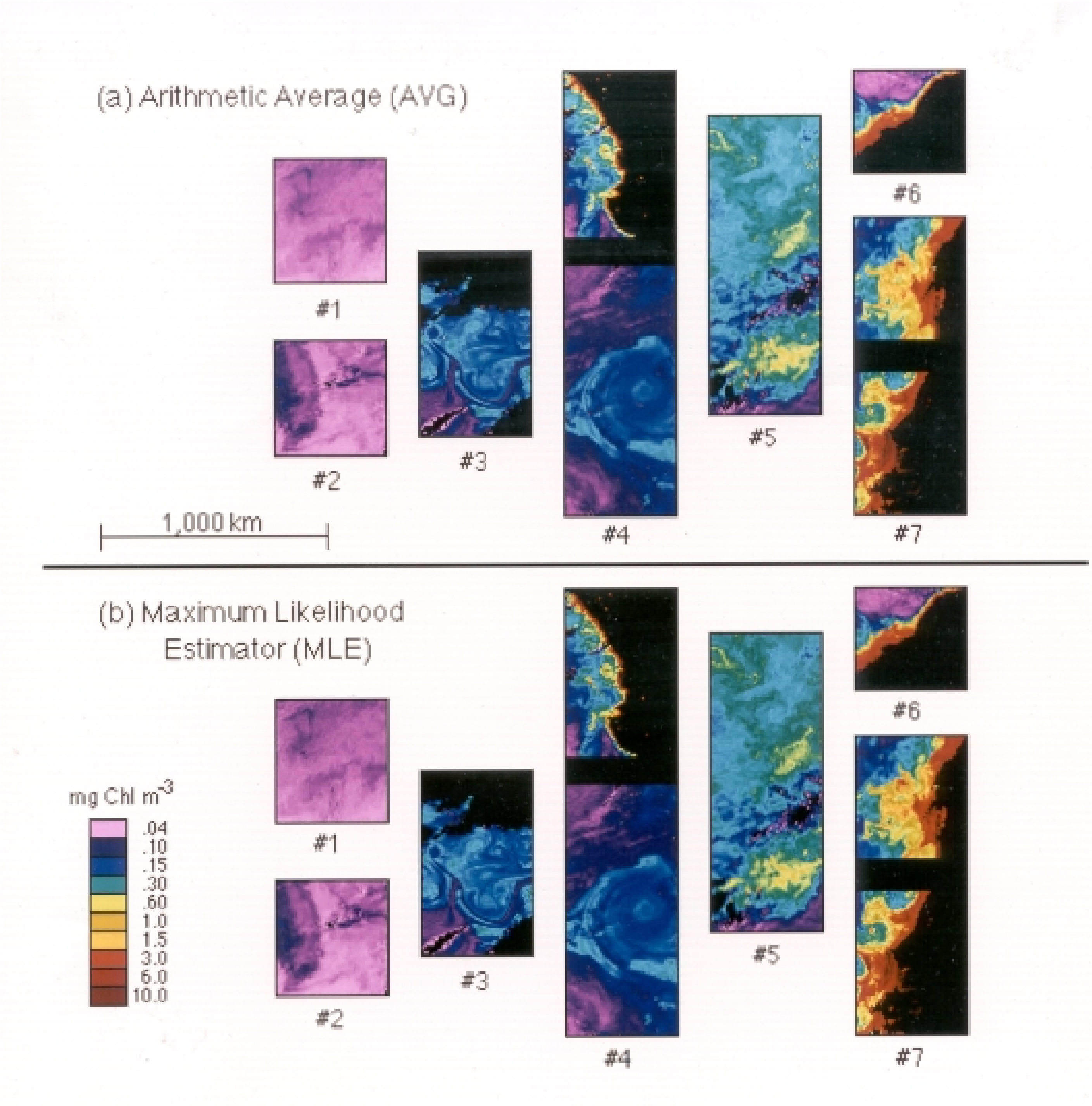


PLATE 1. Mean CHL images derived from the AVG and MLE estimators for the seven CZCS scenes listed in Table 1.

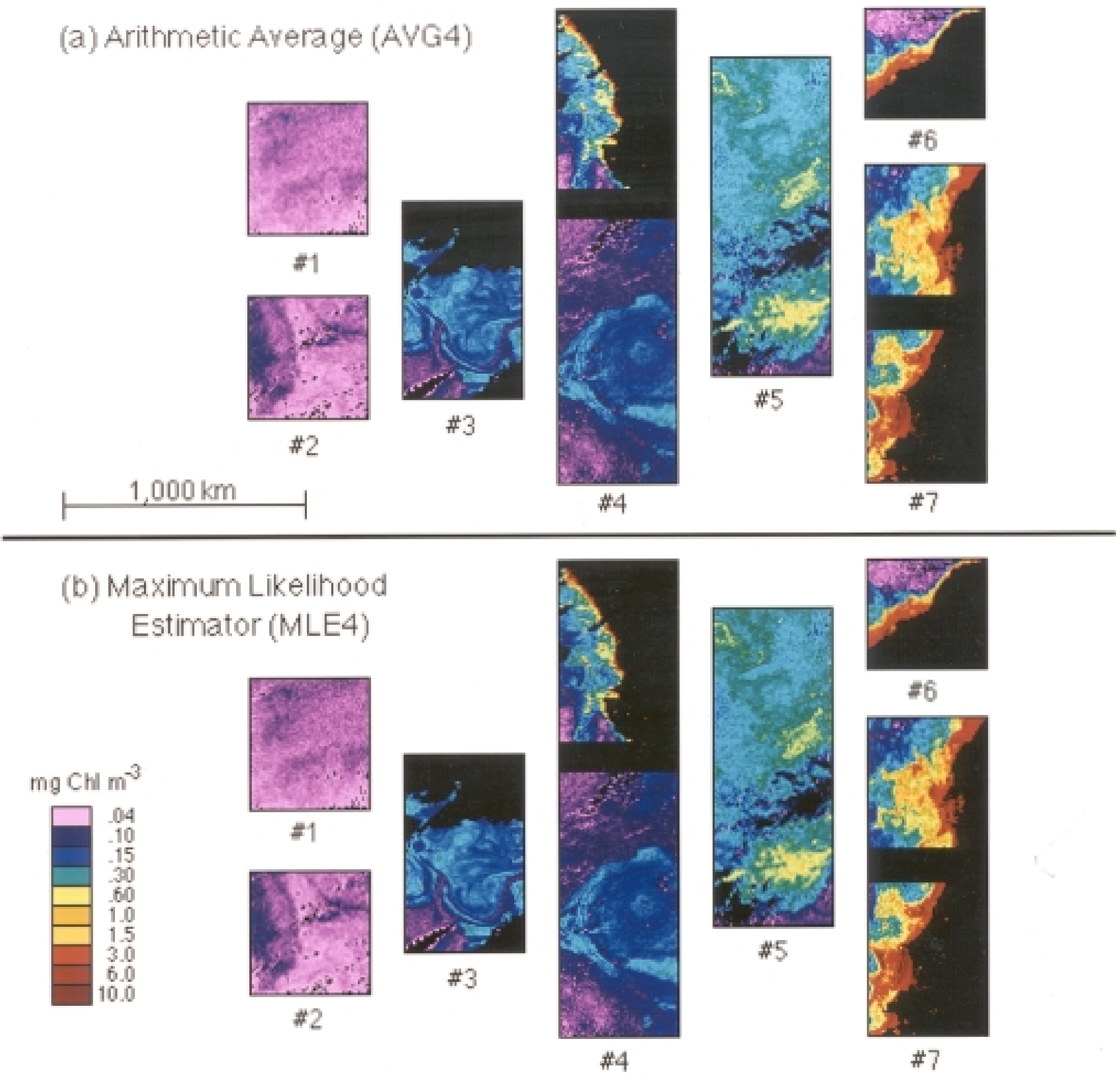


PLATE 2. Mean CHL images derived from the AVG4 and MLE4 estimators for the seven CZCS scenes listed in Table 1.

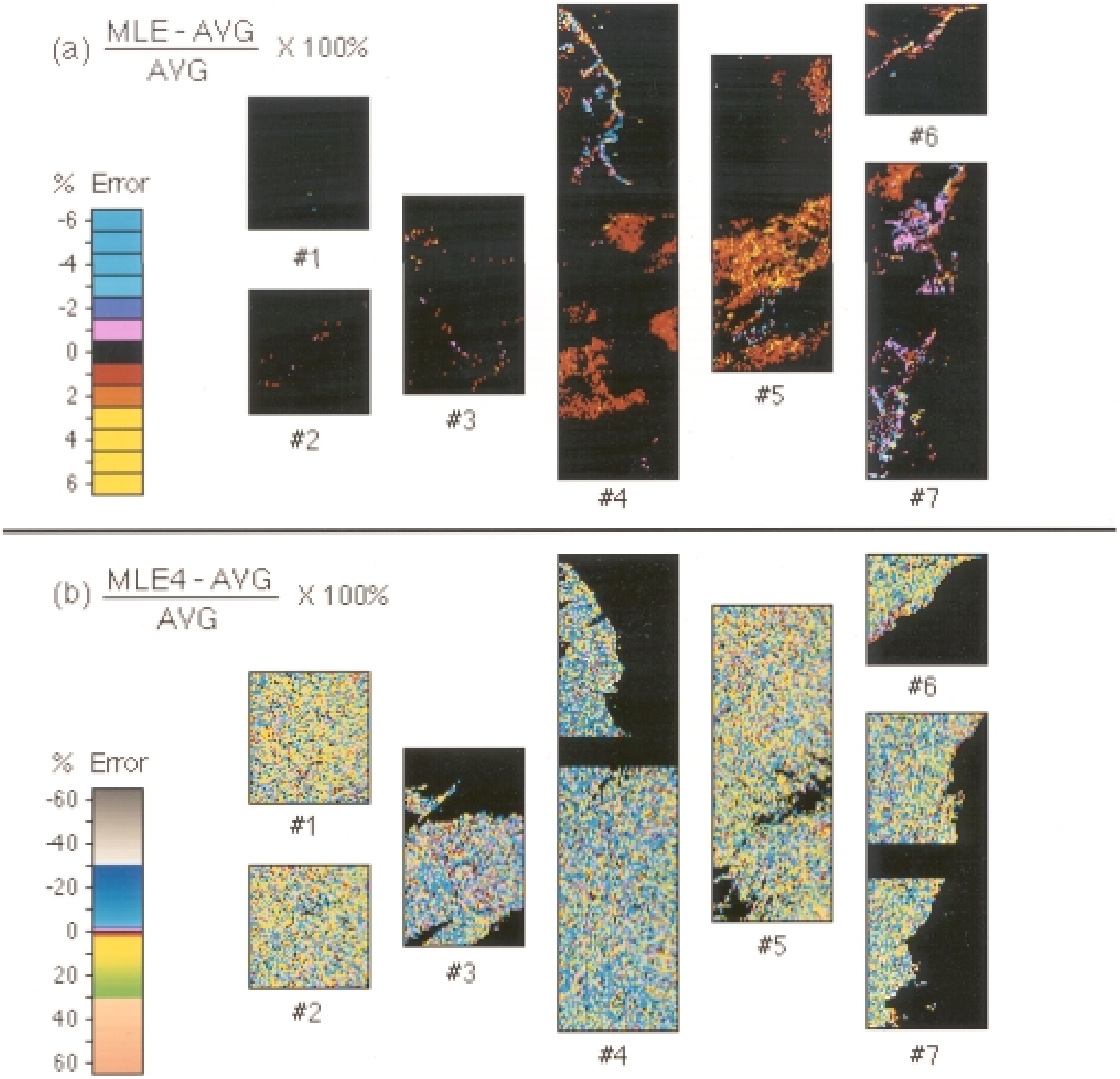

(a) $\frac{\text{MLE - AVG}}{\text{AVG}}$ X 100%
% Error
-6
-4
-2
0
2
4
6
#1
#2
#3
#4
#5
#6
#7
(b) $\frac{\text{MLE4 - AVG}}{\text{AVG}}$ X 100%
% Error
-60
-40
-20
0
20
40
60
#1
#2
#3
#4
#5
#6
#7

PLATE 3. Differences between level-3 means derived from MLE and AVG estimators (upper images) and difference between MLE4 and AVG4 estimates (lower images).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** July 1995	**3. REPORT TYPE AND DATES COVERED** Technical Memorandum

4. TITLE AND SUBTITLE

SeaWiFS Technical Report Series
Volume 32–Level-3 SeaWiFS Data Products:
Spatial and Temporal Binning Algorithms

5. FUNDING NUMBERS

Code 970.2

6. AUTHOR(S)

Janet W. Campbell, John M. Blaisdell, and Michael Darzi

Series Editors: Stanford B. Hooker and Elaine R. Firestone
Technical Editor: James G. Acker

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Laboratory for Hydrospheric Processes
Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PERFORMING ORGANIZATION REPORT NUMBER

95B000116

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, D.C. 20546–0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

TM–104566, Vol. 32

11. SUPPLEMENTARY NOTES

Elaine R. Firestone, John M. Blaisdell, and Michael Darzi: General Sciences Corporation, Laurel, Maryland; Janet W. Campbell: University of New Hampshire, Durham, New Hampshire; and James G. Acker: Hughes STX, Lanham, Maryland

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified–Unlimited
Subject Category 48
Report is available from the Center for AeroSpace Information (CASI),
7121 Standard Drive, Hanover, MD 21076–1320; (301)621-0390

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The level-3 data products from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) are statistical data sets derived from level-2 data. Each data set will be based on a fixed global grid of equal-area bins that are approximately 9x9 km^2. Statistics available for each bin include the sum and sum of squares of the natural logarithm of derived level-2 geophysical variables where sums are accumulated over a binning period. Operationally, products with binning periods of 1 day, 8 days, 1 month, and 1 year will be produced and archived. From these accumulated values and for each bin, estimates of the mean, standard deviation, median, and mode may be derived for each geophysical variable. This report contains two major parts: the first (Section 2) is intended as a users' guide for level-3 SeaWiFS data products. It contains an overview of level-0 to level-3 data processing, a discussion of important statistical considerations when using level-3 data, and details of how to use the level-3 data. The second part (Section 3) presents a comparative statistical study of several binning algorithms based on CZCS and moored fluorometer data. The operational binning algorithms were selected based on the results of this study.

14. SUBJECT TERMS

SeaWiFS, Oceanography, Data Products, Algorithms, Spatial Binning, Temporal Binning
Level-3, Statistics, Empirical Basis

15. NUMBER OF PAGES

73

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** Unlimited