Case Study 19

Determining Dynamic Biogeographic Regions using Remote Sensing Data

E. Santamaría-del-Ángel*[1], A. González-Silvera[1], R. Millán-Núñez[1], M. E. Callejas-Jiménez[1], R. Cajal-Medrano[1]

19.1 Background Information

One of the most important topics in natural resource management is the determination of biogeographic regions (BGR). Each BGR is the result of a particular combination of physical, chemical and biological factors and the variation of one of these factors can have different effects on each region. This characteristic trait makes biogeographic regionalization a basic tool of environmental modeling. Topics such as the carbon cycle and global climate change, and it effects on fisheries (IOCCG, 2008; 2009a) can be addressed using BGRs to bring together *in situ* data and data obtained from remote sensor observations (IOCCG, 2009b). Defining a BGR is simpler in terrestrial ecosystems than in marine environments because ecotypes such as forests, jungles and deserts can be used to define a BGR. The highly-dynamic nature of the marine environment makes defining BGRs a great challenge, which requires different approximations.

There are two basic approximations that can be used to define marine BGR. The first uses a large, *in situ* database (direct approximation) and the second uses data obtained from remote sensor observations (indirect approximation). Applying the direct approximation requires a database that is robust in both space and time. Such a database is generated by monitoring programs such as CalCOFI (California Cooperative Oceanic Fisheries Investigations) within the California Current systems. Millán-Núñez et al. (1997) surveyed the area using the CalCOFI database and determined six BGR. Not all areas are candidates for direct approximation, because of the high costs. For areas that do not have continuous monitoring stations, the only way to determine BGR is to use indirect approximation. The basic idea of this technique is to use satellite images as virtual maps to generate a database of the surface layer of the oceans. A BGR can be determined by the association patterns

[1]Facultad de Ciencias Marinas. Universidad Autónoma de Baja California (FCM-UABC)[1] Carretera Tijuana-Ensenada Km 106 C.P. 22800 Ensenada BC. Mexico **Email address:* santamaria@uabc.edu.mx

yielded by concurrent temporal and spatial variation between geo-locations.

Different mathematical approximations can be used to determine the association patterns between geo-locations. Santamaría-del-Ángel et al. (1994) defined 14 BGRs in the Gulf of California using a Principal Component Analysis (PCA) derived from data of weekly composites of CZCS images with a resolution of 4-km per pixel. In a survey of the same region, with the same sensor, but with monthly compositions of 9-km per pixel, Lunch-Cota (2000) determined only two BGRs using a cluster analysis. The difference between these two approximations is that monthly composites mask the effects of tides, which are one of the main components of turbulent kinetic energy in the Gulf of California. The relationship between turbulent kinetic energy and BGR is directly proportional to the number of BGRs, but inversely proportional to the size of the BGR. In other words, locations with high turbulent kinetic energy should have many, small BGRs, and locations with low turbulent kinetic energy should contain few, large BGRs.

One of the main characteristics of these approximations is that they feature static borders, which do not reflect the dynamic nature of the ocean. González-Silvera et al. (2006) proposed a new approximation to determine dynamic biogeographic regions (DBGRs) in the subtropical convergence zone in the southwest Atlantic ocean. Their approximation was based on a combination of images from SeaWiFS (Chl-*a*) and the Advanced Very High Resolution Radiometer (AVHRR, SST) in monthly composites with 1-km per pixel resolution. They proposed that the numerical combination of these two variables, represented by the first Standardized Empirical Orthogonal Function (SEOF), integrates the biological traits of the area (represented by Chl-*a*) with the physical traits (SST) and thus describes the synchronous variation between regions. Dynamic region determination (IOCCG, 2009b) can be based on the SEOF approximation, a powerful tool that accounts for the space/time variations of an area. It also allows daily determination of DBGR. This case study gives a step-by-step description of how to carry out dynamic regionalization using the SEOF approximation, highlighting what they are, how they are calculated, and how to interpret results.

To understand an SEOF, we must focus on the Empirical Orthogonal Function (EOF) and how it is calculated. EOF is the basis of Principal Component Analysis (PCA) and Factor Analysis (FA) and was first proposed by Kendall in 1930. EOFs can be described by first-degree polynomial equations. There are as many EOFs as variables, and the first EOF is the best linear combination of the collection of variables, as it has a higher total variance than the rest of the EOFs derived from the data.

Let us consider a hypothetical case with j variables, each with n_i occurrences (matrix M_1, see Table 19.1). The idea is to group the variables that display patterns of simultaneous variation or similarity. The original development considers the covariance (*Cov*) between two variables that is calculated with the formula shown in Equation 19.1.

Table 19.1 Matrix M_1 representing j variables each with n_i observations:

$$\begin{bmatrix} A & B & C & & j \\ n_{1A} & n_{1B} & n_{1C} & \cdots & n_{1j} \\ n_{2A} & n_{2B} & n_{2C} & & n_{2j} \\ & \vdots & & \ddots & \vdots \\ n_{iA} & n_{iB} & n_{iC} & \cdots & n_{ij} \end{bmatrix}$$

$$Cov = \frac{\sum((y_i - \overline{y}) * (x_i - \overline{x}))}{n_i - 1} \tag{19.1}$$

where x_i and y_i represent each of the two variables, while $\overline{x}$ and $\overline{y}$ are the respective means of the variables. Calculating *Cov* for all the variables from A to j, we can construct a matrix of *Cov*. However, since the variables can be of a different nature (e.g., Chl-*a*, SST or water leaving radiance), it is necessary to normalize the original values to the same scale. The transformations must also respect the integrity of the original data distribution. One of the transformation that easily fulfills these requirements is the Z transformation (Equation 19.2);

$$Z = \frac{(x_i - \overline{x})}{SD_x} \tag{19.2}$$

where x_i represents each of the variables, $\overline{x}$ is the mean and SD_x is the standard deviation of the variable. Applying Equation 19.2 to the original data (M_1), we obtain a matrix of normalized data M_2 (Table 19.2). If *Cov* is calculated with the normalized variables, we can construct the corresponding matrix of normalized variables (M_3, Table 19.3).

Table 19.2 Matrix M_2 representing normalized variables.

$$\begin{bmatrix} Z_A & Z_B & Z_C & & Z_j \\ Z_{1A} & Z_{1B} & Z_{1C} & \cdots & Z_{1j} \\ Z_{2A} & Z_{2B} & Z_{2C} & & Z_{2j} \\ & \vdots & & \ddots & \vdots \\ Z_{iA} & Z_{iB} & Z_{iC} & \cdots & Z_{ij} \end{bmatrix}$$

M_3 is a symmetrical square matrix (3×3), with a diagonal of one. Note that the symmetry is with respect to the principal diagonal and that M_3 is its own transposed

Table 19.3 *Cov* matrix of the normalized variables M_2. To simplify this explanation, $j = 3$ thus yielding three variables to associate.

$$\begin{array}{c} Z_{var} \\ Z_A \\ Z_B \\ Z_C \end{array} \begin{array}{c} \begin{array}{ccc} Z_A & Z_B & Z_C \end{array} \\ \begin{bmatrix} 1 & Cov_{Z_B Z_A} & Cov_{Z_C Z_A} \\ Cov_{Z_A Z_B} & 1 & Cov_{Z_C Z_B} \\ Cov_{Z_A Z_C} & Cov_{Z_B Z_C} & 1 \end{bmatrix} \end{array}$$

matrix (For example: $M_3 = M_3 t$) and is therefore a mirror matrix. The matrix has a diagonal of one because $Cov_{AA} = SD_A$, which is why $Cov_{Z_A Z_A} = SD_{Z_A} = 1$.

One of the basic theorems that deals with this type of matrix is the *Finite Dimension Spectral Theorem*, which states that every symmetric matrix with real entrances can be characterized (diagonalized) by an orthogonal matrix. A square matrix is diagnosable if its j vectors are linearly independent. Therefore, we can characterize this symmetric square matrix with a diagonal of one (M_3) through the linear algebra resolution of its orthogonal matrix as defined by its own values (*eigenvalues*) and its own vectors (*eigenvectors*).

It is worth noting that the eigenvalues and eigenvectors calculated from the matrix $Cov_{Z_{Var}} M_3$ are the same as those derived from a symmetric square matrix of correlation of the original variables (M_1) that avoid normalization (M_Z). The correlation matrix can be derived by Pearson's linear parameter correlation coefficient (r_P, Equation 19.3):

$$r_P = \frac{Cov_{AB}}{SD_A * SD_B}, \tag{19.3}$$

where Cov_{AB} = covariance of variables A and B, SD_A = standard deviation of A, and SD_B = standard deviation of B. r_P denotes the rate of covariance between the dispersion product of each of the variables. It can also express the percentage of total variability of the data that can be explained by linear models. Another correlation coefficient that yields the same result is Spearman's non-parametric correlation coefficient (r_S, Equation 19.4):

$$r_S = \frac{Cov_{RA,RB}}{SD_{RA} * SD_{RB}}, \tag{19.4}$$

where $Cov_{RA,RB}$ = covariance of the ranges of variables A and B, SD_{RA} = standard deviation of the range of A, and SD_{RB} = standard deviation of the range of B. r_S denotes the general degree of association (which involves several models, including linear ones) between the two variables.

The eigenvector of an EOF can be defined as the coefficients of a polynomial, remembering that an EOF can be expressed as a first-degree polynomial of order j. Therefore the EOF can be written as:

$$EOF_1 = (b_{1A} * A) + (b_{1B} * B) + (b_{1C} * C) + \cdots \cdots + (b_{1j} * j) \tag{19.5}$$

$$EOF_2 = (b_{2A} * A) + (b_{2B} * B) + (b_{2C} * C) + \cdots \cdots + (b_{2j} * j) \tag{19.6}$$

$$\cdots$$

$$EOF_j = (b_{jA} * A) + (b_{jB} * B) + (b_{jC} * C) + \cdots \cdots + (b_{jj} * j) \tag{19.7}$$

where the data set b_1 constitutes *eigenvector*$_1$ and b_2 constitutes *eigenvector*$_2$. Developing each polynomial, we obtain a number of components equal to the initial number of variables (j), each having the same n as the original variables. Calculating eigenvalues and eigenvectors is not an easy task. If λ is an eigenvalue of a symmetric square matrix (A), then a vector (v) exists such that:

$$Av = \lambda v \qquad Av - \lambda v = 0$$

The terms λv and $\lambda I v$ are given to facilitate the factoring of v. I is the matrix of A:

$$Av - \lambda I v = 0 \qquad (A = \lambda I)v = 0.$$

These equations have a non-trivial solution (eigenvector v exists) if $det(\mathrm{A}-\lambda I) = 0$. Since the determinant of a matrix is not affected by multiplying it by a non-null scalar, we can write $det(\lambda I - \mathrm{A})$=0. To calculate the eigenvalues of A, we can construct a characteristic polynomial:

$$\Delta_\lambda = det(\lambda I - A),$$

and find its roots. Each root of Δ_λ will be an eigenvalue of A. The calculation of eigenvectors is based on the eigenvalues and implies the arbitrary solution of j equations (polynomials), each with j unknowns (coefficients).

EOF_1 is the best numerical combination of the variables of interest and explains the greatest percentage of eigenspace variability. EOF_2 is the second best numerical combination and so on, until EOF_j is the combination that explains the least percentage of variability. To determine if the EOF is statistically significant, we can set the selection rule that all components with eigenvaluesless than 1.0 are not significant (Santamaría-del-Ángel *et al.* , 1994), and therefore the EOF associated with them are not valid because the percentage of variability explained is infinitesimal.

Considering EOF_1 as the best representation of the variables of interest, we shall focus on it to determine the DBGR. SEOF_1 (Equation 19.8 below) is the result of applying Equation 19.5 (EOF_1) with the variables normalized by Z (Equation 19.2).

$$\mathrm{SEOF}_1 = (b_{1Z_A} * Z_A) + (b_{1Z_B} * Z_B) +)b_{1Z_C} * Z_C) + \cdots \quad \cdots + (b_{1Z_j} * Z_j) \tag{19.8}$$

The values of SEOF_1 are easily interpreted. A value of zero represents the point where all the variables have their average value; a value of 1 indicates that the

variables are 1 SD above the mean; -1 implies 1 SD below the mean, and so on. It was observed that the maximum values of $SEOF_1$ (± 3.5 to 4) were found in areas with low variability. As the variability increases, the $SEOF_1$ values can be reduced to ± 1 to 2.

19.1.1 Demonstration section

Part of the California Current System (CCS) that covers the Southern California Bight, encompassing the CalCoFi area (http://www.calcofi.org) was selected for this study (Figure 19.1). To visualize the dynamic nature of DBGR, images from 12-13 April 2006 were selected (Figures 19.2 and 19.3). Daily SeaWiFS and MODIS-Aqua LAC images (1-km per pixel resolution) were used with equidistant cylindrical projections, generating the products Chl-*a*, nLw_{412} and nLw_{488} (MODIS-Aqua) or nLw_{490} (SeaWiFS, Table 19.4) using standard algorithms.

Table 19.4 Identification of the spectroradiometric products (Chl *a*, nLw_{412}, nLw_{488} , nLw_{490}) according to sensor and day of scan.

Day	April 12 2006		April 13 2006	
Sensor	SeaWiFS	MODIS-Aqua	SeaWiFS	MODIS-Aqua
Chl *a*	S2006102203905.L3_LAC.chla.hdf	A2006102204000.L3_LAC.chla.hdf	S2006103211916.L3_LAC.chla.hdf	A2006103212500.L3_LAC.chla.hdf
nLw_{412}	S2006102203905.L3_LAC.412.hdf	A2006102204000.L3_LAC.412.hdf	S2006103211916.L3_LAC.412.hdf	A2006103212500.L3_LAC.412.hdf
nLw_{490} nLw_{488}	S2006102203905.L3_LAC.490.hdf	A2006102204000.L3_LAC.488.hdf	S2006103211916.L3_LAC.490.hdf	A2006103212500.L3_LAC.488.hdf

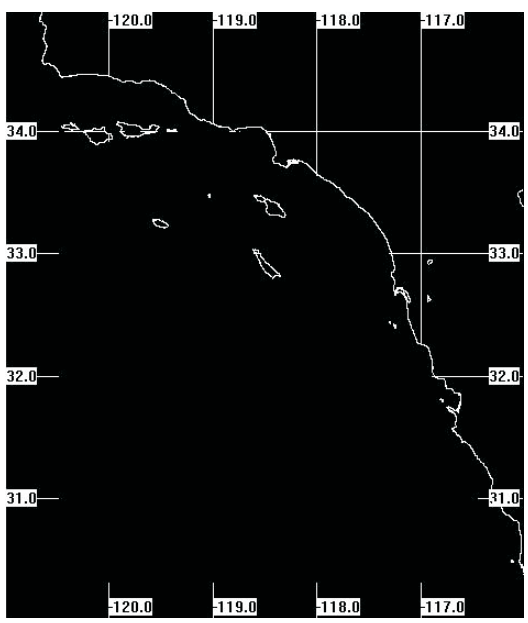

Figure 19.1 Area of study covering part of the CCS that covers the Southern California Bight (SCB) encompassing the CalCoFi area.

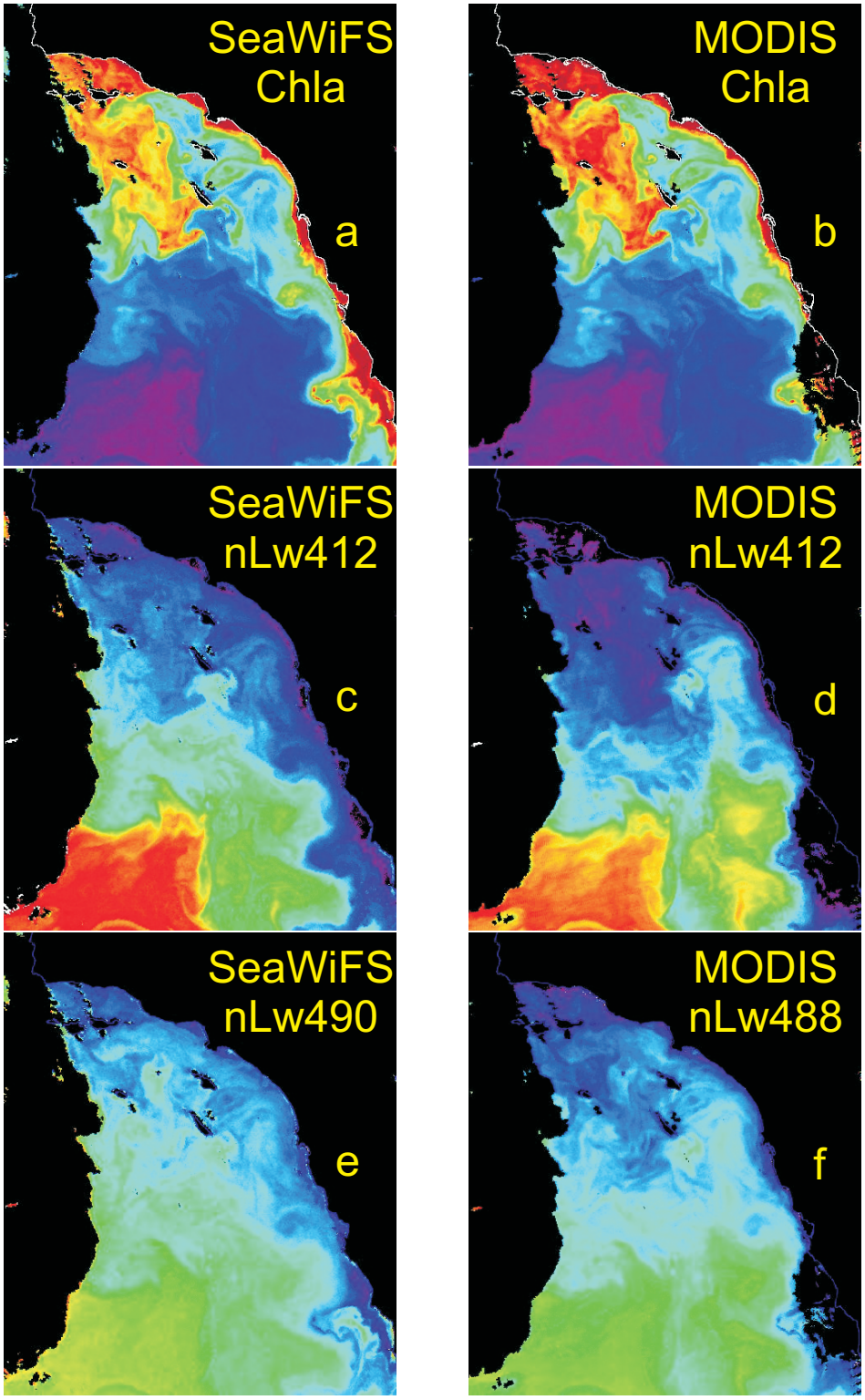

Figure 19.2 Daily LAC images (1-km pixels) from 12 April 2006 for SeaWiFS showing Chl-*a* (a); nLw_{412} (c); and nLw_{490} (e) and for MODIS-Aqua showing Chl-*a* (b); nLw_{412} (d); and nLw_{488} (f).

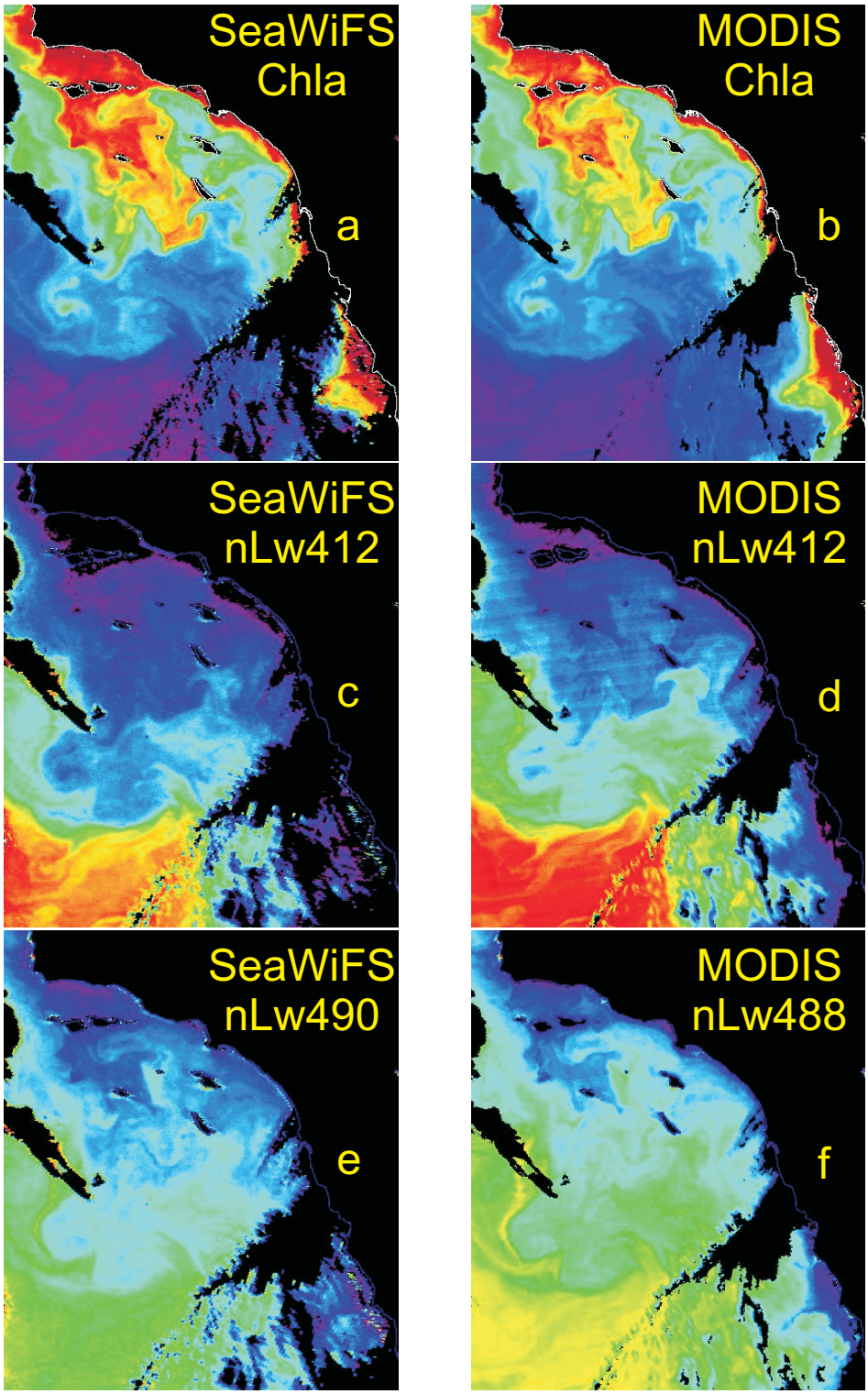

Figure 19.3 Daily LAC images (1-km pixels) from 12 April 2006 for SeaWiFS showing Chl-*a* (a); nLw_{412} (c); and nLw_{490} (e) and for MODIS-Aqua showing Chl-*a* (b); nLw_{412} (d); and nLw_{488} (f).

In this case study, DBGRs were determined based on the first SEOF calculated from nLw_{412}, which was specifically selected because it is a measure of the signal of yellow substances frequently associated with the coast; its current importance has been described in IOCCG (2000; 2004; 2006; 2009b). nLw_{488} (or 490 for SeaWiFS) was also used because it is the closest wavelength to 500 nm, where all the components in the water column absorb the least, such that it can be used to classify coastal and oceanic waters (IOCCG, 2000; 2006). These bio-optical concepts were investigated by Conversi and McGowan (1992) in waters off the coast of San Diego, California.

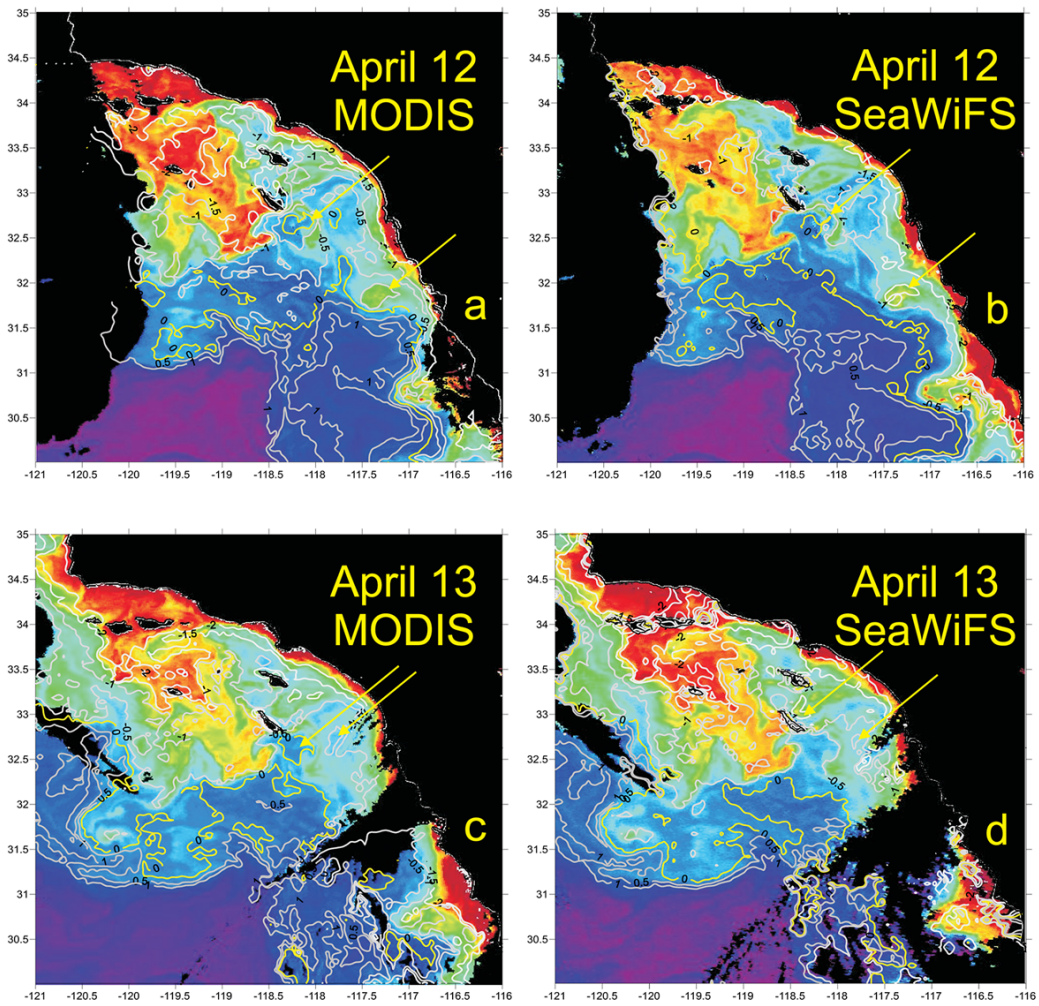

Figure 19.4 Daily LAC images (1-km pixels) of Chl-*a* for 12 and 13 April 2006 obtained from MODIS-Aqua (a, c) and SeaWiFS (b, d). The corresponding $SEOF_1$ derived from the combinations nLw_{412} vs. nLw_{488} or nLw_{490} are superimposed on the LAC images. The yellow arrows indicate the mesotrophic or eutrophic pools derived from the displacement of coastal whirlpools.

Finally, $SEOF_1$ was calculated for each combination of nLw for each day and

sensor (Equation 19.8). The detailed procedure for calculating $SEOF_1$ and deriving DBGR is described in the following section. The resulting values are graphed and superimposed on the images of Chl-*a* (Figure 19.4). The isolines were found to follow the main oceanographic structures. Previous work using approximations based on SST and Chl-*a* in areas where both variables had a strong gradient (González-Silvera et al., 2006; Flores-de-Santiago et al., 2007) also observed these isoline-defined areas with low temperature (negative spatial anomalies) and high Chl-*a* concentrations (positive spatial anomalies). These observations occurred because there is an inverse relationship between Chl-*a* and SST.

nLw_{412} and nLw_{488} (or nLw_{490}) are directly related, causing an isoline of zero in the $SEOF_1$, which divides areas with positive normalized spatial anomalies from areas with negative normalized anomalies. Since the relationship between the variables is positive, and the coefficients described in $eigenvector_1$ are positive (0.707 for all the calculations of $SEOF_1$ for every combination per day per sensor), the values of ± 1 in the $SEOF_1$ imply that the variables included in the model operate at $\pm$ 1 standard deviation values.

Interpolations of the $SEOF_1$ values (Figure 19.4) describe the main oceanographic structures at meso-scale and imply that there is a direct relationship between the dynamic nature of the medium (e.g., winds, surges, and currents) and the number of structures. In fact, it can be seen that closer to the coastline, the isolines describe a greater number of structures. Some structures are well-defined, small areas (Figure 19.4 yellow arrows), which can be the result of small coastal eddies (tens to hundreds of kilometers in size). These eddies are associated with the coastline morphology and mixing processes; these structures can move outside the coastal area as mesotrophic or eutrophic pools. According to Strub et al. (1997) they can be found up to 500 km from the shore.

Three isolines can be highlighted in general: those with values of -2, -1, and 0 (Figures 19.4 and 19.5). A value of 0 indicates the separation of oligotrophic waters from mesotrophic and eutrophic waters. A value of -2 denotes coastal influence, and a value of -1 represents the Ensenada Front. With the exception of the -2 value, the values of 0 and -1 are very similar for MODIS and SeaWiFS data (Figure 19.5). The -2 value is more stable for the MODIS data and appears as a band that fluctuates with the coastline, whereas SeaWiFS data represent this coastal region as a collection of several pools that are near each other, but well defined. When the pattern of these "-2" pools is observed collectively, a distribution similar to that of MODIS can be seen. These observations are due to the differences between the values of the statistical descriptors (Table 19.5) that present each of the values involved in the construction of the $SEOF_1$. It should be noted that the 12 April SeaWiFS nLw have higher mean values and distributions than those obtained with MODIS. Interestingly, this pattern is reversed for the data collected on 13 April.

The Ensenada Front described by $SEOF_1 = -1$ is a very noteworthy oceanographic structure in the south California Current System and divides the area of study

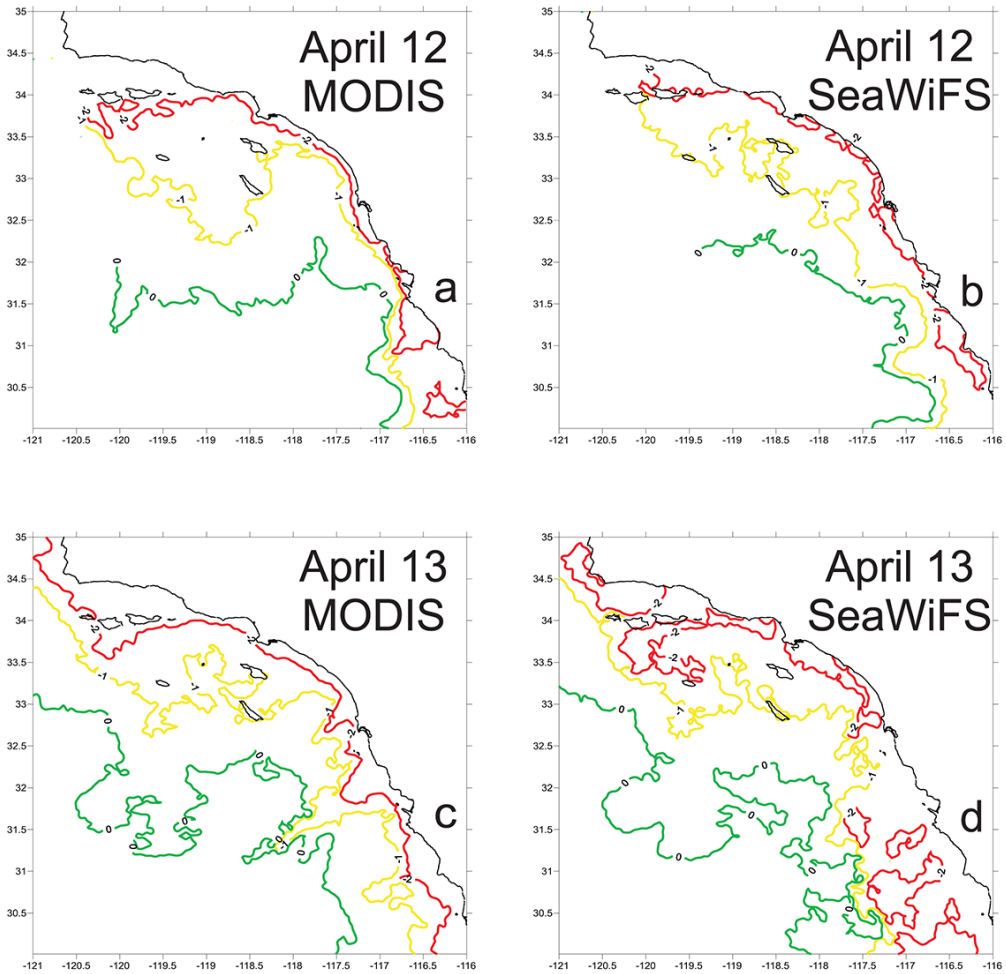

Figure 19.5 Isoline scheme of the $SEOF_1$ = -2 (red), -1 (yellow) and 0 (green) for 12 and 13 April 2006 for MODIS-Aqua (a,c) and SeaWiFS (b,d).

latitudinal. The southern portion has low pigment concentrations and moderate structure while the northern portion has higher concentrations and better-defined patterns (Peláaez and McGowan, 1986; Thomas and Strub 1990; Kahru and Mitchell, 2000; Santamaría-del-Ángel et al., 2002). Comparing the coastal influence region (-2), the Ensenada Front (-1), and the division between oligotrophic waters from mesotrophic-eutrophic waters (0) (Figures 19.5 and 19.6), the daily changes in shape and position of each isoline can be seen. These changes represent the fluid nature of the medium and demonstrate the dynamic concept of this technique of bioregionalization.

The same approximation can be applied to weekly or monthly composites, with smoother data, thus eliminating outlying values and minimizing the small pools caused by whirlpools, but reinforcing and stabilizing the description of the coastal

Table 19.5 Basic statistical descriptors of the marine pixels (without clouds) for nLw for each sensor and day of scan. n is the number of valid data points; $\overline{x}$ is the mean value; SD is the standard deviation of the data.

	April 12 2006				April 13 2006			
	SeaWiFS		MODIS-Aqua		SeaWiFS		MODIS-Aqua	
	nLw_{412}	nLw_{490}	nLw_{412}	nLw_{488}	nLw_{412}	nLw_{490}	nLw_{412}	nLw_{488}
n	136928	136928	127574	127574	142242	142242	157671	157671
$\overline{x}$	0.83257	0.81007	0.77688	0.77725	0.68933	0.76958	0.83115	0.91187
SD	0.45187	0.21130	0.43608	0.20628	0.44125	0.20194	0.46731	0.22176
Minimum	0.00100	0.02600	0.00100	0.10700	0.00100	0.06100	0.00100	0.23200
Maximum	6.37200	5.82800	2.81300	2.28200	2.26300	2.17400	1.99000	3.29000

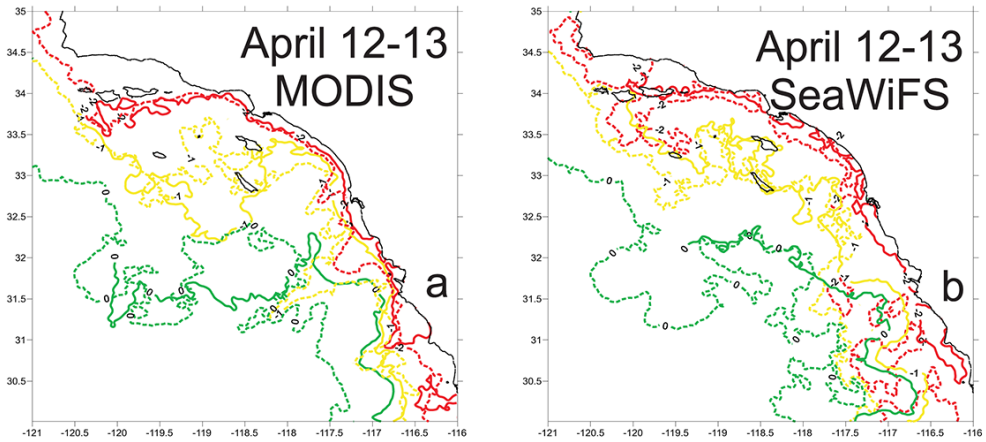

Figure 19.6 Isoline scheme of the $SEOF_1$ = -2 (red), -1 (yellow) and 0 (green) for 12 April (solid lines) and 13 April (dashed lines) 2006 for (a) MODIS-Aqua and (b) SeaWiFS.

domain defined by a value of -2. It should be noted that the selection of variables used to calculate SEOF to determine DBGR is very important; they must feature a wide spatial gradient to enable greater precision in defining the mean value and more robust calculation of the normalized spatial anomalies. Chl-*a* and SST are especially well suited to areas of surge, oceanographic fronts, and enclosed seas, while nLw is well suited for oceanic or oligotrophic areas (although applications based on nLw can also be applied to very productive areas, such as the California Current System presented in this study. Currents are known to play an important role in the formation, evolution and changes of oceanographic structures through time, causing the CCS to be highly dynamic and complex, as manifested in the meanders, whirlpools and filaments that constantly overpower one another.

Haury et al. (1993) defined the Ensenada Front as a recurring border that divides warm oligotrophic waters from cold eutrophic waters, with a gradient ranging from a temperature of 16°C to the north and 19°C to the south. Santamaría-del-Ángel et al. (2002) analyzed CZCS images and defined the Ensenada Front as a recurring structure

shaped like an "M", with an oceanographic signature defined by the isobar of 0.25 mg pigment m^{-3}. Among the peaks, there is a valley located between 119° – 121°W that is associated with Concepción Point, California. In southern Baja California, Concepción Point is a true inflection point in the coastal morphology that promotes surge events (Dugdale et al. 1997) of cold waters, which are rich in nutrients and low in Chl-*a* when recently upwelled. Subsequently, the cold waters are displaced as plumes or filaments, with higher Chl-*a* concentrations and temperatures, and lower nutrient concentrations (Chereskin et al. 2000). Batteen (1997) and Cornuelle et al. (2000) described the cyclonic Southern California Eddy (SCE) to the south of the Concepción Point. This eddy is cyclonic and is thought to originate as a result of interactions between the California Countercurrent, the wind and the coastline. The SCE transports water from the surges to the south. The "M" shape of the Ensenada Front, and especially its valley, is the result of these processes combining with the oligotrophic waters that come from the south.

The changes in the location and shape of each isobar can be monitored from day to day by comparing the coastal influence (with a value of -2), the Ensenada Front (with a value of -1) and the separation of oligotrophic waters from the mesotrophic-eutrophic waters (with a value of 0; Figures 19.5 and 19.6). These changes represent the dynamic nature of the medium and represent the dynamic concept in this technique of bioregionalization. The Ensenada Front displays a latitudinal displacement that varies on a seasonal and interannual scale (Haury et al. 1993). During the cold months in the off-season of an El-Niño year, the Ensenada Front is located to the south of Ensenada Baja California. During the summer, the Ensenada Front can move to the north of San Diego, California, mainly due to the introduction of more oligotrophic waters. In the following months, this influx of oligotrophic waters wanes, causing the Ensenada Front to move to the south again. This fluctuation generally occurs from 30 – 31°N up to 32°30′N.

19.2 Training

Using the program WIM (Windows Image Manager http://www.wimsoft.com), each hdf nLw image was transformed into latitude, longitude, and geophysical value files in txt format with the extension ".dat". These files were arranged into two ".csv" files (2006102DATABASE.csv and 2006103DATABASE.csv, April 12 and 13, 2006, respectively) which can be downloaded from the IOCCG website at: http://www.ioccg.org/handbook/biogeo/ and which can be opened with Microsoft Excel. The ".csv" files have six columns, each with 262,432 rows that represent all of the image pixels in addition to the heading of each column. All the images must have the same resolution so that the pixels can be easily geo-differentiated. Equidistal cylindric projection allows the mathematical construction of the latitude-longitude-geophysical value matrices with greater ease. This is because the degrees of latitude

Table 19.6 Pearson's correlation matrices of the original variables. April 12, 2006: M_4, MODIS-Aqua; M_5, SeaWiFS. April 13, 2006: M_6, MODIS-Aqua; M_7, SeaWiFS.

	MODIS	nLw_{412}	nLw_{488}		SeaWiFS	nLw_{412}	nLw_{490}
M_4 =	nLw_{412}	1	0.95618	M_5 =	nLw412	1	0.94365
	nLw_{488}	0.95618	1		nLw_{490}	0.94365	1
	MODIS	nLw_{412}	nLw_{488}		SeaWiFS	nLw_{412}	nLw_{490}
M_6 =	nLw_{412}	1	0.94600	M_7 =	nLw_{412}	1	0.91790
	nLw_{488}	0.94600	1		nLw_{490}	0.91790	1

and longitude always have the same distance whether at the equator or at the poles. The six columns are latitude (A); longitude (B); MODIS nLw_{412} (C); MOIDS nLw_{488} (D); SeaWiFS nLw_{412} (E); and SeaWiFS nLw_{490} (F). Not all data-handling software can process such a large amount of data, e.g. Microsoft Excel 2003 can only hold 60,000 rows of data (Microsoft Excel 2007 SP1 can analyze a greater volume of data). You will note several rows of zeroes in columns C to F, indicating *NaN* (Not a Number) values, which are the result sensor saturation (due to land or clouds). The next step is to eliminate *NaN* values (land and clouds), leaving only marine data with nLw values. The filtered data is presented in the files "2006102withoutnan.csv" and "2006103withoutnan.csv", and the basic statistical descriptors of each variable per sensor are presented on Table 19.5. The data from each sensor contain a different number of valid points for the same day because *NaN* values are dependent on conditions, such as cloud coverage and sea breeze, that change in a matter of hours, and each sensor scans at a different time of day (SeaWiFS at noon and MODIS-Aqua at 1:30 PM).

The next step is to construct the symmetric square matrix of Pearson's linear correlations (Equation 19.3) with the original data (M_4, M_5, M_6, M_7, Table 19.6). The original values are normalized to Z (Equation 19.2) using the basic descriptors of each variable (Table 19.5) (the normalized variables can be found in the files "Z2006102withoutnan.csv" and "Z2006103withoutnan.csv") with which the *Cov* matrix can be calculated using Equation 19.1 (M_8, M_9, M_{10},M_{11}, Table 19.7). At this point, the Pearson's correlation matrices of the original variables (Table 19.6) are identical to the matrices of the normalized variables (Table 19.7).

We shall use matrix M_4 (which is equal to matrix M_8) as an example of the calculation of $eigenvector_1$ that has the coefficients that consider the normalized variables of Equation 19.8 used to calculate $SEOF_1$. The orthogonal solution of this type of matrix can have many approximations, depending on the number of variables. For example, seven variables would imply solving seven simultaneous equations, each with seven unknowns (eigenvector coefficients). To show one of the many approximations available for the calculation of eigenvectors and eigenvalues,

Table 19.7 *Cov* matrices normalized to Z. April 12, 2006: M_8, MODIS-Aqua; M_9, SeaWiFS. April 13, 2006: M_{10}, MODIS-Aqua; M_{11}, SeaWiFS.

MODIS	$ZnLw_{412}$	$ZnLw_{488}$	SeaWiFS	$ZnLw_{412}$	$ZnLw_{490}$
$M_8 = ZnLw_{412}$	1	0.95618	$M_9 = ZnLw_{412}$	1	0.94365
$ZnLw_{488}$	0.95618	1	$ZnLw_{490}$	0.94365	1
MODIS	$ZnLw_{412}$	$ZnLw_{488}$	SeaWiFS	$ZnLw_{412}$	$ZnLw_{490}$
$M_{10} = ZnLw_{412}$	1	0.94600	$M_{11} = ZnLw_{412}$	1	0.91790
$ZnLw_{488}$	0.94600	1	$ZnLw_{490}$	0.91790	1

the simplest symmetric square matrix (a matrix of order $2 - 2 \times 2$, as derived in this study) is used. If M_4 =A;

$$A + \begin{bmatrix} 1 & 0.956182 \\ 0.95618 & 1 \end{bmatrix}$$

Since the determinant of a matrix is not affected when the matrix is multiplied by a non-null scalar, we can write $\det(\lambda I - A) = 0$ (see Background Information). To calculate the eigenvalues of A, we can construct a characteristic polynomial:

$$\Delta_\lambda = \det(\lambda I - A)$$

The matrix $B=(\lambda I - A)$ is constructed;

$$B = (\lambda I - A) = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0.9561822 \\ 0.9561822 & 1 \end{bmatrix} = \begin{bmatrix} (\lambda - 1) & 0.956182 \\ 0.9561822 & (\lambda - 1) \end{bmatrix} = 0$$

and its determinant is calculated. With lower order matrices (2nd or 3rd order), the determinant is calculated with simple rules derived from Laplace's theorem. The determinants of a second order matrix are calculated with the formula:

$$B = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}$$
$$\det(B) = [(x)]_{11} * x_{22}) - (x_{12} * x_{21})$$

that has a solution in the form of the characteristic polynomial if it is developed.

$$\Delta_\lambda = \det(B) = (\lambda - 1)^2 - 0.836 = (\lambda^2 + 2\lambda) - 0.836 = 0$$

The solutions of characteristic polynomial Δ_λ would be the eigenvalues $\lambda_1 = 1.9562$ and $\lambda_2 = 0.0438$. The vectors associated with $\lambda_1 = 1.9562$ must satisfy:

$$Av_1 = \lambda_1 v_1$$

$$\begin{bmatrix} 1 & 0.95618 \\ 0.95618 & 1 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix} = 1.9562 \begin{bmatrix} v_{11} \\ v_{21} \end{bmatrix}$$

$$\begin{bmatrix} v_{11} + 0.95618v_{21} \\ 0.95618v_{11} + v_{21} \end{bmatrix} = \begin{bmatrix} 1.9562v_{11} \\ 1.0962v_{21} \end{bmatrix}$$

An equation system with infinite solutions is then created, and in this case it is composed of two simultaneous equations, each with two unknowns:

$$\begin{cases} v_{11} + 0.95618v_{21} = 1.9562v_{11} \\ 0.95618v_{11} + v_{21} = 1.9562v_{21} \end{cases}$$

The solution of these equations is not trivial, since many possible solutions can be found. Furthermore, many different values would have to be tested. The most popular method to obtain a vector associated with $\lambda_1 = 1.9562$ is to randomly assign a value to v_{11} or v_{21} and test them to see if they fit the solution. This is why, although the theory for the development of EOF was described a long time ago, it did not become popular until the arrival of FORTRAN, which allowed the development of software that could run several approximations interactively, finding the best solutions for this simultaneous equation system. Currently, software programs like MINITAB (http://www.minitab.com), MATLAB (http://www.mathworks.com) and/or R (http://www.r-project.org), which handle matrices, eigenvalue, and eigenvector calculations to determine the eigenspaces related to the collection of combined variables in Pearson's correlation matrix or the *Cov* matrix (for normalized variables), speed up and increase the precision of the calculation process. In this case study, *eigenvector*$_1$ (v_1) is defined by:

$$v_1 = \begin{bmatrix} 0.707 \\ 0.707 \end{bmatrix}$$

SEOF_1 will be calculated using the general form of Equation 19.8 and the normalized data (Z2006102withoutnan.csv):

$$\text{SEOF}_1 = (0.707 * \text{ZnLw}_{412}) + (0.707 * \text{ZnLw}_{488}) \tag{19.9}$$

SEOF_1 has the same number of observations as the combined variables (127,574 in this case). When determining DBGR, the values of SEOF_1 can be interpreted as normalized spatial anomalies that can be used to generate isoline maps that will determine different areas with similar characteristics to the variables being combined (Figures 19.4, 19.5 and 19.6).

19.3 Questions

Q1: What is the Ensenada Front?

Q2: Why is it called Ensenada Front?

Q3: What is its oceanographic importance in the CCS?

Q4: How can it be detected?

Q5: How can the coastal influence within the study area be determined?

Q6: Why did EOF only become popular recently when the theory was developed a long time ago?

Q7: Why are there small differences in DBGR determined with SEOF-based approximations using daily LAC data obtained with MODIS-Aqua and SeaWiFS?

Q8: When and where can we expect DBGR to be identical using data from MODIS or SeaWiFS or another sensor, such as MERIS?

Q9: Which areas are inadequate for the determination of DBGR from the SEOF obtained from the observations of Chl-*a* and SST?

Q10: What is the meaning of $\mathrm{SEOF}_1 = 0$?

Q11: Why is it important to record all images at the same resolution?

19.4 Answers

A1: The Ensenada Front (EF) is one of the most obvious oceanographic structures that divide the area latitudinally into two parts; the southern part, which has low pigment concentrations with a moderate structure, and the northern part, which has higher pigment concentrations and more well-defined patterns. The EF is a recurring structure that is shaped like an "M".

A2: Haury et al. (1993) named the EF based on the location of the front near Ensenada, Baja California in the CZCS image where it was first described by Peláez and McGowan (1986).

A3: First, the EF clearly divides the study area into two parts: an oligotrophic region to the south and a mesotrophic/eutrophic region. The position of the EF is also influenced by events like El Niño, and the EF is more prominent to the east than to the north in the off-season of an El Niño year. The EF is formed as a result of the combined effects of the physical barriers (Concepción Point), the mixing processes (surges) and the variability of the California Countercurrent.

A4: In this study, the $SEOF_1 = -1$ condition can be detected, but previous studies based on ocean colour images detected the EF based on the isobar of 0.25 mg pigment m^{-3}.

A5: The coastal influence can be detected with SEOF = −2. Superimposing the SEOF onto Chl-*a* maps shows that the SEOF determines a band parallel to the coast and, on certain occasions, defines small mesotrophic pools that displace themselves to the south of the EF.

A6: EOF implies the solution of a system of n simultaneous equations, each with n unknowns. The solution of these equations is not trivial because this equation system has an infinite number of solutions. To find some of the possible solutions, many values would have to be tested simultaneously. Therefore, although the theory for the development of EOF was described a long time ago, its use did not become popular until the arrival of FORTRAN, which allowed the development of software programs that can interactively test several approximations, finding the best solution to this system of simultaneous equations.

A7: The differences are due to the differences in the statistical descriptor values that are related to each of the variables involved in the construction of $SEOF_1$. Different mean values and standard deviations yield transforms with different geolocalized pixels with values of 0 and ± 1 or 2.

A8: This can be expected if the analysis is performed with weekly, monthly, seasonal, or average-climate compositions because the approximations based on these compositions smooth the data, removing outlying values (high noise). This would result in a data dispersion calculation (standard deviation) that is more consistent, which would translate to better-defined regions. It should be noted that these differences are prevalent in areas with high Chl-*a* concentrations, so that temporal composition processes would smooth out the outlying values that are commonly found by the algorithms used in areas with high Chl-*a* concentrations.

A9: In order to determine DBGR, it is very important that the variables used for the construction of SEOF have a large number of data, with a large gradient, to define the mean value more precisely and thus be able to calculate more robust normalized spatial anomalies. For example, if Chl-*a* and SST are selected as the variables to be combined, it must be ensured that both variables have a strong gradient (such as surge areas, oceanographic fronts and enclosed seas). In oligotrophic oceanic or coastal areas (e.g. equatorial), variables must have a considerable gradient to calculate robust descriptive factors that will consolidate the normalized spatial anomalies. The spectroradiometric products of nLw can be used to carry out this approximation, whether in pairs (as in this study) or using all the nLw available.

A10: The isolines of $SEOF_1$ values are generally related to the combinations of

normalized anomalies. A value of zero (0) indicates that the variables combined with the SEOF are operating at their mean values. This is the isoline that shows the most striking contrasts. González-Silvera et al. (2006) and Flores-de-Santiago et al. (2007) found that this isoline separated areas with low temperature (negative spatial anomalies) from those areas with a high Chl-*a* concentration (positive spatial anomalies) due to the inverse relationship between Chl-*a* and SST. In this case, the relationship between nLw_{412} and nLw_{488} (or nLw_{490}) is direct, such that the zero line of $SEOF_1$ divides areas with positive normalized anomalies from the areas where negative normalized anomalies are found.

A11: It is very important that all images have the same resolution because it makes it easier to create a software program that extracts data based on latitude-longitude-geophysical value matrices. If the images had different resolutions, the matrices would each have different dimensions, which would cause the introduction of another variable into the data extraction process. Equidistant cylindrical projection allows the mathematical construction of the latitude-longitude-geophysical value matrices with the greatest ease because the degrees of latitude and longitude have the same distance at both the equator and the poles.

19.5 References

Batteen ML (1997) Wind-forced modeling studies of currents, meanders, and eddies in the California Current system. J Geophys Res 102: 985-1010

Conversi A, McGowan JA (1992) Variability of Water Column Transparency, Volume Flow and Suspended Solids Near San Diego Sewage Outfall (California): 15 Years of Data. Chem Ecol 1029-0370, 6(1):133-147

Cornuelle B, Chereskin TK, Niiler PP, Morris MY (2000) Observations and modeling of a California undercurrent eddy. J Geophys Res 105:1227-1243. Spatial and temporal characteristics of the mesoscale circulation of the California Current from eddy-resolving moored and shipboard measurements. J Geophys Res 105: 1245-1269

Dugdale RC, Davis CO, Wilkerson FP (1997) Assessment of new production at the upwelling center at Punta Conception, California, using nitrate estimated from remotely sensed sea surface temperature. J Geophy Res 102:8573-8585

Flores-de-Santiago FJ, Santamaría-del-Ángel E, González-Silvera A, Martínez-Díaz-de-León A, Millán-Núñez R, Kovacs J (2007) Assessing dynamics micro-regions in the Great Islands of the Gulf of California based on MODIS aqua imagery products. In: Frouin RJ, Lee ZP (eds) Coastal Ocean Remote Sensing, Proc SPIE 6680:668010-1 doi: 10.1117/12.732574

González-Silvera A, Santamaría-del-Ángel E, García VMT, García CAE, Millán-Núñez R, Muller-Karger F (2004) Biogeographical regions of the tropical and subtropical Atlantic Ocean off South America: classification based on pigment (CZCS and chlorophyll-a (SeaWiFS) variability. Cont Shelf Res 24, 983-1000

González-Silvera A, Santamaría-del-Ángel E, Millán-Núñez R (2006) Spatial and temporal variability of the Brazil-Malvinas confluence and the La Plata Pluma as seen by SeaWiFS and AVHRR imagery. J Geophys Res 111:C06010

Haury LR, Venrick E, Fey CL, McGowan JA, Niler PP (1993) The Ensenada Front: July 1985. CalCOFI Rep 34:69-88

Hickey BM (1998) Coastal oceanography of western North America from the tip of Baja California to Vancouver Island: Coastal Segment. In: Robinson AR and Brink KH (eds), The Sea, Vol 11. John Wiley and Son, 345-391

Hope K (1982) Manual práctico de estadística avanzada. Ed. Trillas. México D.F., 111 pp

IOCCG (2000) Remote sensing of ocean colour in coastal, and other optically-complex, waters. In: Sathyendranath S (ed) Reports of the International Ocean-Colour Coordinating Group, No 3, IOCCG, Dartmouth, Canada

IOCCG (2004) Guide to the Creation and Use of Ocean-Colour, Level-3, Binned Data Products. Antoine D (ed) Reports of the International Ocean-Colour Coordinating Group, No 4, IOCCG, Dartmouth, Canada

IOCCG (2006) Remote Sensing of Inherent Optical Properties: Fundamentals, Tests of Algorithms, and Applications. Lee Z-P (ed) Reports of the International Ocean-Colour Coordinating Group, No 5, IOCCG, Dartmouth, Canada

IOCCG (2008) Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology. Platt T, Hoepffner N, Stuart V, Brown C (eds) Reports of the International Ocean-Colour Coordinating Group, No 7, IOCCG, Dartmouth, Canada

IOCCG (2009a) Remote Sensing in Fisheries and Aquaculture. Forget M-H, Stuart V, Platt T (eds) Reports of the International Ocean-Colour Coordinating Group, No 8, IOCCG, Dartmouth, Canada

IOCCG (2009b) Partition of the Ocean into Ecological Provinces: Role of Ocean-Colour Radiometry. Dowell M, Platt T (eds) Reports of the International Ocean-Colour Coordinating Group, No 9, IOCCG, Dartmouth, Canada

Kahru M, Mitchell G (2000) Influence of the 1997-1998 El Niño on the surface chlorophyll in the California Current. Geophy Res Lett 27:2937-2940

Kahru M, Mitchell G (2001) Seasonal and nonseasonal variability of satellite-derivated chlorophyll and colored dissolved organic matter concentrations in California Current. J Geophy Res 106:2517-2529

Lluch-Cota SE (2000) Coastal upwelling in the Eastern Gulf of California. Oceanol Acta 23 (6):731-740

Marchesiello P, McWillians JC, Shchepetkin A (2003) Equilibrium Structure and Dynamics of the California Current System. J Phys Oceanogr 33: 753-783

Millán-Núñez R, Alvarez-Borrego S, Trees CC (1997) Modeling the vertical distribution of chlorophyll in the California Current System. J Geophys 102(C4):8587-8595

Pelaez J, McGowan JA (1986) Phytoplankton pigment patterns in the California Current as determined by satellite. Limnol Oceanogr 31(5):927-950

Santamaría-del-Ángel E, Álvarez-Borrego S, Müller-Karger FE (1994) Gulf of California bio-geographic regions based on coastal zone color scanner imagery. J Geophys Res 99 (C4): 7411-7421

Santamaría-del-Ángel E, Millán-Núñez R, González-Silvera AG, Muller-Karger FE (2002) The Color Signature of the Ensenada Front and its seasonal and interannual variability. CaCoFi Rep 43; 156-161

Strub PT, Chereskin TK, Niiler PP, James C, Levine MD (1997) Altimeter derived variability of surface velocities in the California Current System 1. Evaluation of TOPEX altimeter velocity resolution, J Geophys Res 102(C6), 12727-12748

Thomas AC, Strub PT (1990) Seasonal and interannual variability of pigment concentration across a California Current frontal zone. J Geophy Res 95:13023-13042

Traganza ED, Nestor DA, McDonald AK (1980) Satellite observations of nutrients upwelling off the coastal of California. J Geophy Res 85:4101-4106

19.5.1 Further Reading

Devred E, Sathyendranath S, Platt T (2007) Delineation of ecological provinces using ocean colour radiometry. Mar Ecol Prog Ser 346:1-13

Djavidnia S, Mélin F, Hoepffner N (2010) Comparison of global ocean colour data records. Ocean Sci 6:61-76 http://www.ocean-sci.net/6/61/2010/os-6-61-2010.html

Esaias EW, Iverson RL, Turpie K (2000) Ocean provincia classification using ocean color data: observing biological signatures of variations in physical dynamics. Global Change Biol 6: 39-55

Hardman-Mountford NJ, Hirata T, Richardson KA, Aiken J (2008) An objective methodology for the classification of ecological pattern into biomes and provinces for the pelagic ocean. Remote Sens Environ 112:3341-3352

Longhurst A, Sathyendranath S, Platt T, Caverhill C (1995) An estimate of global primary production in the ocean from satellite radiometer data. J Plankton Res 17 (6), 1245-1271

Oliver MJ, Irwin AJ (2008) Objective global ocean biogeographic provinces, Geophys Res Lett 35, L15601, doi:10.1029/2008GL034238

Preisendorfer RW (1988) Principal Component Analysis in Meteorology and Oceanography. In: Eslinger DL, O'Brien JJ, Iverson RL (1989) Empirical Orthogonal Function Analysis of Cloud-Containing Coastal Zone Color Scanner Images of Northeastern North American Coastal Waters. J Geophys Res 94(C8): 10884-10890

Saraceno M, Provost C, Piola AR, Bava J, Gagliardini A (2004) Brazil Malvinas Frontal System as seen from 9 years of advanced very high resolution radiometer data. J Geophys Res 109: C05027, doi:10.1029/2003JC002127

Saraceno M, Provost C, Piola AR (2005) On the relationship between satellite-retrieved surface temperature fronts and chlorophyll a in the western South Atlantic. J Geophys Res 110: C11016, doi:10.1029/2004JC002736

Saraceno M, Provost C, Lebbah M (2006) Biophysical regions identification using an artificial neuronal network: A case study in the South Western Atlantic. Adv Space Res 37:793-805