SOURCE CODE FOR
BIOLOGY AND MEDICINE

METHODOLOGY   **Open Access**

Using the value of Lin's concordance correlation coefficient as a criterion for efficient estimation of areas of leaves of eelgrass from noisy digital images

Héctor Echavarría-Heras[1*], Cecilia Leal-Ramírez[1], Enrique Villa-Diharce[2] and Oscar Castillo[3]

Abstract

Background: Eelgrass is a cosmopolitan seagrass species that provides important ecological services in coastal and near-shore environments. Despite its relevance, loss of eelgrass habitats is noted worldwide. Restoration by replanting plays an important role, and accurate measurements of the standing crop and productivity of transplants are important for evaluating restoration of the ecological functions of natural populations. Traditional assessments are destructive, and although they do not harm natural populations, in transplants the destruction of shoots might cause undesirable alterations. Non-destructive assessments of the aforementioned variables are obtained through allometric proxies expressed in terms of measurements of the lengths or areas of leaves. Digital imagery could produce measurements of leaf attributes without the removal of shoots, but sediment attachments, damage infringed by drag forces or humidity contents induce noise-effects, reducing precision. Available techniques for dealing with noise caused by humidity contents on leaves use the concepts of adjacency, vicinity, connectivity and tolerance of similarity between pixels. Selection of an interval of tolerance of similarity for efficient measurements requires extended computational routines with tied statistical inferences making concomitant tasks complicated and time consuming. The present approach proposes a simplified and cost-effective alternative, and also a general tool aimed to deal with any sort of noise modifying eelgrass leaves images. Moreover, this selection criterion relies only on a single statistics; the calculation of the maximum value of the Concordance Correlation Coefficient for reproducibility of observed areas of leaves through proxies obtained from digital images.

Results: Available data reveals that the present method delivers simplified, consistent estimations of areas of eelgrass leaves taken from noisy digital images. Moreover, the proposed procedure is robust because both the optimal interval of tolerance of similarity and the reproducibility of observed leaf areas through digital image surrogates were independent of sample size.

Conclusion: The present method provides simplified, unbiased and non-destructive measurements of eelgrass leaf area. These measurements, in conjunction with allometric methods, can predict the dynamics of eelgrass biomass and leaf growth through indirect techniques, reducing the destructive effect of sampling, fundamental to the evaluation of eelgrass restoration projects thereby contributing to the conservation of this important seagrass species.

Keywords: Eelgrass leaf, Area estimations, Noisy digital images selection criterion, Concordance correlation coefficient

* Correspondence: heheras@icloud.com
[1]Centro de Investigación Científica y de Estudios Superiores de Ensenada, Carretera Ensenada-Tijuana No. 3918, Zona Playitas, Apdo. Postal 360, Código Postal 22860 Ensenada, B C, México
Full list of author information is available at the end of the article

Echavarría-Heras *et al. Source Code for Biology and Medicine* (2014) 9:29

Page 2 of 10

Background

Seagrass meadows are highly productive plant communities that grant valuable ecological services in estuaries and near-shore environments worldwide. Seagrasses provide food and shelter for a myriad of economically and ecologically valued marine organisms [1-3], play an important role in nutrient cycling [4,5], favor the stabilization of the shoreline as roots and rhizomes compact the substrate, preventing erosion [6,7], participate in the foundation of the detrital food web [8], and play also, a fundamental role in carbon sequestration [9]. Eelgrass (*Zostera marina L.*) is particularly relevant not only because it is the dominant seagrass species along the coasts of both the North Pacific and North Atlantic [10], but also, because eelgrass communities have been traditionally recognized as among the richest and most varied in the abundance of sea life [11]. Indeed, this cosmopolitan macrophyte was found to produce up to 64% of the total primary production of an estuarine system [12].

The forcing of *Zostera marina* dynamics by environmental variables is well documented in the literature [13-18]. Light availability, temperature, and dissolved nutrients are the most important variables for explaining the observed variability [18,19]. But even when light and nutrients are not limiting, temperatures ranging above the upper limit tolerated by eelgrass can provoke severe negative effects on its growth [20]. Indeed, the onset of warm ENSO events has been shown to dramatically diminish eelgrass growth [20]. Therefore, the productivity of *Zostera marina* populations could be diminished by global climate change, which is expected to result in warming and rising seas, thereby reducing the availability of both light and nutrients underwater [21]. Another concern for the health of eelgrass populations pertains to increasing deleterious anthropogenic influences. The loss of eelgrass habitat has been noted worldwide, with major losses in the past few decades [22-25]. Within restoration strategies, replanting plays an important role [26-28]. The monitoring of these efforts is fundamental for the evaluation of the effectiveness of restoration of functions and values of natural populations. Accurate measurements of the standing crop and productivity of transplanted populations at a given time constitute an important input for evaluating the restoration of the ecological functions and values of natural populations. Although traditional assessment methods do not cause damage to natural populations, their invasive nature could significantly alter the development of transplanted populations. Echavarria-Heras et al. [29] and Echavarria-Heras et al. [30] propose allometric methods that reduce eelgrass biomass and leaf growth rate estimations to measurements of leaf length or area. Besides, the use of digital imagery could provide leaf area estimations which avoid invasive effects. But in some cases noise effects could lead to misidentification of pixels placed on the peripheral contour of leaves images (see Figure 1). This could spread uncertainty on leaf area estimations that ultimately could render imprecise allometric projections of biomass and leaf growth rates. Therefore, for accurateness we must rely on an image selection method that produces an unambiguous identification of the sequence of pixels that form the peripheral contours of digitalized eelgrass leaves. In order to achieve this task, there are techniques developed on the basis of the concepts of adjacency, vicinity, connectivity and tolerance of similarity between pixels (see Appendix). Using this framework Leal-Ramirez and Echavarria-Heras [31] introduced a direct comparison method aimed to discriminate the interval of tolerance of similarity that produces the most accurate estimations of length, width or area of eelgrass leaves from digital images with noise induced by humidity contents. For a given interval of tolerance of similarity, the process initially identifies the peripheral contour of the images of leaves and then measures the concomitant lengths widths and areas. Next, individual deviations between leaf area measurements taken from images and those obtained directly from leaves are used to produce statistics aimed to obtain the proportions of leaves for which image assessments underestimate or overestimate observed values. The ratio of these proportions defines a selection index whose smallest value provides criterion for choosing the interval of tolerance of similarity that yields the most accurate image related measurements. The implementation of the direct comparison method uses lengthy computational stages that include various statistical inferences on deviations between observed an image obtained leaf areas. In this contribution, we present an alternative criterion for the selection of the named interval of tolerance of similarity. The present procedure called the concordance correlation method; is simpler to implement than the direct comparison method. It only requires calculating the values of the Concordance Correlation Coefficient (CCC) for the reproducibility of observed leaf areas through proxies obtained from corresponding images. The present criterion proposes the use of the interval of tolerance of similarity that yields the maximum value of the aforementioned CCC for consistent digital image estimations of eelgrass leaves areas. Our results show that on spite of its simplicity the present selection criterion yields highly reliable levels of accuracy.

In section two, we present a brief review of the direct comparison method. Section three formally explains the present concordance correlation method. Section four describes the results of this study and discusses the advantages and possible drawbacks of the present approach.

The Direct Comparison Method (DCM)

In this section we briefly describe the steps of the direct comparison method as conceived by Leal-Ramirez and

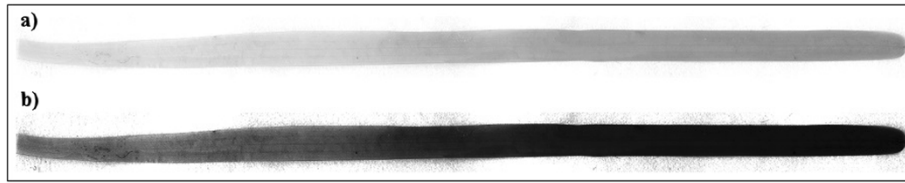Echavarría-Heras *et al. Source Code for Biology and Medicine* (2014) 9:29

Page 3 of 10

Figure 1 A digital image of a *Zostera marina* leaf. a) An image of *a Zostera marina* leaf exhibiting the typical belted shape. Related area is commonly approximated by the product of length and average width. **b)** The display of the image of the leaf using a darker tonality reveals pixels placed beyond the peripheral contours, which do not belong to the image and whose presence is explained by humidity- noise-related effects. Improper identification of the peripheral contour of the leaf image due to spurious entries can lead to miscalculation of related area.

Echavarria-Heras [31]. Initially, the DCM chooses a positive integer n and uses it to fix a tolerance level $q = (l_{max}/n)$, being l_{max} the maximum observed leaf length. This yields a covering for the range $[0, l_{max}]$ by a collection of n disjoint intervals of the form $I_k = [q(k-1), qk)$, with $1 \le k \le n$. Subsequently, for each value of the index k the procedure identifies the group $G_k(l)$ of n_k leaves whose lengths are contained in I_k. An index j such that $1 \le j \le n_k$ labels leaves in $G_k(l)$ while the symbols l_{oj}^k, h_{oj}^k and a_{oj}^k denote respectively the straight length, width and area of the jth leaf in $G_k(l)$. Particularly, estimations a_{oj}^k of the leaf areas in $G_k(l)$ can be obtained by using the length times width proxy [32]. Digital images of leaves in the $G_k(l)$ groups are processed by a specified color format with a number C_{max} of colors and via intervals of tolerance of similarity $ST(r) = [0, r]$, being r, $0 \le r \le C_{max} - 1$, the number of different tonalities used for pixel identification. By keeping $ST(r)$ fixed, a routine selects a starting point within the image of the jth leaf in $G_k(l)$ and detects all adjacent pixels falling within the selected interval of tolerance of similarity $ST(r)$. This task which is achieved using equations (A1), (A2) and (A3) identifies the peripheral contour of the leaf image, and allows the measurements of the concomitant proxies for the length $l_{dj}^k(r)$, width $h_{dj}^k(r)$ and area $a_{dj}^k(r)$ of the leaf. Afterwards the method obtains the deviations for leaf length $e_{lj}^k(r)$, width $e_{hj}^k(r)$ and area $e_{aj}^k(r)$, given by: $e_{lj}^k(r) = l_{oj}^k - l_{dj}^k(r)$, $e_{hj}^k = h_{oj}^k - h_{dj}^k(r)$, and $e_{aj}^k(r) = a_{oj}^k - a_{dj}^k(r)$. This produces respective average deviation values taken over groups $G_k(l)$. These are denoted by means of $\bar{\delta}_l^k(r)$, $\bar{\delta}_h^k(r)$, $\bar{\delta}_a^k(r)$, their corresponding averages taken over the whole collection of groups $G_k(l)$ by means of $\bar{\delta}_l(r)$, $\bar{\delta}_h(r)$, $\bar{\delta}_a(r)$ and the associated standard deviations through $\sigma_{\delta l}(r)$, $\sigma_{\delta h}(r)$ and $\sigma_{al}(r)$ respectively. Then, for each range of similarity $ST(r)$, the technique identifies the leaves satisfying the conditions

$$\bar{\delta}_h(r) \ge 0, \tag{1}$$

$$\bar{\delta}_l(r) \ge 0, \tag{2}$$

$$\bar{\delta}_l(r) - \sigma_{\delta l}(r) \le \bar{\delta}_l^k(r) \le \bar{\delta}_l(r) + \sigma_{\delta l}(r), \tag{3}$$

$$\bar{\delta}_h(r) - \sigma_{\delta h}(r) \le \bar{\delta}_h^k(r) \le \bar{\delta}_h(r) + \sigma_{\delta h}(r) \tag{4}$$

and

$$e_{aj}^k(r) \ge 0 \tag{5}$$

and use their area values $a_{dj}^k(r)$ to calculate $\lambda_a(r)$, which stands for the proportion of images of leaves for which a_d produces consistent estimations of observed leaf areas a_0. This proportion is calculated according to the formula,

$$\lambda_a(r) = \frac{\sum_{k=1}^{n} \sum_{j=1}^{n_k} [a_{dj}^k(r) \mid \text{leaves in } G_k(l) \text{ that comply with conditions (1) through (5)}]}{\sum_{k=1}^{n} \sum_{j=1}^{n_k} a_{oj}^k} \tag{6}$$

Then the method obtains the proportion $\beta_a(r)$ of images of leaves for which a_d estimations overestimates observed leaf areas a_o which is calculated through $\beta_a(r) = 1 - \lambda_a(r)$, and use $\lambda_a(r)$ and $\beta_a(r)$ to calculate the value of the image selection index $IS(r)$, formally defined by

$$IS(r) = \beta_a(r)/\lambda_a(r) \tag{7}$$

Finally, the DCM proposes the use of the $ST(r)$ interval producing the smallest value of $IS(r)$ for reliable estimation of the areas of leaves of eelgrass using images whose peripheral contour is distorted by noise induced by humidity contents.

The Concordance Correlation Method (CCM)
The Concordance Correlation Coefficient symbolized by mean of ρ [33,34] is used to determine reproducibility, as it measures the agreement between the variables x and y by appraising the extent to which they fall on the 45° line through the origin. Its numerical value is represented in terms of the ratio of the expected orthogonal squared distance from the diagonal $y = x$ to the expected

Echavarría-Heras *et al. Source Code for Biology and Medicine* (2014) 9:29

Page 4 of 10

orthogonal squared distance from the diagonal $y = x$ assuming independency. The value of ρ, is commonly used to assess how well a new set of observations y reproduce an original set x. When ρ is computed on a m-length data set (i.e., two vectors $(x_1, x_2, \cdots, x_m)$ and $(y_1, y_2, \cdots, y_m)$ the resulting statistics is denoted by means of $\hat{\rho}$ and calculated through

$$\hat{\rho} = \frac{2s_{xy}}{s_x^2 + s_y^2 + (\bar{x} + \bar{y})^2} \, , \tag{8}$$

being

$$\bar{x} = \frac{1}{m}\sum_{j=1}^{m} x_j \tag{9}$$

$$s_x^2 = \frac{1}{m}\sum_{j=1}^{m} \left(x_j - \bar{x}\right)^2 \tag{10}$$

and

$$s_{xy} = \frac{1}{m}\sum_{j=1}^{m} \left(x_j - \bar{x}\right)\left(y_j - \bar{y}\right) \tag{11}$$

In the present work the value of $\hat{\rho}$ will provide a criterion for the incumbent digital image selection process. The linked CCM does not require the sorting of observed leaf lengths into the $G_k(l)$ groups of the DCM. As it is done in the DCM, in the present CCM, the digital images of sampled leaves are primarily processed by a specified color format with a number C_{max} of colors and using intervals of tolerance of similarity $ST(r) = [0, r]$ with $0 \leq r \leq C_{max} - 1$. Again by keeping $ST(r)$ fixed and within the *jth* leaf image, a routine selects a starting point, and using Eqs. (A1), (A2) and (A3) detects all adjacent pixels connected within the realm of the designated interval of tolerance of similarity $ST(r)$. This device identifies the peripheral contour of the leaf image allowing associated measurements of length $l_{dj}(r)$ and width $h_{dj}(r)$ whose product for $1 \leq j \leq m$, yields image estimated leaf areas $a_{dj}(r)$. Instead of performing the statistical steps required to calculate $IS(r)$, simply for r fixed in equations (9), (10) and (11) we make x_j stand for observed leaf area measurements $(a_{01}, a_{02}, \cdots, a_{0m})$ and let y match digital image produced estimations $(a_{d0}(r), a_{d1}(r), \cdots, a_{dm}(r))$. Then equation (8) yields the resulting value of the Concordance Correlation Coefficient. In the present settings this will be denoted through by means of the symbol $\hat{\rho}(r)$ to emphasize its dependence on r, that is, changing $ST(r)$ produces different pairs of observed and image calculated leaf areas $(a_{0j}, a_{dj}(r))$, $1 \leq j \leq m$, as well as different values of the associated $\hat{\rho}(r)$. After all values of r in the chosen color format are exhausted, we select the tolerance of similarity interval $ST(r)$ that produces the highest value for $\hat{\rho}(r)$ for

efficient estimation of eelgrass leaves area from digital images with noise related to environmental factors.

Results and discussion

For the purposes of the present study, we used a data set obtained by randomly sampling 5 shoots biweekly from January through December 2009 in a *Zostera marina* field at Punta Banda estuary, a shallow coastal lagoon located near Ensenada, Baja California, Mexico (31° 43–46 N and 116° 37–40 W). For each sampled leaf, a millimeter ruler was used to obtain leaf length measurements l_o to the nearest $1/10\,mm$ taken as the distance from the top of the sheath to the leaf tip. Meanwhile, observed leaf width h_o was measured at a point halfway between the top of the sheath and the tip [32]. Observed leaf area estimations a_o were calculated by means of length times width proxy $a_o = l_o \cdot h_o$.

We obtained $l_{max} = 460\,mm$. For the data grouping required by the DCM we choose $n = 46$ so we acquired $q = 10\,mm$, and for the interval $[0, l_{max}]$ we formed a partition P_0^{460} of disjoint intervals I_k of the form $I_k = \{l \mid q(k-1) \leq l < qk\}$, with $1 \leq k \leq 46$. Hence, for each value of the index k, we formed a group $G_k(l)$ containing leaves with sizes varying in the interval I_k. Longer and older leaves displayed darker tonalities than younger and shorter ones, but leaves with lengths varying on a given partition interval I_k displayed a similar color distribution. For some of the partition intervals there was at most one leaf with length placed in the linked variation range. Therefore, these groups are not taken into account because they do not provide information for the statistical analysis.
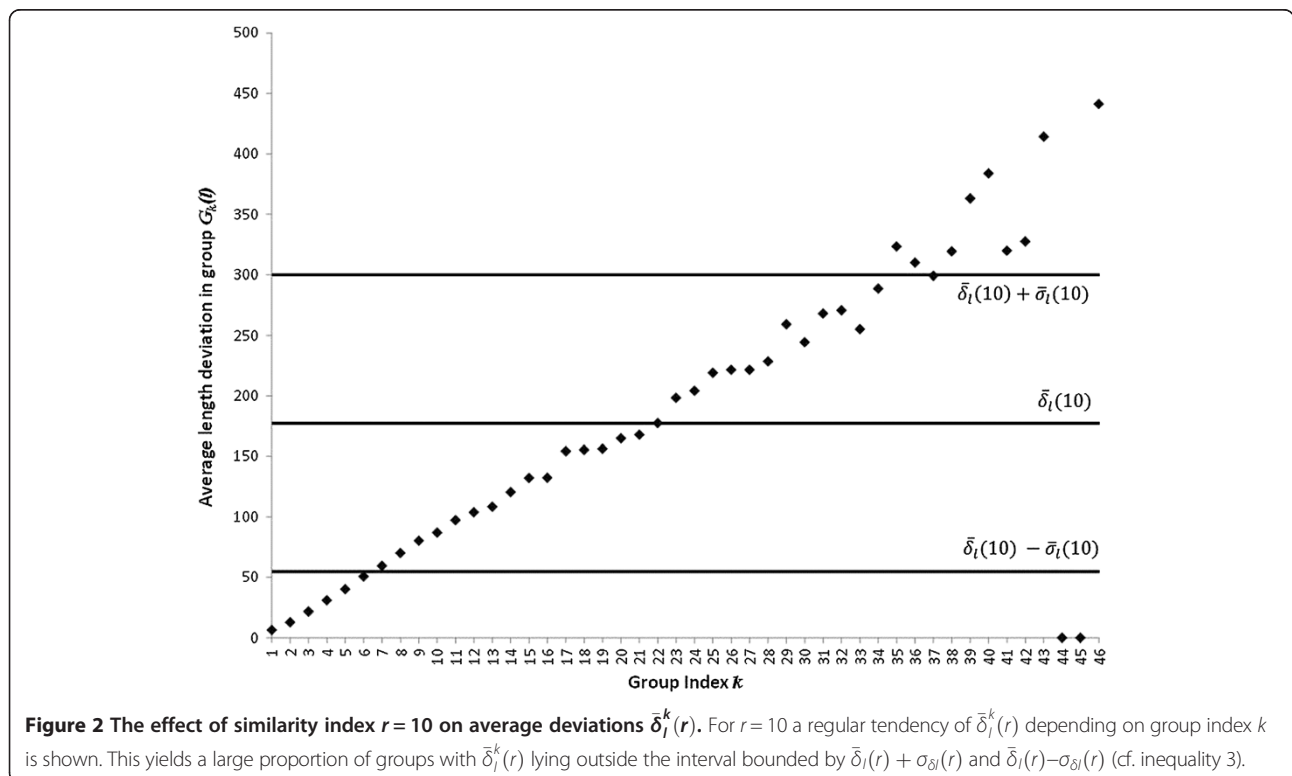
According to the DCM, for each leaf belonging to the group $G_k(l)$ we obtained its digital image. For dealing out with all these individual images we selected an RGB color format with a number C_{max} of 256 colors. For processing each one of the available leaves images, we choose different tolerance of similarity levels $ST(r) = [0, r]$ with the upper bound r satisfying $0 \leq r \leq C_{max} - 1$. Then for a given $ST(r)$ range, we selected a starting point inside the considered leaf image, and identified using equations (A1), (A2) and (A3) all adjacent pixels falling within the named similarity range $ST(r)$. This recognizes the outer contour of the digital blade, and produce concomitant leaf width, length and area estimations. The next step in the DCM concerns the calculation of the selection index $IS(r)$ which depends on the value of the $\lambda_a(r)$ statistics. But according to equation (6) obtaining the value of $\lambda_a(r)$ requires counting the number of leaves in each group $G_k(l)$, that comply with conditions (1) through (5) and these numbers depend on the chosen value of r. Moreover, for small values of r the number of different tonalities included in $ST(r)$ is limited so identification of pixels within an image

can be expected to be imprecise. This is handily clarified by Figure 2, produced using $r = 10$ and that shows a systematic tendency for average length deviations $\bar{\delta}_l^k(r)$ depending on group index k. As a result we can observe a large number of $\bar{\delta}_l^k(r)$ values lying above the $\bar{\delta}_l(r) + \sigma_{\delta l}(r)$ and beyond the $\bar{\delta}_l(r) - \sigma_{\delta l}(r)$ thresholds in inequality (3). The bigger the value of r, the greater the number of color tonalities included in the interval $ST(r)$ and precision in image contour identification improves. This is observed in Figure 3, produced using $r = 128$ and which does not display the above quoted systematic tendency, but a reduced number of groups of leaves with average length deviations $\bar{\delta}_l^k(r)$ lying outside the interval bounded by $\bar{\delta}_l(r) + \sigma_{\delta l}(r)$ and $\bar{\delta}_l(r) - \sigma_{\delta l}(r)$. Consequently, for small values of r we can expect reduced values of $\lambda_a(r)$ and as a result according to equation (7) large values of $IS(r)$. Additionally, when the interval of tolerance of similarity becomes wider, smaller values of $IS(r)$ can be expected. In fact, as shown in Figure 4, the DCM captures this effect in a consistent way, with small values of r leading to large values for the selection index $IS(r)$. Moreover, through the interval $1 \leq r < 128$, $IS(r)$, decreases reaching a minimum value of 0.91, attained at $r = 128$. Meanwhile, for $r \geq 128$, the values of the selection index $IS(r)$ steadily increased towards a value of 1.84, attained at $r = 255$. Therefore, according to the DCM selection criterion $ST(128)$ must be

chosen for efficient estimation of areas of eelgrass leaves using images with noise induced by humidity contents.

Now for the CCM, since a small value of r fails to recognize some pixels in the digital image, we might expect a low reproducibility of directly obtained measurements $(a_{01}, a_{02}, \cdots, a_{0m})$ by means of digitally obtained proxies $(a_{d0}(r), a_{d1}(r), \cdots, a_{dm}(r))$. This is indeed shown in Figure 5. Moreover , the larger the value of r, the greater the number of color tonalities included in the interval $ST(r)$, as a result exactness in image contour identification increases, and reproducibility improves, this explaining why Figure 5 shows increasing values of $\hat{\rho}(r)$ through the interval $1 \leq r < 128$. Moreover, through the domain $128 \leq r < 178$ the values of $\hat{\rho}(r)$ are maintained within a plateau of slight variation around $\hat{\rho}(128) = 0.90$, but for $178 \leq r \leq 255$, $\hat{\rho}(r)$ decreases dropping to a value of 0.8464, attained at $r = 255$. Thenceforth, intervals of tolerance of similarity, wider than $ST(128)$ do not improve reproducibility of observed values of leaves areas by means of their image obtained surrogates. Thus, for the sake of accuracy and simplicity, $ST(128)$ should be used for image selection when noise due to environmental factors is present and efficient estimations of eelgrass leaf area taken from these images are required.

In order to assess robustness of the CCM, we performed a resampling experiment. We chose a sample size index $p = 1, 2, ..., 8$ then for each value of p a set $s(p)$

Figure 2 The effect of similarity index $r = 10$ on average deviations $\bar{\delta}_l^k(r)$. For $r = 10$ a regular tendency of $\bar{\delta}_l^k(r)$ depending on group index k is shown. This yields a large proportion of groups with $\bar{\delta}_l^k(r)$ lying outside the interval bounded by $\bar{\delta}_l(r) + \sigma_{\delta l}(r)$ and $\bar{\delta}_l(r) - \sigma_{\delta l}(r)$ (cf. inequality 3).

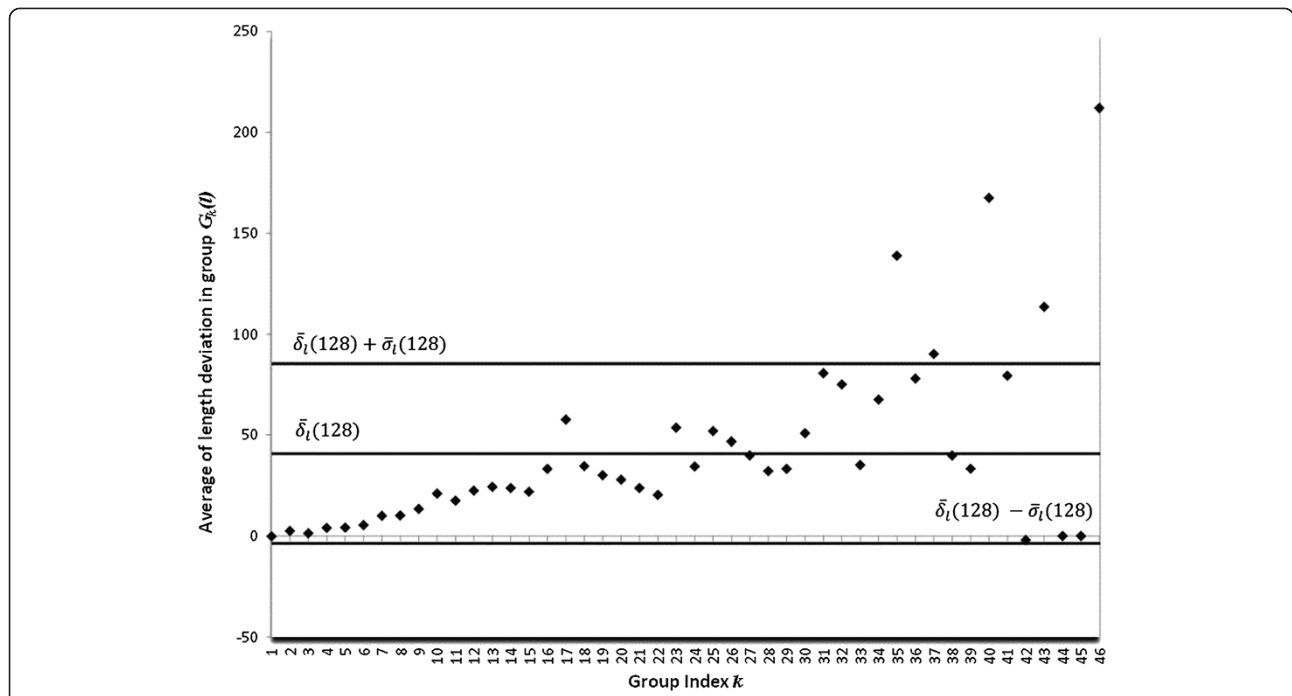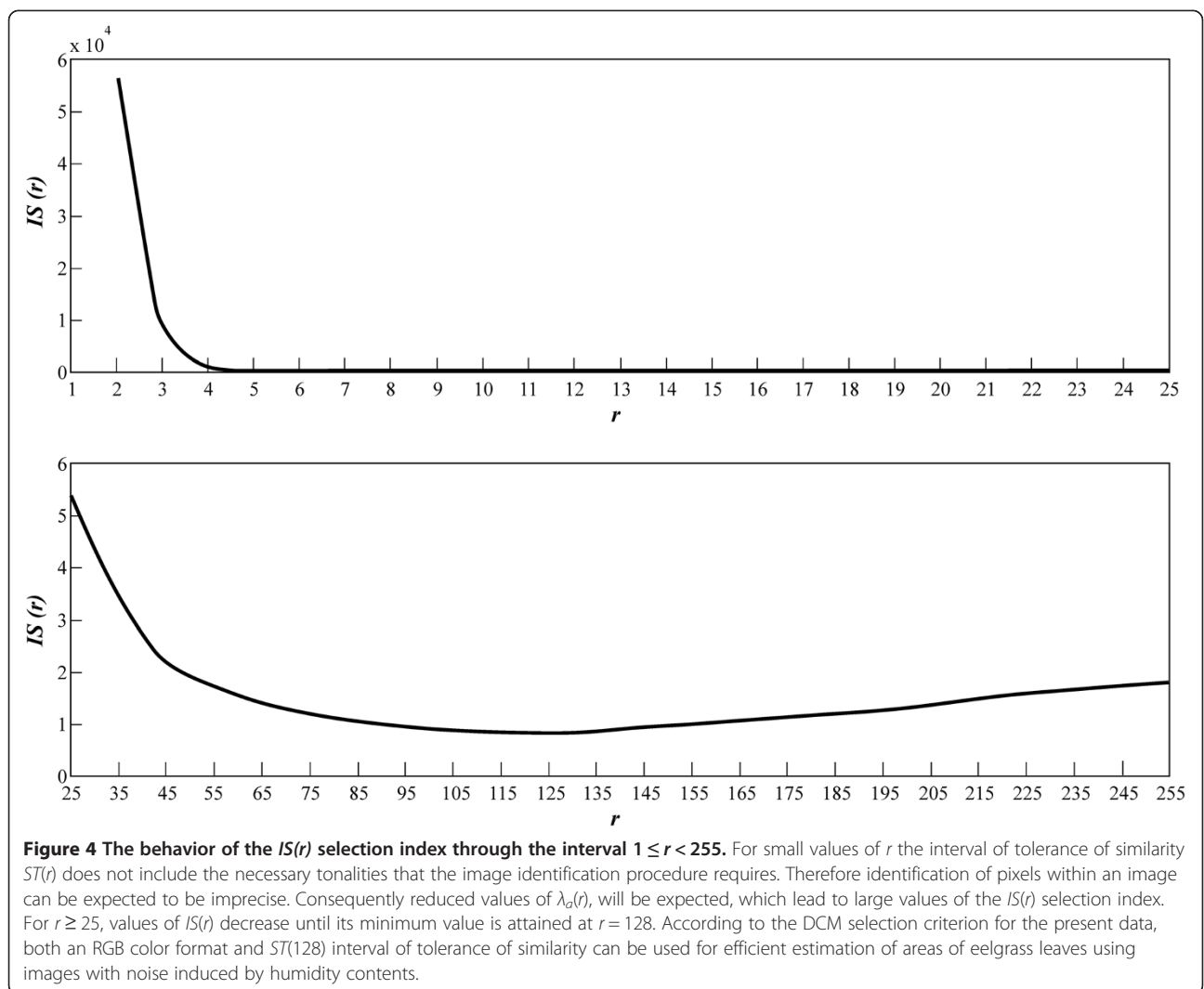Echavarría-Heras *et al. Source Code for Biology and Medicine* (2014) 9:29

Page 6 of 10

Figure 3 The effect of similarity index $r = 128$ on average deviations $\bar{\delta}_l^k(r)$. For $r = 128$ the regular tendency of increasing $\bar{\delta}_l^k(r)$ values shown figure 2 is no longer observed and a reduced number of $\bar{\delta}_l^k(r)$ values lying outside the interval bounded by $\bar{\delta}_l(r) + \sigma_{\delta l}(r)$ and $\bar{\delta}_l(r) - \sigma_{\delta l}(r)$ is observed.

of samples of size $100p$ each were uniformly drawn from the $(a_{01}, a_{02}, \cdots, a_{0m})$ population. Next, we selected one of the $s(p)$ samples, for each value of r through the interval $1 \le r \le 255$, we designated the matching areas obtained from digital images and calculated the concomitant Concordance Correlation Coefficient values $\hat{\rho}(r)$. We recorded the value of r at which the maximum for $\hat{\rho}(r)$ was attained for the selected sample. We repeated this procedure for all the samples in the set $s(p)$ and then averaged the obtained r values for maximum $\hat{\rho}(r)$. Figure 6 displays the obtained averages for the different values of the sample size index p. The maximum values of $\hat{\rho}(r)$ per sample were also averaged over the $s(p)$ sets. These last average values are shown in Figure 6. The results of this study show that the optimal interval of tolerance of similarity, as well as, the reproducibility of observed leaf areas by means of their digital image surrogates can be considered independent of sample size. Therefore, the CCM can be regarded as a robust procedure.

According to our results, both methods sustain the same conclusion regarding the choosing of $ST(128)$ on behalf of accuracy. However, in comparison to the complicated multi-stage procedures of the DCM, using $\hat{\rho}(r)$ values provide a direct and simpler criterion for choosing an interval of tolerance of similarity $ST(r)$ for reliable digital image related assessments of eelgrass leaf area under the specified noise effects. But the main advantage

of the CCM resides on the fact that it allows a straightforward interpretation of the addressed digital image selection procedures in terms of a measure of reproducibility. Indeed the plateau in $\hat{\rho}(r)$ values linked to the domain $128 \le r \le 178$, and the subsequent decreasing mode associated to $r \ge 178$ shown in Figure 5 indicate that intervals of tolerance of similarity wider than $ST(128)$ will fail to improve reproducibility of observed values of leaf area by means of their image produced proxies. In other words for $r \ge 128$, $ST(r)$ includes more tonalities than those contained within the real image, thereby favoring the incorporation of spurious entries appearing beyond its peripheral contour and within the framing of the image. Thus including more color tonalities than necessary in the image processing task could not grant a gain in accuracy, but instead, depending on the severity of the noise effects (Figure 1), and on the size of the framing enclosing the peripheral contour of the image (Figure 1), more spurious pixels could be taken in to account by the image processing devise, which could lead to increased miscalculation of leaf area obtained from images. Meanwhile, our analysis confirms that when noise induced into images by the humidity contents of the leaves reduces the accuracy of estimations of the associated areas we could use a RGB color format, an $ST(128)$ interval of tolerance of similarity and equations (A1), (A2) and (A3) to identify the peripheral contour of leaves images for optimal reproducibility.

Echavarría-Heras *et al. Source Code for Biology and Medicine* (2014) 9:29

Page 7 of 10

Figure 4 The behavior of the *IS(r)* selection index through the interval $1 \leq r < 255$. For small values of *r* the interval of tolerance of similarity *ST(r)* does not include the necessary tonalities that the image identification procedure requires. Therefore identification of pixels within an image can be expected to be imprecise. Consequently reduced values of $\lambda_a(r)$, will be expected, which lead to large values of the *IS(r)* selection index. For $r \geq 25$, values of *IS(r)* decrease until its minimum value is attained at $r = 128$. According to the DCM selection criterion for the present data, both an RGB color format and *ST*(128) interval of tolerance of similarity can be used for efficient estimation of areas of eelgrass leaves using images with noise induced by humidity contents.

Conclusions

The results of the present digital image selection procedure provide simple, unbiased and non-destructive measurements of eelgrass leaf area. These measurements in conjunction with allometric methods [35] can predict the dynamics of biomass and leaf growth through indirect techniques, reducing the destructive effect of sampling and simplifying time consuming methods in the laboratory [36]. Nevertheless, it is worth to emphasize, that leaves removed from a shoot readily begin to lose water and degrade, so changes in shape may occur [37]. Therefore, even though humidity contents could certainly induce noise effects, an efficient digitalizing of a *Zostera marina* blade requires the maintenance of an optimal humidity for increased image fidelity. By taking this into account we can assert that the apparent similarity of values of $\hat{\rho}(r)$ linked to the interval $128 \leq r \leq 178$ could not be exhibited as a weakness of the CCM, that is, the plateau shown in Figure 5 does not associate

to vagueness in the imbedded selection criteria. Indeed in this study both the preparation of lives before digitalization procedures and the framing used to bound the area surrounding the peripheral contour of the digital leaves was effective (1) for reducing inconsistencies attributable to a biased mapping of leaf shape into images, (2) by lessening bias due to the inclusion of spurious entries linked to noise into images and (3) because the framing size used in the present identification procedure further limited the participation of spurious entries in image processing tasks. Therefore, $r = 128$ (that is, the entrance threshold for the plateau of maximum $\hat{\rho}(r)$ values in Figure 5) includes the required number of different tonalities for the processing of the present set of images and we choose it for a consistent estimation of the pertinent leaf area. Although, in the present settings the aforementioned bias reduction practices explain why values of *r* beyond $r = 128$ sustain the same selection criterion, using $r \geq 128$ could lead to extended time consuming computational
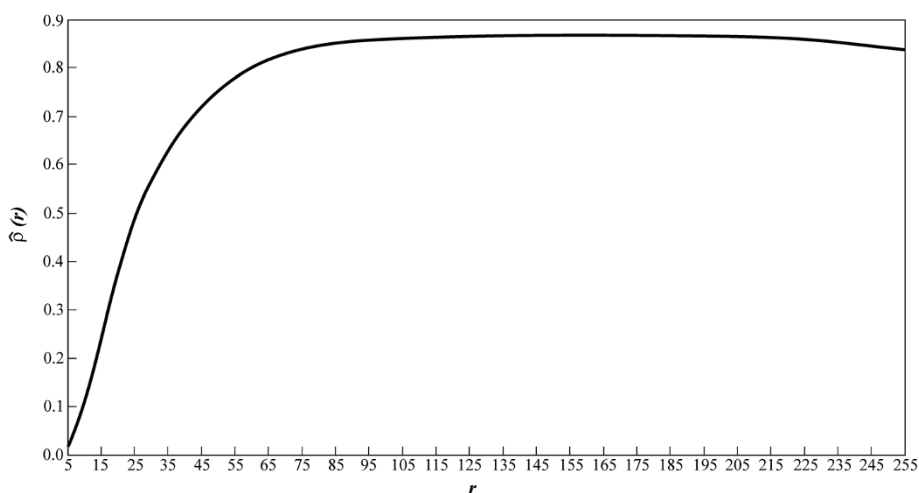
Figure 5 The behavior of the concordance correlation coefficient $\hat{\rho}(r)$, through the interval $1 \leq r < 255$. Increasing values of $\hat{\rho}(r)$ through the interval $1 \leq r < 128$ are displayed. This means that the wider the interval of similarity $ST(r)$ the greater the reproducibility of observed leaf areas by image proxies becomes. Interestingly through the domain $128 \leq r < 178$ values of $\hat{\rho}(r)$ are maintained within a plateau of slight variation around $\hat{\rho}(128) = 0.90$. Afterwards, for $178 \leq r \leq 255$ values of $\hat{\rho}(r)$ decrease slightly until $\hat{\rho}(r)$ drops to a value of 0.8464 attained at $r = 255$. Then for values of r larger than $r = 128$ reproducibility is not improved and coinciding with the criterion in the DCM for the present data, both an RGB color format and the $ST(128)$ interval of tolerance of similarity could be used for image selection when noise due to humidity contents is present and efficient estimations of eelgrass leaf area taken from these images is required.

procedures, because more than necessary tonalities will be included in the identification undertaking. It is also worth to highlight that in further applications, before the CCM could provide consistent results, care should be taken in order to ensure that the handling of samples be performed in an efficient way for reducing bias in the overall image selection procedures. Indeed we could anticipate that in settings where points (1) through (3) above are disregarded, the inherent bias could seriously reduce reproducibility. Nevertheless, this could not be exhibited as a
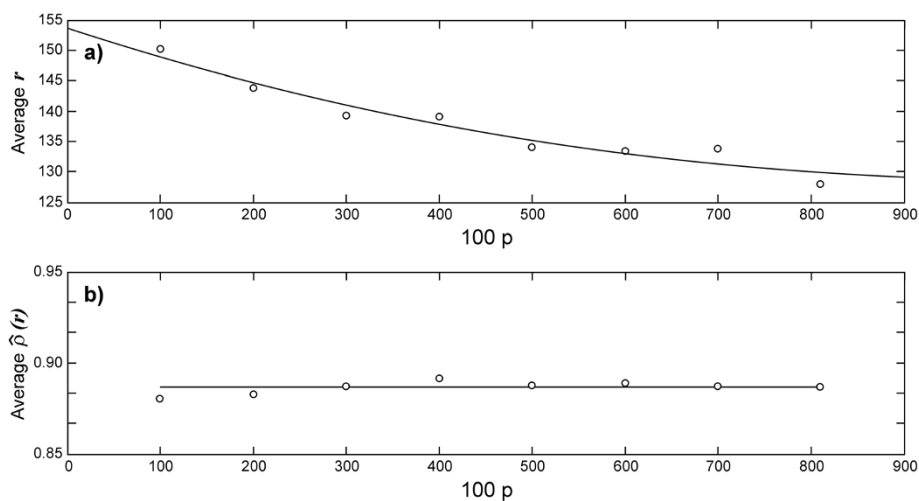

Figure 6 Dependence of both the value of r for maximum $\hat{\rho}(r)$ and the maximum value of $\hat{\rho}(r)$ itself on sample size. For each value of the sample size index $p = 1, 2, \ldots 8$ a fixed number $s(p)$ of samples of size $100p$ each were uniformly drawn from the population of observed leaf areas. For each one of the samples in a set $s(p)$, we iterated values of r through the interval $1 \leq r \leq 255$, and for each one of these r values we obtained the concomitant concordance correlation coefficient values $\hat{\rho}(r)$. We repeated this procedure for all the samples in the set $s(p)$ and averaged the r values at which $\hat{\rho}(r)$ attained its maximum value, the obtained averages for the different values of the sample size index p are shown in **a)**. The maximum values that $\hat{\rho}(r)$ obtained in a sample were also averaged over the $s(p)$ sets. These average values depending on sample size are correspondingly shown in **b)**. The results of this study shows that neither the optimal interval of tolerance of similarity $ST(r)$ or the reproducibility of observed leaf areas by means of their digital image surrogates depend on sample size, therefore the CCM can be considered as, a robust procedure.

weakness of the present CCM, since the DCM itself as well as any other image selection procedure is subject to the same bias effects. In summary, the CCM, not only provides a simplified and robust image processing device, besides, (a) this criterion offers a conceptual substantiation for the DCM itself by linking the minimum values of the selection index $IS(x)$, to the maximum values of the Concordance Correlation Coefficient $\hat{\rho}(r)$, and (b) even though here we applied the CCM to account solely for the effects of noise linked to humidity contents, it is worth to mention that since the core of the CCM criterion is the evaluation of reproducibility, its scope directly embraces the treatment of any kind of noise effects that can reduce the accurateness of digital image proxies of areas of eelgrass leaves.

Studies of seagrass communities such as those composed of *Zostera marina* show that these systems are among the most productive marine systems [38]. The characterization of the dynamics of such ecosystems is important from both a scientific and conservation perspective. Moreover, the methods sustained by the present research may be fundamental to the evaluation of eelgrass restoration projects and could thereby contribute to the conservation of this important seagrass species.

Appendix

We describe here the conceptual and formal framework for digital image processing. Two pixels are adjacent if, and only if, they share one of their borders, or at least one of their corners. Two pixels are neighbors if they fulfill the definition of adjacency. Formally, the vicinity $V_p(x, y)$ of the point $P(x, y)$ is defined through

$$V_p(x,y) = \left\{ \begin{array}{c} (x+1,y), (x-1,y), (x,y+1), (x,y-1), \\ (x+1,y+1), (x+1,y-1), (x-1,y+1), (x-1,y-1) \end{array} \right\} \tag{A1}$$

Without loss of generality, we explain the notion of tolerance of similarity, by referring to the Reed, Green and Blue (RGB) color space. This allows quantifying tonality in terms of the intensities of the constituting primary colors: red, green, and blue. To indicate at which amount each one of these colors is mixed, to produce a given tonality a value is assigned to each prime color, for example, the value 0 means that a given primary color does not appear in the mix, but if a chief color component is non-vanishing it means that it contributes to the mix in a given intensity. We introduce C_{max} which identifies the number of colors to be used through the whole image processing task. For an RGB color space we have $C_{max} = 256$. Usually, the intensity of each of the primary colors appearing in a mix is measured on a scale ranging from 0 to $C_{max} - 1$. The set of all color intensities can be represented in the form of a cube in the Cartesian coordinate system, where each color is a point on the surface or in its interior. Given points $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ in an RGB color space, we will define the distance $d_E(P, Q)$ between them through,

$$d_E(P, Q) = \sqrt{\sum_{i=1}^{n} (p_n - q_n)^2} \tag{A2}$$

Moreover, given a point P in an RGB color space, a second one Q with the greatest similarity to P is the one placed at the smallest distance $d_E(P, Q)$. Furthermore, let $ST(r) = [0, r]$ be a color tonality range, being r the number of different colors included. Then, we must have $1 \leq r \leq C_{max} - 1$ and we will say that two pixels P and Q are similar to a tolerance limit $ST(r)$ if the inequality

$$d_E(P, Q) \leq r \tag{A3}$$

is satisfied. The range $ST(r)$ is called "interval of tolerance of similarity" and the upper bound r can be interpreted as the maximum distance that two points located within the extent of an object can attain in a RGB color space in order to be considered similar. Connectivity between pixels is used to identify the limits in objects and regions in an image. We will say that two pixels P and Q are connected with tolerance of similarity $ST(r)$ if they fulfill the definition of adjacency and also if inequality (A3) holds.

Author details
[1]Centro de Investigación Científica y de Estudios Superiores de Ensenada, Carretera Ensenada-Tijuana No. 3918, Zona Playitas, Apdo. Postal 360, Código Postal 22860 Ensenada, B C, México. [2]Centro de Investigación en Matemáticas, A.C. Jalisco s/n, Mineral Valenciana, Guanajuato Gto. Código Postal 36240, México. [3]Instituto Tecnológico de Tijuana, Tijuana, Baja California, México.

References
1. Holmquist JG, Powell GVN, Sogard SM: **Decapod and stomatopod assemblages on a system of seagrass-covered mud banks in Florida Bay.** *Mar Biol* 1989, **100**:473–483.
2. Montague CL, Ley JA: **A possible effect of salinity fluctuation on abundance of benthic vegetation and associated fauna in northeastern Florida Bay.** *Estuar Coast Shelf Sci* 1993, **16**:703–717.
3. Plummer ML, Harvey CJ, Anderson LE, Guerry AD, Ruckelshaus MH: **The role of eelgrass in marine community interactions and ecosystem**

Echavarría-Heras *et al. Source Code for Biology and Medicine* (2014) 9:29

Page 10 of 10

services: results from ecosystem-scale food web models. *Ecosystems* 2013, **16**(2):237–251.

4. Blackburn TH, Nedwell DB, Weibe WJ: **Active mineral cycling in a Jamaican seagrass sediment.** *Mar Ecol Prog Ser* 1994, **110:**233–239.

5. Park SR, Li WT, Kim SH, Kim JW, Lee KS: **A comparison of methods for estimating the productivity of Zostera marina.** *J Ecol Field Biol* 2010, **33**(1):59–65.

6. Terrados J, Borum J: **Why are Seagrasses Important? Goods and Services Provided by Seagrass Meadows.** In *European Seagrasses: an Introduction to Monitoring and Management*. Edited by Borum J, Duarte CM, Krause-Jensen D, Greve TM. The M&MS project; 2004:88. http://www.vliz.be/en/imis?module=ref&refid=70489&pp=print. ISBN 87-89143-21-3.

7. Newell IER, Koch EW: **Modeling seagrass density and distribution in response to changes in turbidity stemming from bivalve filtration and seagrass sediment stabilization.** *Estuar Coast Shelf Sci* 2004, **27**(5):793–806.

8. Liu X, Zhou Y, Yang H, Ru S: **Eelgrass detritus as a food source for the Sea Cucumber Apostichopus japonicus Selenka (Echinodermata: Holothuroidea) in Coastal Waters of North China: an experimental study in flow-through systems.** *PLoS One* 2013, **8**(3):e58293.

9. Kennedy H, Beggins J, Duarte CM, Fourqrean JW, Holmer M, Marbà N, Middelburg JJ: **Seagrass sediments as a global carbon sink: isotopic constraints.** *Global Biogeomechanical Cycles* 2010, **24:**GB4026.

10. Short FT, Coles RG, Pergent-Martini C: **Global Seagrass Distribution.** In *Global Seagrass Research Methods*. Edited by Short FT, Coles RG. Amsterdam, The Netherlands: Elsevier Science B.V; 2001:5–30.

11. Phillips RC: **Temperate Grass Flats.** In *Coastal Ecological Systems of the United States*. 2nd edition. Edited by Odum HT, Copeland BJ, Mc Mahan EA. Washington DC: Conservation Foundation; 1974:244–299.

12. Williams RB: **Nutrient Level and Phytoplankton Productive in the Estuary.** In *Proceedings of the Coastal Marsh and Estuary Management Symposium*. Edited by Chabreck RA. Baton Rouge, USA: Louisiana State University; 1973:59.

13. Jacobs RPWM: **Distribution and aspects of the production and biomass of eelgrass, Zostera marina L., at Roscoff, France.** *Aquat Bot* 1979, **7:**151–172.

14. Mukai H, Aioi K, Ishida Y: **Distribution and biomass of eelgrass (Zostera marina L.) and other seagrasses in Odawa Bay, central Japan.** *Aquat Bot* 1980, **8:**337–342.

15. Phillips RC, Backman TW: **Phenology and reproductive biology of eelgrass (Zostera marina L.) at Bahia Kino, Sea of Cortez, Mexico.** *Aquat Bot* 1983, **17:**85–90.

16. Dennison WC, Alberte RS: **Role of daily light period in the depth distribution of Zostera marina (eelgrass).** *Mar Ecol Prog Ser* 1985, **25:**51–61.

17. Bulthuis DA: **Effects of temperature on photosynthesis and growth of seagrasses.** *Aquat Bot* 1987, **27:**27–40.

18. Solana-Arellano E, Echavarria-Heras HA, Ibarra-Obando SE: **Leaf size dynamics for Zostera marina L, in San Quintin Bay, Mexico: a theoretical study.** *Estuar Coast Shelf Sci* 1997, **44:**351–359.

19. Nadezhda Z, Sfriso A, Voinov A, Pavoni B: **A simulation model for the annual fluctuation of Zostera marina biomass in the Venice Lagoon.** *Aquat Bot* 2001, **20:**135–150.

20. Echavarria-Heras H, Solana-Arellano E, Franco-Vizcaino E: **The role of increased sea surface temperature on eelgrass leaf dynamics: onset of El Niño as a proxy for global climatic change in San Quintín Bay, Baja California.** *Bull Southern Calif Acad Sci* 2006, **105:**113–127.

21. Short FT, Neckles HA: **The effects of global climate change on seagrasses.** *Aquat Bot* 1999, **63:**169–196.

22. Orth RJ, Moore A: **Chesapeake Bay: an unprecedented decline in submerged aquatic vegetation.** *Science* 1983, **222:**51–53.

23. Short FT, Wyllie-Echeverria S: **Natural and human-induced disturbance of seagrasses.** *Environ Conserv* 1996, **23:**17–27.

24. Short FT, Burdick DM, Granger S, Nixon SW: **Long-Term Decline in Eelgrass Zostera marina L. Linked to Increase Housing Development.** In *Seagrass Biology, Proceedings of an International Workshop*. Edited by Kuo J, Phillips RC, Walker DI, Kirkman H. Western Australia: Rottnest Island: 1996:291–298.

25. Lee KS, Park JI: **An effective transplanting technique using shells for restoration of Zostera marina habitats.** *Mar Pollut Bull* 2008, **56:**1015–1021.

26. Orth RJ, Harwell MC, Fishman JR: **A rapid and simple method for transplanting eelgrass using single, unanchored shoots.** *Aquat Bot* 1999, **64:**77–85. ISSN 0304-3770, http://dx.doi.org/10.1016/S0304-3770(99)00007-8.

27. Campbell ML, Paling EI: **Evaluating vegetative transplanting success in Posidonia australis: a field trial with habitat enhancement.** *Mar Pollut Bull* 2003, **46:**828–834.

28. Fishman JR, Orth RJ, Marion S, Bieri J: **A comparative test of mechanized and manual transplanting of eelgrass, Zostera marina, in Chesapeake Bay.** *Restoration Ecol* 2004, **12:**214–219.

29. Echavarría-Heras H, Lee AKS, Solana-Arellano ME, Franco-Vizcaino E: **Formal analysis and evaluation of allometric methods for estimating above-ground biomass of eelgrass.** *Ann Appl Biol* 2011, **159**(3):503–515.

30. Echavarría-Heras HA, Solana-Arellano ME, Leal-Ramírez C, Franco-Vizcaíno E: **An allometric method for measuring leaf growth in eelgrass, Zostera marina, using leaf length data.** *Botánica Marina* 2013, **56**(3):275–286.

31. Leal-Ramirez C, Echavarria-Heras H: **A method for calculating the area of Zostera marina leaves from digital images with noise induced by humidity content.** In *The Scientific World Journal.* ; 2014:11.

32. Echavarría-Heras HA, Solana-Arellano ME, Leal-Ramírez C, Franco-Vizcaino E: **The length-times-width proxy for leaf area of eelgrass: criteria for evaluating the representativeness of leaf-width measurements.** *Aquat Conserv Mar Freshwater Ecosystems* 2010, **21**(7):604–613.

33. Lin LIK: **A concordance correlation coefficient to evaluate reproducibility.** *Biometrics* 1989, **45:**255–268.

34. Lin LIK: **Assay validation using the concordance correlation coefficient.** *Biometrics* 1992, **48:**599–604.

35. Echavarrıa-Heras H, Solana-Arellano E, Franco-Vizcaıno E: **An allometric method for theprojection of eelgrass leaf biomass production rates.** *Math Biosci* 2010, **223**(1):58–65.

36. Solana-Arellano E, Borbón-González DJ, Echavarria-Heras H: **A general allometric model for blade production in Zostera marina L.** *J Calif Acad Sci* 1998, **97:**39–48.

37. Juneau KJ, Tarasoff CS: **Leaf area and water content changes after permanent and temporary storage.** *PLoS One* 2012, **7**(8):1–6.

38. Zieman JC, Wetzel RG: **Productivity in Seagrasses: Methods and Rates.** In *Handbook of Seagrass Biology: an Ecosystem Perspective*. Edited by Phillips RC, McRoy P. New York and London: Garlanda STPM Press; 1980:87–115.