Assessing goodness of fit for CSTT model (ECASA internal paper) Napier University

Enrique Portilla, Paul Tett

January 31, 2008

1 Introduction

ECASA WP4's objective 2 cover the development of operational tools, especially models, which capture the functional relationship between environment and aquacultural activities, and which embody the chosen indicators. The WP's deliverable is a"Toolpack" report on the merits of the chosen indicator set including best methodologies for collection, analysis and interpretation, and on the recommended set of models, including criteria for choice of models depending on spatial scale and farm size, and guidance on the use of models to estimate site and water body assimilative capacity and sustainable production, and on the reliability of model predictions.

Here, we present the results of the simplified procedures to assess the *reliability of model predictions* presented to all ECASA modellers. Modellers were asked to use this procedure, so that the project can report a comparable set of model (and indicator) assessments. Nevertheless, due to the nature of the model that occupies us here (The CSTT (1994, 1997) model or the worst case scenario of chlorophyl concentration), a slightly different methodology was used that the one proposed initially. Nevertheless, for the interest of the inter-comparability, we will also report on the standard proposed methodology.

Therefore, this report applies both the standard test for the accuracy and precision of all ECASA models and the new approach. Both techniques are based on the comparison between simulated data and observed data using linear regression. However the way models are judged and methods to find the regression line differ. In the first case, the standard methodology will evaluate the model on the basis of (i) the proportion of variance explained by the regression, and (ii) the regression coefficients. These criteria are similar to those described by Mesple et al. (1996). In the second, although a regression line is fitted and the coefficients will be evaluated, the methodology to find the line is different from the previous method. We will use instead the quantile regression (Cade et al., 1999).

We will follow Oreskes et al. 1994, and will judge a model as *excellent, good, fair* or *poor*, accepting their argument that this terminology encourages model use and improvement.

This document has two sections. The first part presents the result of the application of procedures agreed by all ECASA modellers, to simulations made with the CSTT model for 1975 and 2003. The second part is Appendix A with goodness of fit tabulated according to the agreed format.

2 Material and Methods

2.1 Introduction

We have confronted simulations with the Comprehensive Studies Task Team (CSTT) (CSTT, 1994, 1997) model with data from Loch Creran in two different years, 1975, prior to fish-farming, and 2003. The CSTT maximum concentration on microplankton (phytoplankton plus pelagic microheterotrophs) in sea lochs. The output is a single value of Clorophyll, in a zone B scale water body. The physics is normally basic assuming constant exchange rate in the box. When possible the CSTT model was coupled with the results from FJORDENV-like physics (Gillibrand and Inall, 2006; Portilla and Tett, 2006) for some ECASA sites having time series of nutrient concentration and exchange in the box. More precisely, for the case of Loch creran,instead of a single maximum value for the summer as reported in (Tett et al., 2003), we provide with the maximum possible concentration of chlorophyll as a function of nutrient concentration.

By including FJORDENV-like physics (Gillibrand and Inall, 2006) and estimating the seasonal concentration enhancement of some nutrients (Portilla and Tett, 2006), we are able to populate the observation versus predictions graph with data.

2.2 Variables used in the test

The variables simulated by the CSTT model maximum concentration for values of 1 layer box Chlorophyll. The model aims to predict the worst case scenario as a function of available nutrients for a single micro-plankton compartment. The methodology for predicting the maximum concentration of chlorophyll for given concentration of limiting nutrients in equilibrium in a Region of Restricted Exchange (RRE) can be found in CSTT (1994, 1997); Tett et al. (2003). The key parameter to obtain the predictions is conversion factor of from Nitrogen to chlorophyll q. The worst case scenario for the concentration of chlorophyll (X_{max})

is obtained when all the nutrient inputs is converted to phytoplankton.

$$X_{max} = X_0 + q \cdot S_{eq} \tag{1}$$

where the equilibrium enhanced concentration of nutrient (S_{eq}) in the Region of Restricted Exchange (RRE) is the nutrient concentration in the absence of local consumption by phytoplankton, X_0 is the concentration of chlorophyll in the boundary conditions and q is the yield of chlorophyll from Nitrogen q and equal to 1.2 (Edwards et al., 2005). As announced above, it was possible to estimate time series of nutrient concentration enhancement (S_{eq}) for Loch Creran using the output from the seasonal Equilibrium Concentration Enhacement model Portilla and Tett (2006). In those cases, time-series of X_{max} where possible obtain.

The CSTT model was used to simulate chlorophyll in the fjordic loch Creran in 1975 and 2003, with forcing from daily values of simulated limiting nutrient concentration enhancement. The model was also tested in two other ECASA sites, the Venice Lagoon in Italy (45.1253 N 12.4525 E) and Ria Farmosa in Portugal (37 00.350 N 07.56.917 E). Finally, Predictions from the CSTT model and observed values from the OAERRE project where also gathered (Tett et al., 2003) .

For loch Creran in 2003, a salmon farm was adding significant N and P to the loch, in daily amounts calculated by the 'Blackfish' routine in CSTT from data on monthly amounts of feed. For the Venice Lagoon and Ria farmosa, the addition of N is quite similar, and about 200 kilo-moles per day. The ammount of external imput of nutrients of the other OARRE points is diverse and can be found in Tett et al. (2003).

For Loch Creran, the samples were collected from the upper layer of the main basin in 1975 and 2003. Data for 1975 were taken from the thesis by Jones (1979), and data for 2003 are those reported by Laurent et al. (2006). In the Venice lagoon, the data of chlorophyll was sampled between July 2005 and May 2005, and in Ria farmosa between June 2005 and October 2006.

For the two years in loch Creran, Chlorophyll observations were plotted with the corresponding simulated maximum values of chlorophyll for the upper layer. We displayed the observations and the prediciton in a time series graph and confronted the observations and predicitons of the the model in the same graph. Linear regressions were fitted using the well-known Ordinary Least Squares (OLS) method. For all the bulk of data we perfomed a quantile regression. We will confront the observed values and predicted maximum chlorophyll in all the location. The regression line that will summarise the cloud of points will not be a line goring throught the points, but leaving most of them above the line.

Table 1: Criteria established to classify the model predictions (X) in different grades when compared with data (Y). Four groups are formed according to the estimates of the slope (β_1) and intercept (β_0) from the regression line of the form $Y = \beta_1 X + \beta_0 + \epsilon$ (being ϵ the error term of the regression model). Within each of the groups the quality of the simulation increases with increasing values of R^2 (the coefficient of determination). The statement $\neq$ corresponds to the result of a t-test being *not significantly different from*, at the level $\alpha = 0.05$

Model category	Slope	Intercept	Interpretation
Excellent	$\beta_1 = 1$	$\beta_0 = 0$	The model is regarded as perfectly simulating, on average, the observations
Good (effect 1)	$\beta_1 = 1$	$\beta_0 \neq 0$	The model over or underestimate, on average, the observations
Good (effect 2)	$\beta_1 \neq 1$	$\beta_0 = 0$	The difference between model predictions and observation is proportional to the values predicted.
Fair	$\beta_1 \neq 1$	$\beta_0 \neq 0$	Effects 1 and 2 combined
Poor	$\beta_1 = 0$		There is no relationship between model and observations

2.3 The quantile regression

The second method used here has its origin on Cade et al. (1999), which was basead on the principle that the response of species to some limiting factors (such as habitat conditions) is greatly influenced by some unmeasured limiting factors (such as weather or disease). In fact, changes in biomass will not exceed the limits imposed by habitat conditions, but can be reduced by non habitat factors. We assume here that the CSTT model is able to predict the maximum concentration of chlorophyl with simple assumptions or processes built in, but that can be reduced by the effect of some other *unmeasured limiting factors*, assumptions or processes not taken into account in the simple screening CSTT model. Therefore, when displaying the model prediction (X) in the x-axis versus observed values (Y) in the y-axis, the points will not be distributed around the bisection of the first quadrant. On the contrary, the data will be distributed within a triangle, where the upper limit will be the hypothenuse and the adjacent cathetus the x-axis. The position of the hypothenuse will be defined by the points distribution and the number of points we want to remain above the line. This number of points is defined with by certain percentage or quantile (τ) in the quantile regression Cade et al. (1999) by the user. We have defined a set of new criteria to assess the model when the quantile regression is used (table 2). The criteria are based on the value of the slope of the quantile regression $\tau = 0.95$, and the quantile of the slope which is equal to 1.

Table 2: Criteria established to classify the model predictions (X) in different grades when compared with data (Y). Four groups are formed according to the estimates of the slope (β_1) and the quantile (τ) from the regression line of the form $Y = \beta_{1,\tau=i}X + \epsilon$ (being ϵ the error term of the regression model).

Model category	Slope	Interpretation
Excellent	$\beta_{1,\tau=0.95} = 1$	The model is regarded as perfectly simulate the worst case scenario for predicted chlorophyll.Worst case scenario is higher than observed values
Good	$\beta_{1,\tau=0.95} < 1$	The predicted maximum is is below the perfect agreement. Worst case scenario is higher than observed values
Fair	$\beta_{1,\tau<0.95} = 1$	The predicted maximum is above the perfect agreement. Worst case scenario is below observed values
Poor	$\beta_{1,\tau=0.95} = 0$	There is no relationship between model and observations

3 Results

We assessed the CSTT model for Loch Creran using the (OLS) and the quantile regression for all the data together (table 3). Figure 1 compares the time series of simulations with observations of Chlorohyll, in the upper layer of Loch Creran for 1975 and 2003. Figure 2) compares simulations forced with data from all the locations ad years together for Chlorophyll.

The assessment of the model in Loch Creran in 1975 and 2003 separately produces poor results (figure 1) obtaining small r^2 ($r^2 < 5$). Moreover the test for the slope and intercept for the OLS shows that the CSTT model fail to meet any of the quality test set in table 1. For the two years the model overpredicted chlorophyll concentration in Loch Creran.

When all the data is clumped together and the quantile regression is applied, the CSTT model yield better results (table 3). The plot of Maximum predicted values against observed (figure) shows the majority of the data distributed below the one to one relationship (bisection of the first quadrant). Only two points in loch Creran in 1975 and half of the OAERRE sites are distributed above the one to one line.

We were able to find a regression line with the quantile regression following the criteria of 95 % ($\tau = 0.95$) of the data above that line. Although the line that satisfy $\tau = 0.95$ is statistical significant different from 0, it we also found differences from the 1 to 1 relationship (table 3). Therefore, following the criteria set in table 2, the model was classifed as good.

4 Discussion

We tried two regression methods for comparing observations with simulations for the assessment of goodness–of–fit of the CSTT model. The CSTT model performed badly with the standard methodology for assessing models in ECASA, the Ordinary Least Squares (OLS). The main reason for the model failure is the fact that the model is not intending to reproduce the exact concentration of chlorophyl at a certain time, but to predict the worst case scenario with given nutrient conditions Tett et al. (2003). It is clear therefore, that the relationship expected between model predictions and observed field data will never follow a straight line.

The Quantile regression is a more suitable to test the performance of the CSTT. The performance of the CSTT model was shown to be good, with the 95% of the data being above a certain line which was not too far from the 1 to 1 relationship, but which was statistically different from 0. Actually the regression line for $\tau = 0.975$ was not significantly different from 1.

The behaviour observed is caused by the principles discussed above for CSTT are based. The expected distribution of the data in plot of observed vs predicted is the observed values to be always below the predicted concentration of the CSTT. We chose the Quantile regression because we could define the criteria of the percentage of the data above what we consider to be the main response of the model. The line will not be anymore a line estimated from the OLS but a line that meets the quality criteria of a certain percentage of situations where the CSTT model is not able to predict the worst case scenario.

The CSTT model was evaluated following the standard procedure proposed for all the ECASA models. However, due to its nature, we recommend to retain the reliability information yield by the quantile regression and not the Ordinary Least Squares. Some of the cells of the model cannot be filled because of the new methodology propossed here. more precisely the model was forced to pass throught the origin, therefore, the inference for the intercept of the regression line is lived blank in the tables of apendix A.

We finally encourage other possible researchers to apply a similar method to assess the reliability of their models. The quantile rergession will be useful when the main porpoise of the model predictions is not to predict the reality but a certain threshold, which is likely or unlikely to be passed.

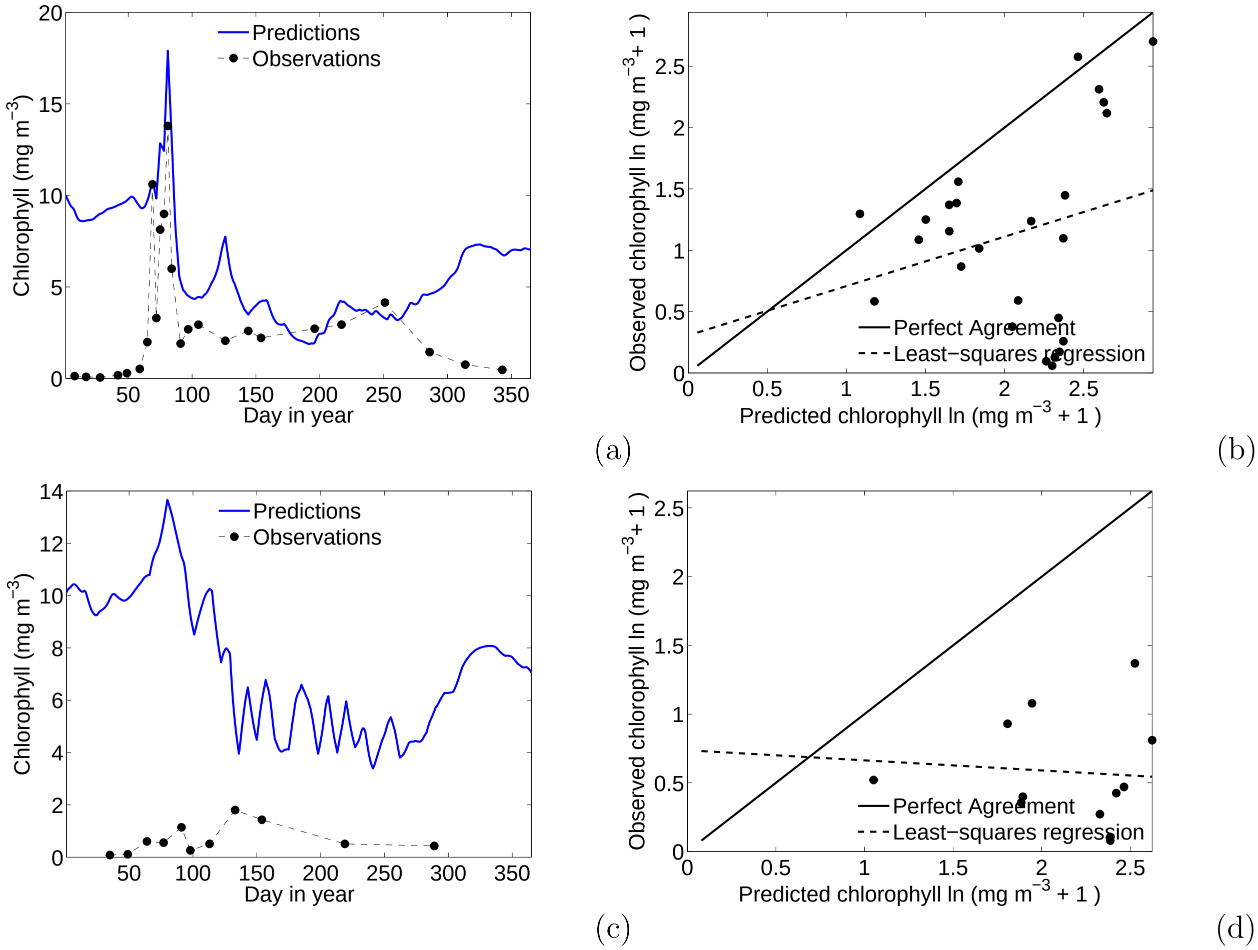

Figure 1: Test of the CSTT model for 1975 (a and b) and 2003(c and d). The left column (a and d) contains the time series of observed values (black) and simulated values (blue) for total chlorophyll. The right column contains plots of simulated values on the x-axis versus observed median values on the y-axis for total chlorophyll (b and c). The chlorophyll data are given as $\ln(\text{mg m}^{-3} + 1)$. The continuous line is the one to one relationship expected in the case of a perfect fit. The broken black line is the relationship found by OLS regression.

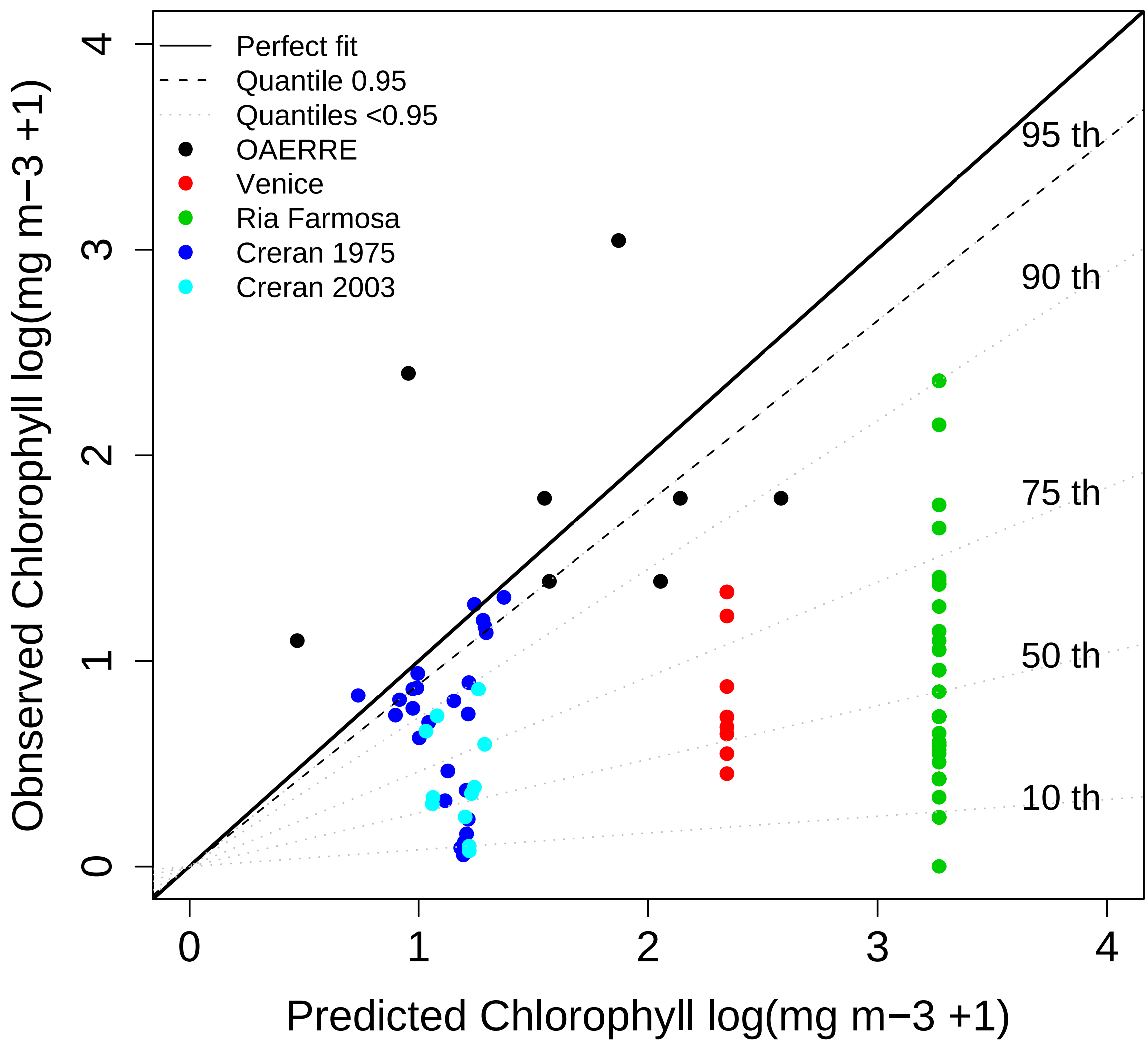

Figure 2: test for the CSTT model using quantile regression. The porpoise of this regression is to find a line that leaves a certain percentage of points underneath the regression line.

Table 3: Statistical results from the assessment of goodness of fit of CSTT simulations against 1975 and 2003 in Loch Creran and the CSTT with all the data together. data. The results shown are the slope (β_1) and intercept (β_0) from the OLS line-fits. The lower and upper values are 95% confidence limits to the estimates. We look for the confidence limits of the intercept to include zero, and the confidence limits of the slope to include 1, in order to say a model fits the data well. for the case of the Quantile regression the tested performed is for the slope of the quantile regression of $\tau = 0.95$ is different from 1. Again, the lower and upper values are 95% confidence limits to the estimates. Model category (Table 1) for CSTT model. The CSTT model was tested under three different situations.

	OLS-1975			OLS-2003			Quantile Regression		
	Lowr	Est	Upr	Lowr	Est	Upr	Lowr	Est	Upr
CHL									
β_0	-1.002	0.305	1.613	-1.196	0.737	2.67	—-	—-	—-
β_1	-0.206	0.402	1.01	-0.929	-0.073	0.782	0.778	0.885	0.993
Model Class									
		Poor			Poor			Good	

* Zero is included in the interval, so no relationship between simulations and observations

References

Cade, B., Terrell, J., Schroeder, R., 1999. Estimating Effects of Limiting Factors with Regression Quantiles. Ecology 80 (1), 311–323. 1, 4

CSTT, 1994. Comprehensive studies for the purposes of article 6 of dir 91/271 eec,the urban waste water treatment directive. report published for the comprehensive studies task team of group coordinating sea disposal monitoring by the forth river purification board. Tech. rep., Edinburgh, Scotland. 1, 2

CSTT, 1997. Comprehensive studies for the purposes of article 6 of dir 91/271 eec,the urban waste water treatment directive. report published for the comprehensive studies task team of group coordinating sea disposal monitoring by the forth river purification board. Tech. rep., Edinburgh, Scotland. 1, 2

Edwards, V. R., Icely, J., Newton, A., Webster, R., 2005. The yield of chlorophyll from nitrogen: a comparison between the shallow ria formosa lagoon and the deep oceanic conditions at sagres along the southern coast of portugal. Est.Coast.Shelf Sci. 62, 391–403. 3

Gillibrand, P. A., Inall, M. E., March 2006. Improving assimilative capacity modelling for scottish coastal waters: I. a model of physical exchange in scottish sea lochs. Tech. Rep. 167, Scottish Association for Marine Science, Dunstaffnage Marine Laboratory, Oban. 2

Jones, K. J., 1979. Studies on nutrient levels and phytoplankton growth in a scottish sea loch. Phd, University of Strathclyde. 3

Laurent, C., Tett, P., Fernandes, T., Gilpin, L., Jones, K. J., 2006. A dynamic cstt model for the effects of added nutrients in loch creran, a shallow fjord. J. Marine. Syst. 61 (3-4), 149–164. 3

Mesple, F., Troussellier, M., Casellas, C., Legendre, P., 1996. Evaluation of simple statistical criteria to qualify a simulation. Ecol. model. 88 (1), 9–18. 1

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263, 641–646. 2

Portilla, E., Tett, P., 2006. Report on the seasonal equilibrium concentration enhacement (sece) model. Tech. rep., School of Life sciecnes, Napier University, Edinburgh. 2, 3

Tett, P., Gilpin, L., Svendsen, H., Erlandsson, C., Larsson, U., Kratzer, S., Fouilland, E., Janzen, C., Lee, J., Grenz, C., et al., 2003. Eutrophication and some European waters of restricted exchange. Cont. Shelf Res. 23 (17/19), 1635–1671. 2, 3, 6

APPENDIX

A Tables

Table 4:

Model Description			
Model name		CSTT	
State variable		Chlorophyll-a	
site at which tested		Loch Creran 1975	
n, number of independent observations used in test		25	
Model Performance			
r^2, % of variance	05.29	p, on null hypothesis	$p = 0.268$
$\hat{\beta}_0$, regression intercept	0.31	$se_{\hat{\beta}_0}$	0.763
$t = (\hat{\beta}_0 - 0)/se_{\beta_0}$	0.40	p	0.692
$\hat{\beta}_1$, regression slope	0.40	$se_{\hat{\beta}_1}$	0.355
$t = (\hat{\beta}_1 - 1)/se_{\beta_1}$	-1.648	p	$p = 0.106$
Model Conclusion †			
Model explaint a signifcant part of variance in observations		NO	
Model reliability Class		4 Poor	

† Delete what does not apply

Table 5:

Model Description			
Model name		CSTT	
State variable		Chlorophyll-a	
site at which tested		Loch Creran 2003	
n, number of independent observations used in test		11	
Model Performance			
r^2, % of variance	0.27	p, on null hypothesis	$p = 0.879$
$\hat{\beta}_0$, regression intercept	0.74	$se_{\hat{\beta}_0}$	1.054
$t = (\hat{\beta}_0 - 0)/se_{\beta_0}$	0.70	p	0.492
$\hat{\beta}_1$, regression slope	-0.07	$se_{\hat{\beta}_1}$	0.492
$t = (\hat{\beta}_1 - 1)/se_{\beta_1}$	-2.30	p	$p = 0.031$
Model Conclusion †			
Model explaint a signifcant part of variance in observations		NO	
Model reliability Class		4 Poor	

† Delete what does not apply

Table 6:

Model Description			
Model name		CSTT	
State variable		Chlorophyll-a	
site at where tested		Loch Creran, OAERRE sites Venice Lagoon, Ria Farmosa	
n, number of independent observations used in test		103	
Model Performance			
r^2, % of variance	—-	p, on null hypothesis	$p < 0.001$
$\hat{\beta}_0$, regression intercept	—	$se_{\hat{\beta}_0}$	—-
$t = (\hat{\beta}_0 - 0)/se_{\beta_0}$	—-	p	—-
$\hat{\beta}_1$, regression slope	0.885	$se_{\hat{\beta}_1}$	0.065
$t = (\hat{\beta}_1 - 1)/se_{\beta_1}$	-2.284	p	$p = 0.013$
Model Conclusion †			
Model explaint a signifcant part of variance in observations		NO	
Model reliability Class		2 GOOD	

† Delete what does not apply