

Different Approaches to Efficiency Analysis. An application to the Spanish Trawl Fleet Operating in Moroccan Waters.

Ines Herrero*

Dep. Economía y Empresa. Universidad Pablo de Olavide

Ctra. de Utrera Km.1, 41013, Sevilla, Spain.

Phone no.: +34 954 349 906. Fax no.: +34 954 349 339

email: ihercha@upo.es

Abstract.

Several approaches exist for the estimation of technical efficiency. In this paper the four most relevant ones (Data Envelopment Analysis, Stochastic Production Frontiers, Panel Data and Distance Functions) have been compared. The different approaches have been applied to the Spanish Trawl fishery that operated in Moroccan waters. The study highlights some of the problems arising from the wrong management of the fishery. It can also be concluded that in most cases, the multi vs single output feature is determinant in producing higher differences in the efficiency estimates, though the effect of either ignoring or taking into account random error also influences the results.

Keywords: Data Envelopment Analysis, technical efficiency, fisheries, Stochastic Frontiers, Distance Functions, Panel Data.

* The study has been carried out with the financial support of the Commission of the European Communities Fifth Framework programme, QLK5-CT1999-01295, "Technical efficiency in EU fisheries: implications for monitoring and management through effort controls" and the support of the Centro de Estudios Andaluces, CentrA. The author thanks the people in the register of the fish market of Huelva for their cooperation during the collection of the data and three anonymous referees.

1. Introduction.

Several approaches can be found in the literature for measuring technical efficiency of firms. The most traditional ones in the stochastic framework are the Panel Data (PD) approach and the Stochastic Production Frontier (SPF) approach, whereas Data Envelopment Analysis (DEA) is the most common among deterministic ones. Another more recent technique is the multi-output stochastic approach, known as stochastic distance functions (DF). The DF approach benefits from allowing both random error in the analysis and multiple outputs.

The use of panel data (PD) techniques is the most traditional approach. The most relevant difference from the stochastic production frontiers (SPF) approach is the estimation procedure. Whereas panel data techniques estimate the inefficiency component by the use of least squares, the stochastic production frontier approach assumes a certain distribution for the inefficiency term and the estimation procedure is based on maximum likelihood. The SPF and PD approaches can only be used in single output production processes or in multi-output cases when the aggregation of all outputs into one is reasonable. On the contrary, the DEA methodology allows the inclusion of multiple outputs in the analysis but it presents the disadvantage of being deterministic.

The more recent multi-output stochastic approach, known as stochastic distance functions (DF), has the advantage of including random error and, at the same time, it possesses the desirable characteristic of including multiple outputs in the analysis. However, it requires restrictive assumptions (like the property of homogeneity of degree one for the outputs). Moreover, if there is evidence that a certain distribution can be assumed for the inefficiency term, then, the SPF approach provides better estimates.

The stochastic approaches present two important disadvantages when compared with DEA: 1) a functional form regarding the production process has to be imposed and 2) a distributional assumption has to be made for the error term. Furthermore, in the case of SPF and DF a distributional assumption has also to be made for the inefficiency term.

In general terms, it cannot be said that any of the abovementioned methodologies is better than the rest. The most adequate methodology to be used in each study depends on the characteristics of the production process, the degree of stochasticity, number of outputs and possibility of aggregation.

In many recent papers, efficiency techniques are used and applied to different fields. Efficiency has been applied not just to measure efficiency itself but also for other purposes like capacity utilisation (Vestergaad et al., 2002, Pascoe et al., 2001), risk analysis (Herrero, 2004), etc. The main goal of this work is to provide a summary of the main techniques for the estimation of technical efficiency and to make a comparison among them when applied to a Spanish fishery. The different efficiency estimation approaches have been applied to the Spanish trawl fishery that operated in Moroccan waters and some interesting conclusions has been drawn from the analysis.

Three different software packages have been used: NLOGIT8.0 (Greene, 2002) for panel data estimators, GAMS for DEA (Brooke *et al.*, 1992) and FRONTIER (Coelli, 1991) for stochastic production functions and for distance functions.

The paper is structured as follows: in section 2 a brief description of the fishery has been provided. Section 3 is dedicated to the analysis of the variables included in the study. Sections 4, 5, 6 and 7 summarised the SPF, PD, DF and DEA techniques (respectively) and presents the results when applied to the case study. The empirical results of the different approaches have been compared in section 8. Finally the main conclusions of the study are discussed in the last section.

2. Brief description of the fishery

The study is based on a sample of 26 vessels from the Spanish Andalusian fleet that operated during 1993-1998. In the years under analysis, a close season was imposed during January and February¹, so that the database is restricted to the period March-December each year. The fleet

¹ This information has been obtained from the fishermen's guilds

consists of deep-water trawl vessels that are based in the Port of Huelva and fish in Moroccan Waters. The target species are crustaceans (deep water rose shrimp, prawns, scarlet shrimp...) and other valuable fish (like European Hake). As the species are all bottom dwelling species they are normally fished together. However, the composition of the catch is not always homogeneous, as the proportion of each species is different by area fished (Garcia and Herrero, 1998) or it may be altered by the effect of the discard of low value species.

The data on individual landings from 1993 to 1998 was collected from the records of the fish auction market in Huelva. The information had to be computerised as it was obtained from written documents. Official Institutions facilitated most of the physical characteristics of the vessels. This information was aggregated into monthly observations and the number of trips per month was considered a new variable. The vessels that operated in this fishery are relatively large (table 1) as they have to navigate from Huelva to Moroccan Waters for fishing. Vessels stay out in the sea for around 12 days.

This fishery had always operated under the different agreements between Morocco and Spain (or, later in time, between Morocco and the EU). The last agreement finished in November 1999 and despite all efforts, the activity of this Spanish fleet in Morocco was ended.

Table 1. Technical characteristics of the fleet (Trawl fishery)

The management of this fishery has always been affected by the different agreements and it has been hardened over time. The regulations have also been more and more complex². Regarding the physical characteristics of the vessels, the different agreements have restricted the total volume of the fleet. Hence, depending on the agreement, there was a limit on the total Gross Registered Tons

² These regulations imply some traditional restrictions (like restricting the size of the mesh) as well as other particular rules like having to land fish in Moroccan ports a certain number of times per month, or having to contract a certain proportion of crew of Moroccan nationality.

(GRT) per year for the whole Spanish trawl fleet operating in Morocco³ (table 2). Furthermore, taxes were imposed per GRT (Barroso, 1998). Contrarily, the engine power of the vessels was never restricted. This type of regulation brought about a decrease in the average volume (partly forced by the management system) and a steady increase of the average engine power of the fleet over time.

Table 2. Maximum volume allowed for the Spanish North Trawl fishery in Moroccan Waters.

3. Variables included in the analysis.

The dependent variables used in all approaches were the total value of the catch of the species over the month. Some aggregation had to be done due to the high number of different species in the catch. As a reasonable number of the most relevant ones could not be found (the number of the most valuable species was too high) they were aggregated into two, crustaceans and finfish, the former being of a much higher price than the latter. For the multi output approaches, the value of these two types of different species was considered, whereas for the single-output approaches, the value of all species was aggregated into the total value. This may produce some differences in the results. However, as one of the key advantages of the multi-output techniques is precisely the fact of being able to include more than one output, this option has been preferred.

In both cases, this measure is effectively an aggregated index of output, with each component of the catch weighted by its price (reflecting its relative importance to the overall objective of profit maximisation). Other studies on this fishery (see Herrero and Pascoe, 2003) compared the use of different output measures (value and weight of the catch) and the results suggest that value is the best measure for the study of production functions though the use of both measures simultaneously can be used to study fishers' behaviour.

³ From 1995, the North Deep Trawl fishery was distinguished from other types of trawl fisheries also operating in Morocco (like the one dedicated exclusively to European Hake). Therefore, in this table the data before 1995 have not been compared with the data from 1995 onwards.

A common problem in fisheries is the lack of any indicator of the stock biomass. Some authors have recommended the use of catch (or value) per unit of effort (CPUE) as an indicator of the stock (Van Hoof, L., and Salz, P., 2001). However, this approach presents an endogeneity problem. Cross sectional models or monthly dummies did not seem a feasible option either due to the high number of months considered (47).

Following Pascoe and Herrero (2001), the output measure (i.e. value) was standardised by a seasonal index to avoid the problem of not having an indicator of stock. This indicator is similar to Malmquist indexes as it is based on shifts in the production frontier over time. The Pascoe and Herrero (2001) stock index has been shown to be more appropriate for accounting for aggregate stock changes in production frontiers for multi-species fisheries. The measure is a "seasonal and stock effect" rather than a stock index *per se*, and an individual measure is derived that relates to individual vessels, taking into account their size and activity levels. The measure is used to normalise the output data rather than act as a separate variable in the analysis.

The index was estimated in terms of value. This index measures how some temporal changes have affected the revenue of each boat given the level of fishing power at which the boat is operating. These changes can be of very different nature, like changes in stock, composition of the catch, market conditions (like price or demand) and other temporal variables, although the most influential ones are more likely to be prices and stock. Therefore, this is just an *individual* indicator of the effect of the seasonal conditions as the measure is more an *individual* index of by how much these changes have affected the output of *each* of the different boats given the scale at which they are operating.

In order not to lose the "individual information" included in the index (of how the stock and temporal changes affects each observation given the scale at which it is operating), the indexes were not averaged and used as an input, but were used to normalise the output measure that is subsequently used in the efficiency analyses. Note that this index is different for each observation, as the effect includes the interaction of the changes with the other physical and effort variables. A more detailed explanation on this index can be found in Pascoe and Herrero (2001) and an application can be found in Herrero and Pascoe (2003).

The available inputs were the size of the vessels (represented by its volume measured by its gross registered tons, or GRT), the engine power of the vessels (in horsepower), and the number of trips each month. Other physical variables, like hold size were not available, but they are probably highly correlated with volume.

The trawling activity is highly machine-orientated (unlike other types of fishing). Therefore, it may not be highly dependent on the number of crew members. While some authors (e.g. Pascoe, Andersen and de Wilde 2001, Pascoe and Coglán 2002) have proved that for some fisheries the number of crew members is highly correlated with the vessel size, others (Herrero, 2004) have shown that this variable is not relevant in the production process of some fisheries. Therefore, the no availability of the number of crew is not likely to represent a problem.

Given the correlation between the variables (bigger boats use bigger engines), the risk of multicollinearity existed and the estimated individual coefficients could be unreliable (a problem associated with multicollinearity), although the overall predictiveness of the model is most likely reasonable. The correlations of both volume and engine power with the number of trips were approximately equal to zero (0.16 and 0.06, respectively). On the contrary, a certain degree of correlation between volume and engine power (equal to 0.68) seemed to exist. However, as it was not very high the use of formal techniques to detect the existence of multicollinearity was applied.

There are several ways to detect the existence of multicollinearity (Mason et al, 1975, Gunst and Mason, 1977). One of the most common methods (Gujarati, 2000) is based on the size (the determinant) of the matrix $X'X$ (where X , as usual, is composed of the different observations). If there exists perfect multicollinearity, this matrix is singular whereas if there is some approximated multicollinearity, then this matrix is “nearly singular”. As $X'X$ is symmetrical, its determinant is related to its eigenvalues.

To avoid problems related with the units of measure, the observations were normalised by either the standard deviation or by $\sqrt{\sum (x_i')^2}$. The square root of the ratio of the maximum over the

minimum eigenvalue of the matrix $X'X$ is called *the condition number of matrix X*. If this number is greater than 20 (Greene, 1999) then the existence of multicollinearity cannot be rejected (though some authors do not consider multicollinearity to be a serious problem if this value is not higher than 25).

The condition number for the observations in this study was found to be equal to 28.7. This value is above the limit considered in most literature. Hence, there could be a problem of multicollinearity.

Several approaches appear in the literature to deal with the problem of multi-collinearity. The most common one is to avoid one (or more) of the variables that produces the multicollinearity so that only one of the correlated variables is left in the model. Then, different analyses can be carried out, each of them using non-correlated variables. This may produce a miss-specification problem. However, this option was preferred in this study in order to estimate, within each model, the effect of the volume and engine power of the vessels on the catch (though none of the models can include the interaction of both variables).

Therefore, in this study two different models were considered. In the first one, the variable for volume was excluded (so that engine power is representative of the physical characteristic of the vessels) and in the second one the engine power was not included in the analysis. Hence in the first model only two inputs were considered (engine power and number of trips). In the second model, the two inputs were volume and number of trips.

All the methodologies were carried out using the same database and the same number of observations, so that results could be compared more consistently. The only difference was that for the single output methodologies, the outputs had to be aggregated. While this may result in some differences among single and multi output approaches, the effect of this on the results was one of the objectives of the present study.

4. Stochastic production frontiers models.

In this section, the SPF approach was used to estimate the relationship between the level of inputs and outputs, and the distribution of technical efficiency in the fleet.

A translog production function was used in the analysis. This is a flexible functional form in which is nested the Cobb-Douglas and CES production functions. The translog production function is given by

$$\ln Y_{it} = \beta_0 + \sum_k \beta_k \ln X_{ikt} + \frac{1}{2} \sum_k \sum_j \beta_{kj} \ln X_{ikt} \ln X_{ijt} + V_{it} - U_{it}$$

where Y_{it} is the standardised output of unit i in time t , X_{ikt} and X_{ijt} are the inputs associated to that unit in the production process, and V_{it} and U_{it} represent the joint error component that consists of the random error and inefficiency terms respectively. The measure of technical efficiency of firm i (TE_i) when working with logged variables is given by (Coelli *et al.*, 1998):

$$TE_i = e^{-U_i}.$$

The error term, V_{it} is assumed to be independently and identically distributed (iid) $N(0, \sigma_v^2)$. In order to separate the stochastic and inefficiency effects in the model, a distributional assumption has to be made for U_{it} . Two main distributional assumptions have been proposed: a normal distribution truncated at zero, $U_{it} \sim |N(\mu, \sigma_u^2)|$ (Aigner, Lovell and Schmidt 1977) and a half-normal distribution truncated at zero, $U_{it} \sim |N(0, \sigma_u^2)|$ (Jondrow *et al.* 1982). In addition, the inefficiency can also be considered to have a time invariant component, such that $U_{i,t} = U_i \exp[\eta(T-t)]$ (Battese and Coelli 1992), where T is the terminal time period (i.e. $U_{i,t} = U_i$ when $t=T$).

For the initial estimation of the model, a truncated normal distribution was assumed for the inefficiency term (as in Battese and Coelli 1992), such that U_{it} are non-negative iid truncated normal random variables of the distribution $N(\mu_i, \sigma_u^2)$ of the form $U_{i,t} = U_i e^{-\eta(t-T)}$. A series of

tests of the functional form and the assumptions regarding the structure of the error components (i.e. random error and inefficiency) were subsequently undertaken (Herrero and Pascoe, 2003).

As was stated before, two different models were considered. In the first one, the variable for volume was excluded and in the second one the engine power was not included in the analysis. In both cases the output considered was the value of the catch (standardised by the seasonal effects). The value of all variables was normalised using its mean value such that the mean logged value of the variables was zero (see Huang and Liu 1994)). The normalisation enables the elasticities associated with each of the inputs to be derived directly from the coefficient relating to their levels in the stochastic production frontier.

Several log-likelihood ratio tests were carried out to test for the functional forms of the model⁴. In the model omitting volume the null hypothesis of the Cobb-Douglas functional form⁵ could not be rejected whereas in the model omitting engine power this hypothesis had to be rejected. The half-normal distribution (i.e., $\mu=0$) was also tested and rejected both models. Similarly the hypotheses of time-invariant inefficiency ($\eta=0$) and the absence of inefficiency effects ($\gamma=0$) were also tested and rejected. Some of these results are shown in Table 3.

The results of both analyses were found to be relatively similar, suggesting that both variables can fairly equally explain the dependent variable. However, the model omitting the engine power of the vessels had more significant variables. Therefore, only the results of this model have been provided in Table 4.

Table 3. Results of the likelihood ratio tests on the SPF model (omitting engine power).

Table 4. Estimated SPF model (omitting engine power).

⁴ The likelihood ratio tests are based on the ratio of the likelihood associated to the null hypothesis and the likelihood associated to the alternative hypothesis.

⁵ This implies that the cross-product coefficients ($\beta_{ij}=0$) are all equal to zero

5. Panel data models.

What has been called the “panel data technique” (PD) is a parametric procedure that is commonly used when the number of firms involved in the analysis is bigger than the number of time periods (Hsiao, 1986). Consider a standard SPF of the form:

$$Y_{it} = \beta \mathbf{x}_{it} + V_{it} - U_{it}$$

where, the same notation as in the SPF model has been used. As before, V_{it} are the random errors, assumed to be independently and identically distributed (iid) $N(0, \sigma_v^2)$; and U_{it} are constant firm-specific variables, accounting for inefficiency. If efficiency for a given unit is assumed to remain invariant over time, the model can be transformed into,

$$Y_{it} = \alpha_i + \beta \mathbf{x}_{it} + V_{it}$$

where α_i stands for the firm’s specific effect. The model above is of the form of the traditional fixed effects panel data model (Hallam and Machado, 1996). The parameters can be estimated using Ordinary Least Squares (OLS).

A measure of inefficiency similar to the stochastic production frontier measure can be estimated by carrying out the transformation:

$$u_i = \max \{ \alpha_j \} - \alpha_i$$

The term u_i will account for the inefficiency, having a value of zero for efficient units and a positive value for inefficient ones.

If the data is logged, then the inefficiency estimates are given by:

$$TE_i = \exp(-u_i)$$

Several approaches can be found in the literature for the estimation of fixed-effects efficiency indicators (for example Corrected Ordinary Least Squares, COLS⁶). A modified method to the standard COLS was proposed by Alvarez-Pinilla and González (1999).

⁶ See for example Greene, 1993.

However, these time invariant constant effects are hard to maintain in a real case. In addition, the firm effects may not be correlated to the explanatory variables. In this case the firm effects can be included as part of the random error terms. Hence, a more realistic approach, but computationally more difficult, is what is known as the random effects model. This model allows for inefficiency to slightly vary over time; in this context, the inefficiency terms, u_i , are non-negative iid random variables having variance σ_i^2 . The standard random effects model is given by:

$$Y_{i,t} = \mu + x'_{i,t}\beta + \alpha_i + v_{i,t}$$

Different approaches can be found to solve this model. If there is no correlation between individual effects and the explanatory variables then the Balestra-Nerlove estimator (Balestra and Nerlove, 1966) obtained from a generalised least squares procedure seems to be the best option.

The Hausman (1978) test -based on the Wald test- can be used to identify which of the two main models -fixed or random effects- is more appropriate for each particular situation (Hsiao, 1986, Judge et al., 1980, Alvarez-Pinilla *et al.*, 2001). In this case, given that some inputs vary across firms but not over time, the most appropriate model is the random effects model, which can be estimated using the Generalised Least Squares procedure. The Breusch and Pagan's Lagrangian multiplier statistic was used to test the absence of inefficiency effects.

These models have been applied to fisheries studies; Squires and Kirkley (1999) used a fixed and a random effect model concluding that the latter one was more appropriate for the trawl fishery in the Pacific Coast. They also concluded that skipper skill came out to be more important for the fishery activity than the input mix employed in it.

As before, two different models of a translog production function were used. The standardised value of the catch was the output of the model as in the previous approach. Regarding the productive inputs, the number of trips per month was used together with either the engine power

or the volume of the vessels. As before, results of the estimation model when omitting engine power have been presented in table 5.

Table 5. Panel data model (omitting engine power).

The Breusch and Pagan's Lagrangian multiplier statistic was used to test the absence of inefficiency effects. The t-statistic was equal to 117.09 with a p-value equal to 0.00. Hence, the null hypothesis of no inefficiency effects was rejected.

6. Distance functions approach.

The distance functions (DF) approach is a multi-output stochastic methodology for the estimation of efficiency. It is similar to the stochastic production function approach but it includes multiple outputs, one of which plays the role of the "dependent" output (Coelli et al, 1998, Kumbhakar and Lovell, 2000).

The output distance function is defined as (Coelli and Perelman, 2000):

$$D_o(x, y) = \min\{\delta : (y/\delta) \in P(x)\}$$

where $P(x)$ is the output production possibility set:

$$P(x) = \{y : x \text{ can produce } y\}$$

which is assumed to verified the usual axioms (null production, impossibility of producing outputs without using any amount of inputs and strong disposability of outputs).

A unit is said to be efficient if $D_o(x, y) = 1$ and inefficient if $D_o(x, y) \leq 1$.

It was used a translog production distance function of the form:

$$D_o(x_{it}, y_{it}) = \alpha_0 + \sum_m \alpha_m \ln x_{mit} + 0.5 \sum_k \sum_m \alpha_{km} \ln x_{kit} \ln x_{mit} + \sum_n \beta_n \ln y_{nit} + \\ + 0.5 \sum_n \sum_l \beta_{nl} \ln y_{nit} \ln y_{lit} + 0.5 \sum_m \sum_n \delta_{mn} \ln x_{mi} \ln y_{nit} + u_{it}$$

where the usual restrictions of symmetry ($\alpha_{km} = \alpha_{mk}$ and $\beta_{nl} = \beta_{ln}$) and homogeneity of degree one in outputs have been imposed:

$$\sum_n \beta_n = 1; \quad \sum_n \beta_{nl} = 0, \forall l; \quad \sum_n \delta_{mn} = 0, \forall m$$

For the estimation of the above model one of the outputs, y_{1i} , had to be chosen arbitrarily as the dependent output and the others had to be standardised by the “dependent” output, i.e.,

$$y^*_{ni} = y_{ni} / y_{1i}$$

The value of crustaceans (standardised by the individual stock index) were considered as the “dependent output” and the value of other fish (also standardised by the individual stock index) were considered as the independent output. All variables were subsequently standardised by their average values. The estimates of the model depend on the selection of the “dependent” output which is one of the major criticisms of this approach.

As distance functions verify homogeneity of degree one in outputs, the model verify:

$$D_o(x_{it}, \lambda y_{it}) = \lambda D_o(x_{it}, y_{it})$$

Hence, if outputs are standardised by output 1 and the above equation is re-arranged, the standard form of a distance function model results:

$$\begin{aligned} \ln y_{it} = & \alpha_0 + \sum_m \alpha_m \ln x_{mit} + 0.5 \sum_k \sum_m \alpha_{km} \ln x_{kit} \ln x_{mit} + \sum_n \beta_n \ln y^*_{nit} + 0.5 \sum_n \sum_l \beta_{nl} \ln y^*_{nit} \ln y^*_{lit} + \\ & + 0.5 \sum_m \sum_n \delta_{mn} \ln x_{mi} \ln y^*_{nit} + v_{it} + u_{it} \end{aligned}$$

The above is the more common structure of an stochastic production frontier (which, in this case, includes the effect of the different outputs) where the terms of efficiency and random error appear. Note that y^*_{ni} , has now the asterisk superscript because it has been standardised by the dependent output, i.e., $y^*_{ni} = y_{ni} / y_{1i}$.

The models were estimated using the stochastic frontiers approach. A truncated normal distribution, $[N(v, \sigma_v^2)]$, was assumed for the inefficiency term, v_i , and a normal distribution, $N(0, \sigma_\mu^2)$ for the random error term, μ_i .

As before, in order to eliminate the effect of multicollinearity two separated models were run. One omitting the volume variable and another one omitting the engine power of the vessels.

Several tests were carried out to obtain a final specification for each of the models (see table 6 for the results when omitting the power of the engine variable). Separability among inputs and outputs were tested⁷. However, this assumption had to be rejected for both models. Hence the cross-products of the inputs with the dependent output do appear in the final models.

A test on homotheticity⁸ was also carried out but the likelihood ratio test was also rejected in both cases. Analogously, the hypothesis of linear homogeneity⁹ in inputs was rejected in both models using the likelihood ratio test.

Table 6. Results of the likelihood ratio tests on the distance functions model (omitting engine power).

The assumption of the time-invariant efficiency ($\eta=0$) could not be rejected in any model. While the half-normal distribution ($\mu=0$) was rejected for the model in which engine power was omitted, it could not be rejected in the case of the model in which the volume variable was not taken into account. Tests on the absence of the inefficiency effects¹⁰ were also carried out and rejected in both cases. The final model when omitting engine power has been presented in table 7.

⁷ Separability requires: $\delta_{mn} = 0, \quad \forall m, n$

⁸ Homotheticity implies the following set of restrictions: $\sum_m \alpha_{km} = 0, \quad \forall k$

⁹ Linear homogeneity in inputs (homogeneity of degree -1) implies setting the following three sets of restrictions: $\sum_m \alpha_{km} = 0, \quad \forall k; \sum_k \alpha_k = -1; \sum_m \delta_{mn} = 0, \quad \forall n$

¹⁰ The null hypothesis in this case is $\gamma=0$

Table 7. Estimated distance functions model (omitting engine power).

7. Data Envelopment Analysis approach.

A standard DEA analysis was also carried out. For consistency and for comparative purposes, a model omitting each of the physical characteristics of the vessels was also carried out, similarly to the other approaches applied in this work.

A standard BBC (Banker, Charnes and Cooper, 1984) output oriented model was used. The BCC output (input) oriented model seeks to maximise (minimise) the proportion of the outputs (inputs) of the firm under evaluation, i_0 , based on a weighted convex combination of the inputs and outputs of the other firms, which outperform firm i_0 -th. The formulation of the output oriented model is given by:

$$\begin{aligned}
 & \text{Max } \theta_0 \\
 & \text{s.t.} \\
 & Y \lambda - s^+ = \theta Y_0 \\
 & X \lambda + s^- = X_0 \\
 & \sum \lambda = 1 \\
 & \lambda, s^+, s^- \geq 0
 \end{aligned}$$

where $X=\{X_{jit}\}$ and $Y=\{Y_{rit}\}$ are the input and output matrixes respectively; X_{jit} being the amount of input j used by the i -th DMU in time t and Y_{rit} the amount of output r (standardised as in the previous approaches) produced by the i -th DMU in time t . X_{it} (Y_{it}) represents the $J \times 1$ ($R \times 1$) column input (output) vector corresponding to the j -th DMU in time t . The zero sub-index stands for the unit being analysed.

Theta (θ) is called the expansion ratio. Inefficient DMUs can be transformed into efficient ones by radially expanding their outputs (or contracting its inputs) by multiplying them by theta for output

(or input) orientation. All the outputs of the unit under analysis should be multiplied by this value to make the unit being analysed efficient.

The technical efficiency score for the unit under analysis is given by the inverse of the value of theta ($TE = 1/\theta$). This value lies between zero and one. A unit is said to be efficient if the optimal value of theta is equal to one.

If at the optimum, the value of one of the lambdas is equal to zero, it means that the observation of the reference set associated to that lambda is not a peer unit for the DMU under analysis. The peer units associated to the unit under analysis are those with a positive value of lambda and they are the units that compose the convex combination that would make the unit under analysis be efficient.

This model was preferred to a CCR (Charnes, Cooper and Rhodes, 1978) model as the latter imposes constant returns to scale whereas the above model does not impose any restriction on the returns to scale. The CCR model formulation is similar to the previous one but omitting the convexity restriction (sum of lambdas equal to one).

When the vessels were only compared to those operating in the same time period as the one under evaluation, the average value of the pure technical efficiency was very high. However, for consistency the output variable was also standardised (as in the previous approaches) and each vessel was compared with all the observations in the sample.

In comparison with the econometrics methods, multi-collinearity is not a big problem in DEA. This issue has been studied by some authors. While Nunamaker (1985) concluded that the existence of correlation among variables would produce an expansion of the efficient frontier, Ahn et al. () proved that his conclusions were wrong showing that the DEA approach was hardly affected by the effect of multicollinearity among variables. Saen *et al.* (2004) analysed the effect of omitting a certain number of correlated inputs in the different DEA models using simulated data. Smith (1997) also studied the relation between actual and estimated efficiency using simulated data under different levels of correlation among inputs and outputs. He draw different conclusions depending on the existence of extraneous variables.

In any case, for consistency in the comparison with the other approaches presented in this study, two models were also carried out, omitting in each of them one of the physical variables.

8. Results and comparison of approaches.

The results of the different approaches used in the analysis were compared. It can be observed (Tables 8a and 8b) that the efficiency estimates are highly correlated when using all approaches.

In general terms, the SPF and DF models when omitting the power of the engines seem to perform better than the models where volume is omitted, in the sense that the former have more significant variables (volume and fishing) than the latter (which is only based on the fishing effort regardless the physical characteristics of the vessels).

As the correlation between the volume variable and the engine power of the vessels is not very high, there were some differences in the estimation of both of the models where either volume or engine power were omitted.

The efficiency scores per boat resulting from the models where either volume or engine power were omitted were averaged. The average TE rate of all the approaches where volume was omitted (0.474) was lower than in the model where engine power was omitted (0.479). In the case of the DF approach, the maximum difference between the efficiency scores of both models was high (0.4) though they were highly correlated (0.84). This implies that the inefficiency seems to be more explained (mainly by this approach) by the volume variable than by the power of the engines. This is consistent with the importance of the coefficients associated to the physical inputs. Moreover, the efficiency scores per boat in the model that included both input variables¹¹ was more correlated to the model in which the engine power was omitted (0.80) than to the model in which the volume variable was omitted (0.60).

¹¹ Results on this model has not been presented to limit the paper to a reasonable size.

The averaged value of the models where one of the physical variables was omitted was found to be similar to the value obtained when the two variables were included in the analysis (with a correlation equal to 0.72). This is reasonable because even if the coefficient estimates are not reliable in the presence of multicollinearity, the estimates of efficiency are expected not to be affected by any distortion.

In any case, the differences existing among the TE scores per boat associated to each of the models indicate that even if a certain level of correlation between the physical variables exists, the correlation is not very high; otherwise, the similarities of the estimates would be more similar.

At present, I am studying other fisheries in which the multicollinearity is much higher (the condition number in that case is 38.7 while in this case is only 28.7) and where the correlation between the efficiency scores of the models omitting one of the physical variables is much higher (0.99) than in this case (0.84) and it results in the maximum difference between efficiency scores being much lower.

When the power of the engines is not considered, the highest correlated approach with distance functions is DEA (0.84), whereas the highest correlated approach with SPF is the PD approach (0.89) and vice-versa. These results suggest that, in this case, one of the most relevant characteristics of the approaches is not the fact of taking into account random variation but the fact of using a single or a multi-output approach.

On the other hand, the least correlated estimators to those obtained from DEA are those resulting from the SPF analysis (when omitting engine power). This result is most likely because they differ in two relevant features. Firstly because DEA does not take random error into account, whereas SPF does and, secondly because the former is a single approach and the latter is a and multi-output approach.

*Table 8a. Correlation among TE estimates from different approaches
(omitting engine power).*

On the contrary, when omitting volume, even if DEA is more correlated to DF than to any other approach, the reverse does not occur. DF estimates are more correlated to the other approaches than to DEA. In any case, the least correlated estimators to those obtained from DEA are those resulting from PD. As before, this may be explained by the fact that they differ in the two main features. DEA is a multi-output approach not considering random error, whereas panel data is a single output approach that takes random error into consideration. In this case, panel data and stochastic frontiers estimates are the highest correlated (as both are stochastic and single output approaches), with a correlation of 0.98 (see table 8b).

*Table 8b. Correlation among TE estimates from different approaches
(omitting volume)*

In general, the estimates of technical efficiency per boat resulting from all the approaches follow similar trends (see figures 1 and 2). On average, the highest TE scores are given by the PD estimates, whereas the lowest are produced by the DEA approach.

Figure 1. Comparison of results per boat (when omitting engine power)

Figure 2. Comparison of results per boat (when omitting volume)

In tables 9a and 9b, the averages and standard deviations of the efficiency scores have been presented for the models including engine power and the models including volume. On average, it seems that the vessels of this fleet are no more efficient when omitting engine power than when omitting volume. It only depends on the approach used, being the average efficiency score very similar if either volume or engine power is omitted.

When the power of the engine variable was omitted, the SPF average efficiency score is lower than the average DF estimate (Table 9b). The reverse occurs for models omitting volume.

The average value of the technical efficiency scores using the multi-output approaches (distance functions and the DEA) is higher for models omitting the power of the engines than for models omitting volume. On the contrary, the average scores resulting from the single output approaches are lower for the former than for the latter.

Table 9a. Averages TE scores for the different approaches (omitting engine power)

Table 9b. Averages TE scores for the different approaches (omitting volume)

Regarding the SPF model, results from a t-test for the hypothesis that the elasticity of the number of trips was equal to 1 could not be rejected in any of the two cases (being the t-statistic equal to 1.595 and 1.076 for the models omitting engine power and volume respectively). The elasticity of engine power was not significantly different from zero. On the contrary, the elasticity of the volume of the vessels (equal to 0.67) was significant at the 1% level.

Using data of the middle eighties, Garcia and Herrero, 1998 found that the elasticity of the engine power of the fleet was significant. Furthermore, it was higher than the elasticity associated to the volume variable. On the contrary, this study shows that for the nineties the elasticity for the engine power of the vessels was not significant whereas the elasticity for the volume is significant. A possible explanation could be that as Morocco has been imposing some restrictions and taxes on volume (leaving the engine power under no control) since the middle eighties, the fleet in the nineties is composed by smaller but more powerful vessels than it was in the eighties (Table 1). This means that input substitution may have occurred by fishers taking the option of installing larger engines in new vessels (as new vessels may have been forced to be smaller in volume).

At present, an increase in the power of the engines does not lead to a proportional increase in the output of the vessels¹². Therefore, the engine power may be under-utilised given that the productivity may be restricted by the volume of the vessels (and maybe by a decrease in the stock

¹² This conclusion is drawn by the non-significance of the elasticity of the engine power, which can be mistaken with a non-relationship between this input and the output of the fishing activity.

of the biomass in the last decade).

This finding is clear from the results of the panel data and the DF approaches. As with the SPF approach, the coefficient for the volume is highly significant in the PD and the DF models, whereas the coefficient (elasticity) for the power of the engines is not significant neither in the PD nor in the DF models. These results confirmed the idea that the power of the engines is under-utilised as its coefficient and all related coefficients again appeared as non-significant whereas the coefficient associated to the volume of the vessels is highly significant.

9. Conclusions.

In this paper some of the most relevant approaches that exist in literature for the estimation of technical efficiency have been compared and applied to analyse the efficiency of the vessels of the Spanish Trawl fleet that operated in Moroccan waters. While some of them allow the inclusion of multi-output measures, others are restricted to the case of single outputs. Stochastic approaches have also been compared with deterministic ones.

As would be expected in most cases, it can be concluded that the multi vs single output feature is determinant in producing differences in the efficiency estimates. Similarly, the effect of either ignoring or taking into account random error also influences the results. Overall, none of the methodologies cannot be said to be better than the rest. The most appropriated methodology depends on the characteristics of the production process, the degree of stochasticity, number of outputs and possibility of aggregation, etc.

From the results it follows that, in general terms, the efficiency estimates, in the present case study, are similar and highly correlated. However, in other case studies characterised by a higher amount of randomness or by a wider variety of outputs results could differ more among the different approaches than in the present study.

Literature regarding the stochasticity nature of different production processes shows different results. Therefore, the stochastic nature of the fishery being analysed should be examined and the most appropriate method should be applied in each case.

From the empirical results, it can also be deduced the importance that management systems control all inputs simultaneously in a fishery. Regulations on the fishery under analysis have been traditionally of high complexity. However, they just control two of the main inputs (time spent fishing and volume) whereas the engine power of the vessels is out of any control. As this was the only uncontrolled main input, input substitution occurred and the fleet increased this input to compensate the lack of volume.

The average power of the engines was very low in the eighties, but the steady increase in this input brought about an increase in the efficiency of the fleet (Garcia-Hoyo and Herrero, 1998). While regulations intended to control the already over-exploited fishing grounds, they got the reverse (i.e., regulations lead to an increase in the vessels efficiency and, as a consequence, an increase in the over-exploitation of the resource). However, from the results of this study, it seems that, years later, an over-dimension of the power of the engines has been produced. During the period analysed in this study, it seems that the power of the engines was being under-utilised probably because of three reasons: the increase in the average engine power, the restrictions imposed on the volume of the vessels and the decrease in the stock of the biomass. This finding for this fishery was firstly analysed by Herrero and Pascoe (2003).

References.

- Ahn, T., Charnes, A., Cooper, W.W., 1988. Using Data Envelopment Analysis to measure the efficiency of Not-for Profit Organizations: a critical evaluation- Comment". *Managerial and Decision Economics*, 9(3), pp. 251-253.
- Aigner D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6, pp. 21-37.
- Alvarez-Pinilla, A. (coordinator), 2001. *La medición de la eficiencia y la productividad* (in Spanish). Pirámide. ISBN 84-368-1586-6.
- Alvarez-Pinilla, A., González, E., 1999. Using Cross-sectional data to adjust technical efficiency indexes estimated with panel data. *American Journal of Agricultural Economics*, 81, pp. 894-901.
- Balestra, P., Nerlove, M., 1966. Pooling cross-section and time series data in the estimation of a Dynamic model: the demand for natural gas. *Econometrica*, 34, pp. 585-612.
- Banker, R.D., Charnes A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in DEA. *Management Science*, 30 (9), pp. 1078-1092.
- Barroso, D., 1998. *La Flota Pesquera Onubense en Marruecos*. PhD Thesis.
- Battese, G.E., Coelli, T. J., 1988. Prediction of firm level technical inefficiencies with a generalised frontier production function and panel data, *Journal of Econometrics*, 38, pp. 387-399.
- Battese, G.E., T.J. Coelli, 1995, A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, 20, pp. 325-332.
- Brooke, A. D., Kendrick, D., Meerhaus, A. *GAMS: A User's Guide*, Scientific Press, California. 1992
- Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, pp 429-444.
- Coelli, T., 1996. A guide to FRONTIER version 4.1: a computer program for frontier production function estimation. CEPA Working Paper 96/07, Department of Econometrics, University of New England, Armidale, Australia.
- Coelli, T.J., Perelman, S., 2000. Technical Efficiency of European railways: a distance function approach. *Applied Economics*, 32, pp. 1967-76.

- Coelli, T.J., Prasada Rao, D.S., Battese, G.E, 1998. An Introduction to Efficiency and Productivity Analysis, Kluwer Academic Publishers: USA.
- Garcia, J.J., Herrero, I., 1998. The production function for the Spanish trawl fleet fishing in Moroccan Waters. In Eide, A. and Vassdal, T. (Eds) IIFET'98 Proceedings, Norwegian College of Fisheries Science, Tromso, pp. 374-379.
- Greene, W. H., 1999. Análisis econométrico. Prentice Hall, Madrid.
- Greene, W., 2002. LIMDEP (Version 8.0): User's Manual and Reference Guide, Econometric Software Inc., New York.
- Gujarati, D. N., 2001. Econometría básica. McGraw-Hill, Santafé de Bogotá, Colombia.
- Gunst, R.F., Mason, R.L., 1977. Advantages of Examining Multicollinearities in Regression Analysis. *Biometrics*, 33, pp. 249-260
- Hallam, D., Machado, F., 1996. Efficiency Analysis with Panel Data: A study of Portuguese Dairy Farms. *European Rev. Of Agricultural Economics*, 23, 79-93.
- Herrero, I., Pascoe, S., 2003. Value vs. volume in the catch of the Spanish South-Atlantic Trawl Fishery. *Journal of Agricultural Economics*, 54(2), pp. 325-342.
- Herrero, I., 2004. Risk and Strategy of Fishers Alternatively Exploiting Sea Bream and Tuna in the Gibraltar Strait Using an Efficiency Perspective. *ICES Journal of Marine Science*, 61(2), pp. 230-236.
- Herrero, I., 2004. An analysis of the Octopus fishery that operates in the Spanish South Atlantic area. Submitted to *ICES Journal of Marine Science*.
- Hsiao, 1986. Analysis of Panel Data. *Econometric Society Monographs*.
- Huang, Liu, 1994. Estimation of a non-neutral stochastic frontier production function. *Journal of Productivity Analysis*, 5(2), pp. 171-180.
- Jondrow, J., Lovell, C.A.K., Materov ,I.S. and Schmidt, P., 1982. On the estimation of technical inefficiency in the stochastic frontier production function model, *Journal of Econometrics*, 19, pp. 223-238.
- Judge, G.G., Griffiths, W.E., Carter Hill, R., Lütkepohl, H., Tsoung-Chao Lee, 1980. The theory and Practice of Econometrics. Wiley and Sons. ISBN 0-471-89530-X

- Kirkley, J. E., Squires, D., Strand, I. E., 1995. Assessing technical efficiency in commercial fisheries: The Mid-Atlantic Sea Scallop Fishery, *American Journal of Agricultural Economics*, 77(3), pp. 686-97.
- Kirkley, J.E., Squires, D., Strand, I.E., 1998. Characterizing managerial skill and technical efficiency in a fishery, *Journal of Productivity Analysis*, 9, 145-160.
- Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restrictions, *Econometrica* 54(5), pp. 1243-1248.
- Kumbhakar, S.C., Lovell, C.A.K., 2000. *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.
- Mason, R.L., Gunst, R.F., Webster, J.T., 1975. Regression Analysis and problems of Multicollinearity. *Communications in Statistics*, A, 4 (3), pp. 277-292.
- Nunamaker, T.R., 1985. Using Data Envelopment Analysis to measure the efficiency of Not-for Profit Organizations: a critical evaluation". *Managerial and Decision Economics*, 6(1), pp. 50-59.
- Pascoe, S., Andersen, J.L., de Wilde, J.W., 2001. The impact of management regulation on the technical efficiency of vessels in the Dutch beam trawl fishery. *European Review of Agricultural Economics* 28(2), pp. 187-206.
- Pascoe, S., Coglan, L., 2002. Contribution of unmeasurable factors to the efficiency of fishing vessels: an analysis of technical efficiency of fishing vessels in the English Channel. *American Journal of Agricultural Economics* 84(3), pp. 45-57.
- Pascoe, S., Coglan, L., and Mardle, S. 2001. Physical vs. harvest measures of capacity: the case of the UK vessel capacity unit. *ICES Journal of Marine Science*, 58 (6): 1243-1252.
- Pascoe, S., Herrero, I., 2004. Estimation of a composite fish stock index using Data Envelopment Analysis. Accepted in *Fisheries Research* (forthcoming).
- Retzlaff-Roberts, D. L., Morey, R. C., 1993. A goal-programming method of stochastic allocative data envelopment analysis. *European Journal of Operational Research* 71(3), pp. 379-397.
- Saen, R.F., Memariani, A., Lofti, F.H., 2004. The effect of correlation coefficient among multiple input vectors on the efficiency mean in data envelopment analysis. *Applied Mathematics and Computation* (forthcoming).
- Sengupta, J. K., 1987. Data Envelopment Analysis for efficiency measurement in the stochastic case, *Computers and Operations Research* 14(2), pp. 117-29.

- Smith, P., 1997. Model misspecification in Data Envelopment Analysis. *Annals of Operational Research*, 73, pp. 233-252.
- Squires, D., Kirkley, J., 1999. Skipper skill and panel data in fishing industries. *Canadian Journal of Fisheries and Aquatic Sciences*, 56, pp. 2011-2018.
- Van Hoof, L., Salz, P., 2001. Applying CPUE as management tool. EAFE Conference. Salerno, Italy, 18-20 April, 2001.
- Vestergaard, N. (coordinator), Hoff, A., Andersen, J., Lindebo, E., Gronbaek, L., Pascoe, S., Tigley, D., Mardle, S., Guyader, O., Daures, F., van Hoof, L., de Wilde, J.W., Smith, J. Measuring capacity in fishing industries using the Data Envelopment Analysis Approach. Research project report, 2002. Department of Environment and Business Economics. University of Southern Denmark and the authors. ISBN 87-90872-05-3.

Table 1. Technical characteristics of the fleet (Trawl fishery)

	1985		1993		Average 1993-99	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Volume (in GRT)	91	36	59	14.9	59	15
Engine power (in HP)	300	94	330	61	324	83
Trips per month	1.6	0.6	1.7	0.7	1.7	0.7

Table 2. Maximum volume allowed for the Spanish North Trawl fishery in Moroccan Waters.

	1/12/95-30/11/96	1/12/96-30/11/97	1/12/97-30/11/98	1/12/98-30/11/99
Max. volume	11,200	10,000	9,000	8,200

(source: Barroso, 1998).

Table 3. Results of the likelihood ratio tests on the SPF model (omitting engine power).

Null hypotheses	Log-likelihood under H_0	Base Log-likelihood	λ -ratio	Critical value	Decision
$H_0:\gamma=0$	-244.620	-148.925	191.391	7.04 ^(a)	Reject H_0
$H_0:\beta_{ij}=0$	-155.928	-148.925	14.006	7.81	Reject H_0
$H_0:\eta=0$	-158.206	-148.925	18.563	3.84	Reject H_0
$H_0:\mu=0$	-155.364	-148.925	12.879	3.84	Reject H_0

(a) This critical value has been calculated using the tables in Kodde and Palm (1986).

Table 4. Estimated SPF model (omitting engine power).

	coefficient	t-ratio	
Intercept	0.761	10.188	***
Lngrt	0.670	2.634	***
Lntrips	1.056	30.049	***
Grt2	0.429	1.616	*
Trips2	-0.140	-1.391	
Grttrips	0.115	0.794	
Sigma-sq.	0.246	7.645	***
Gamma	0.612	15.723	***
Mu	0.776	5.742	***
Eta	-0.006	-3.582	***

Significance level: * 10 per cent; ** 5 per cent; *** 1 per cent.

Table 5. Panel data model (omitting engine power).

	coeff	t-stat
grt	0.255	2.084 ***
trips	1.070	11.165 ***
grt2	0.013	0.031
trips2	0.150	1.009
grttrips	-0.013	-0.050

Significance level: * 10 per cent; ** 5 per cent; *** 1 per cent.

Table 6. Results of the likelihood ratio tests on the distance functions model (omitting engine power).

Null hypotheses	Log-likelihood under H_0	Base Log-likelihood	Decision
Separability	-324.27	-303.79	Reject H_0
Homothet.	-315.48	-303.79	Reject H_0
Homogeneity	-324.64	-303.79	Reject H_0
$H_0: \eta=0$	-313.63	-303.79	Accept H_0
$H_0: \mu=0$	-305.31	-303.79	Accept H_0
$H_0: \gamma =0$	-310.61	-303.31	Reject H_0

Table 7. Estimated distance functions model (omitting engine power).

	coefficient	t-ratio
Intercept	-1.094	-1.569
Fish	-0.028	-1.351
trips	1.114	4.018***
grt	3.278	2.390***
fish2	0.001	9.966***
trips2	-0.262	-2.858***
grt2	-1.598	-2.234***
tripsgrt	0.482	2.425***
tripsfish	-0.052	-6.413***
grtfish	-0.018	-1.315
Sigma-sq.	0.609	3.575***
Gamma	0.671	6.971***
Mu	-	-
Eta	-	-

Significance level: * 10 per cent; ** 5 per cent; *** 1 per cent.

Table 8a. Correlation among TE estimates from different approaches
(omitting engine power).

	DF grt	SPF grt	PD grt	DEA grt
DF grt	1			
SPF grt	0.80	1.00		
PD grt	0.81	0.89	1	
DEA grt	0.84	0.73	0.86	1

Table 8b. Correlation among TE estimates from different approaches
(omitting volume)

	DF hp	SPF hp	PD hp	DEA hp
DF hp	1			
SPF hp	0.87	1		
PD hp	0.84	0.98	1	
DEA hp	0.80	0.78	0.77	1

Table 9a. Averages TE scores for the different approaches (omitting engine power)

	DF grt	SPF grt	PD grt	DEA grt	Average
Average	0.578	0.448	0.584	0.306	0.479
St. Deviat.	0.151	0.139	0.168	0.090	0.137

Table 9b. Averages TE scores for the different approaches (omitting volume)

	DF hp	SPF hp	PD hp	DEA grt	Average
Average	0.424	0.581	0.600	0.291	0.474
St. Deviat.	0.153	0.134	0.183	0.121	0.148

Figure 1. Comparison of results per boat (when omitting engine power)

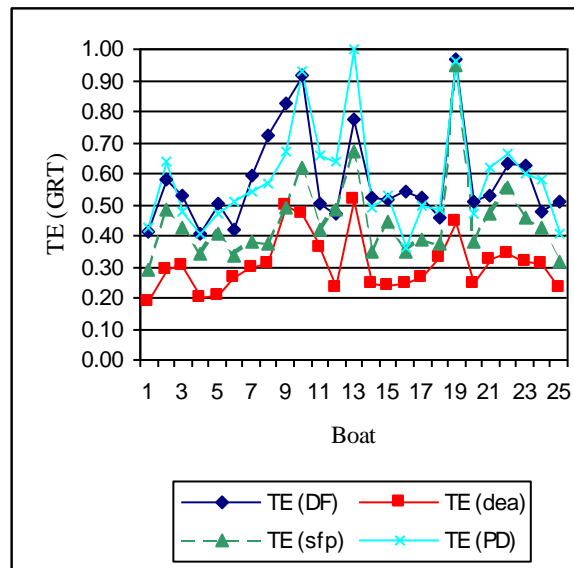


Figure 2. Comparison of results per boat (when omitting volume)

