*Quaderni di Statistica*
*Vol. 5, 2003*

# RegArima models and the Autoregressive metric for Italian fishery time series

Paolo Accadia, Vincenzo Placenti
*IREPA Onlus – Istituto di ricerche economiche per la pesca e l'acquacoltura*
*E-mail: accadia@irepa.org; placenti@irepa.org*

*Summary*: This paper analyses the exploitation dynamics of marine resources by partitioning the Italian fleet into six segments according to fishing method. RegARIMA models are used to identify the main components of the dynamics of catches and revenues per gross tonnage unit, and the Autoregressive metric to identify the distance between time series. The outcome is a classification of fishery segments that can be interpreted in the light of the parametric formulation of their models. This classification can be clearly linked to the regulation activity levels established by law. Such a link would be justified, among other things, by the effect of high levels of regulation on the seasonality of the phenomena analysed.

*Keywords*: AR Metric, RegARIMA Models, Fishery Management.

## *1. Introduction*

The fisheries sector is characterised by both its economic and its biological importance. Compared to other types of economic activity, the fisheries sector requires constant planning and control to preserve the natural resources that it affects. Fish stocks must be renewed in order to compensate for the mortality due to both natural causes and fishing. It is for this reason that the Public Administration (PA) continues to develop planning and intervention policies to restrict exploitation of stocks. Policy interventions do not affect all fishery operators equally, however, owing to their differing impacts on stock impoverishment. The Italian fleet can be conveniently partitioned into homogeneous segments based on bio-economic criteria. Vessel size and fishing equipment are the main indicators used for such partitioning. IREPA, Istituto di Ricerche Economiche per la Pesca e l'Acquacoltura (Fishery and Aquaculture Economic Research Institute), distinguishes among six fishery segments: bottom trawl, mid water pair trawl, purse seine, small-scale fishery, mechanised dredge, and polyvalent (IREPA, 1999).

The purpose of this paper is to analyse the exploitation dynamics of each segment in order to determine potential similarities. This may prove useful to the PA when defining its policy to preserve fishery stocks.

The time series analysed in the paper relate to catches and to revenues per GT (gross tonnage[1]) unit. These series provide an indication of, respectively, the technical and economic efficiency of the above fishery segments. It should be noted that the series are affected by changes in legislation as well as by exceptional weather conditions during data collection. It was therefore necessary to correct the time series for these effects before estimating the similarities among the dynamics of fishery segments. To this end, linear regression models with ARIMA errors, regARIMA models in short (Planas, 1997, Chapter 8), were applied to the data. These models comprise a deterministic component, which measures the trading day and movable feast (Easter) effects as well as the possible anomalous-value ones, and a stochastic component, which is explained using ARIMA models (Box and Jenkins, 1976). Once the series had been corrected for the deterministic effects, and once possible non-linearities had been eliminated by using a logarithmic transformation, the residual stochastic components were compared and classified using the Autoregressive metric introduced by Piccolo (1984, 1990).

The paper is organized as follows. Section 2 describes the time series used for the analysis. Section 3 illustrates the methodology employed to construct and classify the models, while Section 4 deals with analysis of the results. For the purpose of interpreting the resulting clusters, Section

[1] A volume measurement equal to 100 cubic feet, or 2.832 cubic meters.

5 illustrates the basic concepts of fishing activity regulation. Finally, some conclusions are drawn in Section 6.

## *2. Italian fishery time series*

The 12 monthly time series - 6 for captures per GT unit and 6 for revenues per GT unit - analysed in this paper are listed in Table 1 and shown in Figure 1. The time period employed for each series was January 1993-October 2000, giving a total of 94 observations. The time series were obtained from the sampling database compiled by IREPA as part of the programme entitled "Economic Observatory on the Production Structure of Italian Marine Fishing" (Osservatorio economico sulle strutture produttive della pesca marittima in Italia) (see IREPA, 1999).

*Table 1. Time series analysed. Period: 01/1993 - 10/2000. No. of observations: 94.*

| Abbreviation | Time series |
|---|---|
| BT-CAT | Bottom trawl captures by tonnage unit |
| MT-CAT | Mid water pair trawl captures by tonnage unit |
| PS-CAT | Purse seine captures by tonnage unit |
| SF-CAT | Small-scale fishery captures by tonnage unit |
| MD-CAT | Mechanised dredge captures by tonnage unit |
| PO-CAT | Polyvalent captures by tonnage unit |
| BT-REV | Bottom trawl revenues by tonnage unit |
| MT-REV | Mid water pair trawl revenues by tonnage unit |
| PS-REV | Purse seine revenues by tonnage unit |
| SF-REV | Small-scale fishery revenues by tonnage unit |
| MD-REV | Mechanised dredge revenues by tonnage unit |
| PO-REV | Polyvalent revenues by tonnage unit |

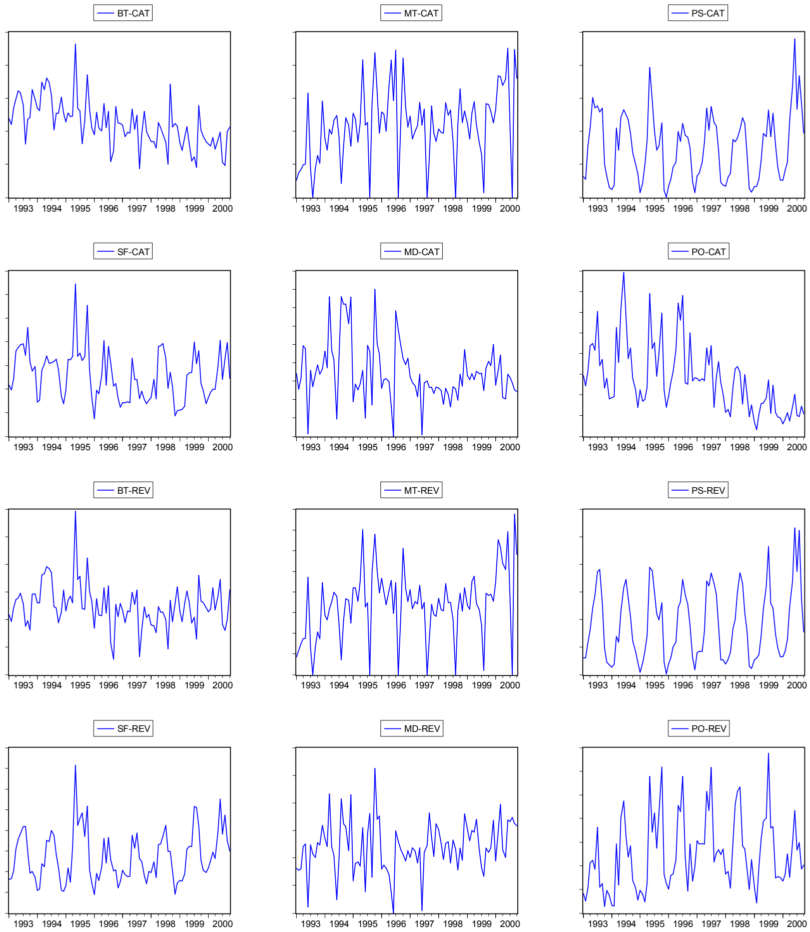


*Figure 1. Time series plots. Captures expressed in kg; revenues expressed in EUR.*

### *3. The Autoregressive metric and the regARIMA model*

The Autoregressive metric proposed by Piccolo (1984; 1990) has been used for various purposes, such as analysing demographic changes (Corduas and Piccolo, 1995), classifying stationary time series

(Maharaj, 1996), assessing the adequacy of seasonal adjustment procedures (Corduas and Piccolo, 1999), and evaluating and optimising the efficiency of environmental monitoring networks (Costanzo and Sarno, 2000). In this paper, we use the AR metric to classify fishing methods and to evaluate similarities among their exploitation dynamics of marine resources.

To introduce the AR metric, we may consider the generic process $Z_t \sim ARIMA(p, d, q)(P, D, Q)_s$:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D Z_t = \theta(B)\Theta(B^s)a_t,$$

where $a_t$ is a normal identically and independently distributed *white-noise* variable $(0,\sigma_a^2)$. When all the roots of the characteristic equation associated with the MA components - that is, $\theta(B)\Theta(B) = 0$ - are external to the unit circle, the $Z_t$ process can be inverted and allows for the linear representation AR (∞):

$$\pi(B)Z_t = a_t,$$

where

$$\pi(B) = \phi(B)\Phi(B^s)\theta^{-1}(B)\Theta^{-1}(B^s) = 1 - \sum_{i=1}^{\infty}\pi_i B^i .$$

To measure the structural diversity between two invertible processes, say $X_t$ and $Y_t$, the AR metric compares their respective sequences of autoregressive weights $\pi_{xj}$ and $\pi_{yj}$ (with $j$ = 1, 2, …), by means of a Euclidean distance:

$$d(X_t, Y_t) = \sqrt{\sum_{j=1}^{\infty}\left(\pi_{xj} - \pi_{yj}\right)^2} .$$

Consideration of a convenient finite approximation of the above AR(∞) representations and suitable estimates of the autoregressive coefficients yield the AR distance estimator:

$$\hat{d}(X_t, Y_t) = \sqrt{\sum_{j=1}^{m} \left(\hat{\pi}_{xj} - \hat{\pi}_{yj}\right)^2} \quad .$$

Its asymptotic distribution is known in the case of AR (Piccolo, 1989) and ARMA (Corduas, 1996) model comparisons under ML estimates, and in the case of MA models under LS estimates (Sarno, 2000, 2001). In particular, Piccolo (1989) showed that the asymptotic distribution of the ML estimator is a linear combination of independent Chi-squared random variables.

The data used to estimate the autoregressive coefficients were previously corrected for deterministic effects. For this purpose we used the regARIMA models developed by the U.S. Census Bureau and described by Findley et al. (1998). Considering the generic time series $y_t$, the model can be written as:

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D(y_t - \sum_{i=1}^{k}\beta_i x_{it}) = \theta(B)\Theta(B^s)a_t$$

where $x_{it}$ are the $k$ deterministic regressors and $\beta_i$ are the regression coefficients. In the analysis reported by this paper, we took account of the regressors most often used in practice, i.e., trading day, movable feast (Easter) and outliers (for more details see Planas, 1997, Chapter 8).

## *4. Analysis and results*

The models were identified and evaluated using the TRAMO procedure[2] (Gomez and Maravall, 1997). The identification step was

[2] TRAMO is the first part of a more complex procedure called TRAMO-SEATS. It is the official procedure used by ISTAT since 1999 for the seasonal adjustment of time series. TRAMO (Time series Regression with Arima noise, Missing observation and Outliers) identifies the statistical models that effectively describe the temporal evolution of the time series; SEATS (Signal Extraction in Arima Time Series) performs the actual seasonal adjustment of the series, using the structure of the statistical models identified by TRAMO.
ISTAT's decision to use this new procedure and to abandon the X11-ARIMA (Istat, 1987) one was taken in the light of recommendations by the Scientific Commission SARA. This choice

carried out by using the automatic module of Demetra 2.0 software, a graphical interface of TRAMO-SEATS for Windows. The default parameters were used, except when correcting for the trading day effect, where the use of a single regressor was preferred to the seven default ones, given that only the difference between working and non-working days (Saturday and Sunday) was deemed to be relevant to the fisheries sector.

*Table 2. Final models*

| Time series | ARIMA models | Average adjustment | Logarithm | Working days effect | Easter effect | Outliers |
|---|---|---|---|---|---|---|
| BT-CAT | (0 1 1)(0 1 1) | No | Yes | Yes | No | AO JUN1999, AO SEP1998, |
| MT-CAT | (0 1 1)(0 1 1) | No | No | Yes | No | TC FEB2000, |
| PS-CAT | (0 1 1)(0 1 1) | No | No | Yes | No | AO JUN2000, AO MAY1995, AO OCT1995, |
| SF-CAT | (0 1 1)(0 1 1) | No | Yes | Yes | No | AO JAN1996, |
| MD-CAT | (0 1 1)(0 1 1) | No | No | No | No | LS JAN1995, TC JUL1996, TC OCT1995, LS AUG1994, AO MAR1994, |
| PO-CAT | (0 1 1)(0 1 1) | No | Yes | Yes | No | |
| BT-REV | (1 0 0)(1 0 0) | Yes | Yes | Yes | No | AO SEP1996, |
| MT-REV | (1 0 1)(1 0 0) | Yes | No | No | No | LS JAN2000, |
| PS-REV | (1 0 0)(0 1 1) | No | No | Yes | No | TC JUN2000, AO JUL1999, AO AUG2000, AO OCT1995, |
| SF-REV | (1 0 0)(0 1 1) | No | Yes | Yes | No | |
| MD-REV | (0 0 0)(1 0 0) | Yes | No | No | No | AO OCT1995, |
| PO-REV | (1 0 1)(0 1 1) | Yes | Yes | Yes | No | |

has been adopted by other national statistical institutes and by the Statistical Office of the European Community (Eurostat). Indeed, the DEMETRA 2.0 software, a graphical interface of TRAMO-SEATS for Windows, developed by Eurostat was used in this paper.

The scientific commission SARA (Seasonal Adjustment Research Appraisal) headed by prof. D. Piccolo, University of Naples Federico II, concluded its work in 1998. For a review of the results achieved see Istat (2000).

By default, TRAMO verifies the need for logarithmic transformation and mean correction, the significance of trading day and Easter effect regressors, and identifies the presence of possible outliers, taking three separate kinds into account: additive outliers (AO), temporary changes (TC), and level shifts (LS).

The models identified were estimated with the same software and using the exact maximum likelihood method (Gomez and Maravall, 1994). They were validated by means of residuals analysis and specification tests.

The 12 final models are shown in Table 2, and the estimate for the autoregressive and moving average parameters in Table 3.

*Table 3. Estimate for the parameters of the ARIMA model*

| Models | AR(1) | MA(1) | SAR(1) | SMA(1) |
|---|---|---|---|---|
| BT-CAT | | 0.82 (0.057) | | 0.37 (0.197) |
| MT-CAT | | 0.78 (0.076) | | 0.34 (0.129) |
| PS-CAT | | 0.69 (0.086) | | 0.67 (0.146) |
| SF-CAT | | 0.57 (0.091) | | 0.98 (0.020) |
| MD-CAT | | 0.71 (0.090) | | 0.42 (0.136) |
| PO-CAT | | 0.67 (0.088) | | 0.90 (0.351) |
| BT-REV | 0.36 (0.099) | | 0.51 (0.091) | |
| MT-REV | 0.93 (0.072) | 0.79 (0.130) | 0.74 (0.080) | |
| PS-REV | 0.39 (0.103) | | | 0.94 (0.038) |
| SF-REV | 0.45 (0.101) | | | 0.86 (0.192) |
| MD-REV | | | 0.32 (0.098) | |
| PO-REV | 0.74 (0.186) | 0.43 (0.243) | | 0.99 (0.003) |

Standard error of the estimate in parentheses

As regards the deterministic component (Table 2), it may be observed that the Easter effect was not statistically significant for all six fishing systems, while the calendar effect was statistically significant for all systems except dredges (MD), and a considerable number of outliers were identified.

The ARIMA models identified and estimated by TRAMO using the series corrected for deterministic effects display a parsimonious parameterisation (Table 2) with the prevalence of the *Airline model,* ARIMA(0 1 1)(0 1 1). For the purpose of comparing and classifying these models, they were re-formulated in their purely AR representation, with $m$ = 48 selected as the truncation point. This value enabled account to be taken of the most “important” coefficients. The results are shown in Tables 4 and 5, from which the dendrograms are obtained (Figures 2 and 3) by using the single linkage as the cluster algorithm.

*Table 4. Matrix of the Euclidean distances for ARIMA models for captures by tonnage series.*

| | BT-CAT | SF-CAT | PO-CAT | MT-CAT | PS-CAT | MD-CAT |
|---|---|---|---|---|---|---|
| BT-CAT | 0 | | | | | |
| SF-CAT | 5.82 | 0 | | | | |
| PO-CAT | 5.07 | 1.45 | 0 | | | |
| MT-CAT | 0.73 | 5.61 | 4.97 | 0 | | |
| PS-CAT | 2.78 | 3.26 | 2.54 | 2.53 | 0 | |
| MD-CAT | 1.61 | 4.96 | 4.43 | 0.97 | 1.91 | 0 |

*Table 5. Matrix of the Euclidean distances for ARIMA models for revenues by tonnage series.*

| | BT-REV | SF-REV | PO-REV | MT-REV | PS-REV | MD-REV |
|---|---|---|---|---|---|---|
| BT-REV | 0 | | | | | |
| SF-REV | 3.66 | 0 | | | | |
| PO-REV | 4.33 | 1.22 | 0 | | | |
| MT-REV | 0.39 | 3.81 | 4.46 | 0 | | |
| PS-REV | 4.09 | 0.61 | 0.75 | 4.22 | 0 | |
| MD-REV | 0.45 | 3.52 | 4.2 | 0.5 | 3.96 | 0 |

Of the two phenomena analysed, the most straightforward classification was the one obtained for the revenues by tonnage. The dendrogram (Figure 3) splits the fishing systems in the following two groups:

- Bottom trawl (BT), mid water pair trawl (MT) and mechanised dredge (MD);
- Small-scale fishery (SF), purse seine (PS) and polyvalent (PO).

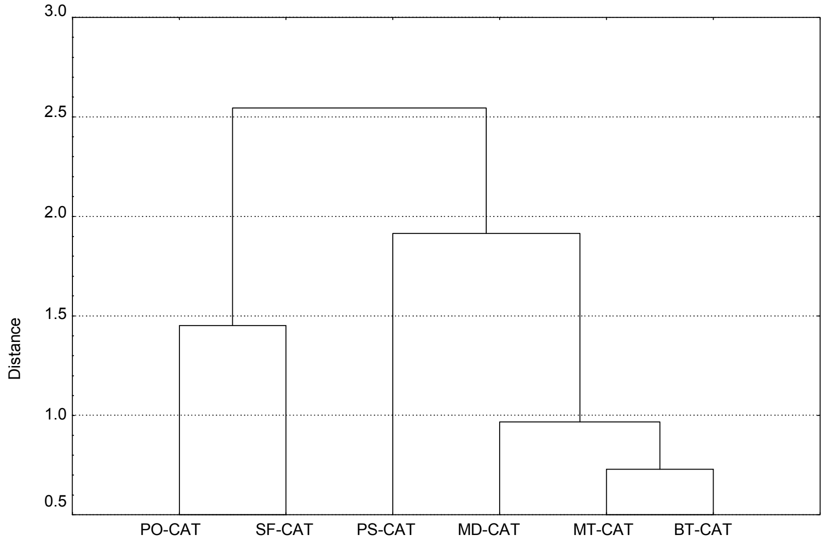


*Figure 2. Dendrogram for the data in Table 4. Euclidean distance – Single linkage.*

The matrix of the Euclidean distances (Table 5) associated with this phenomenon exhibits, for practical purposes, almost identical values between elements of the first group, while in the second the polyvalent (PO) is closer to purse seine (PS) than to small-scale fishery (SF).
The captures by tonnage dendrogram (Figure 2) confirms the presence of the same two groups, save for the purse seine (PS). Indeed, the latter is not included in the second group, but instead occupies an intermediate position. This is evident from the matrix of the Euclidean distances (Table 4), where the purse seine (PS) is equidistant from the mid water pair trawl (MT) and from the polyvalent (PO). It is therefore impossible to classify it in either of the two groups.

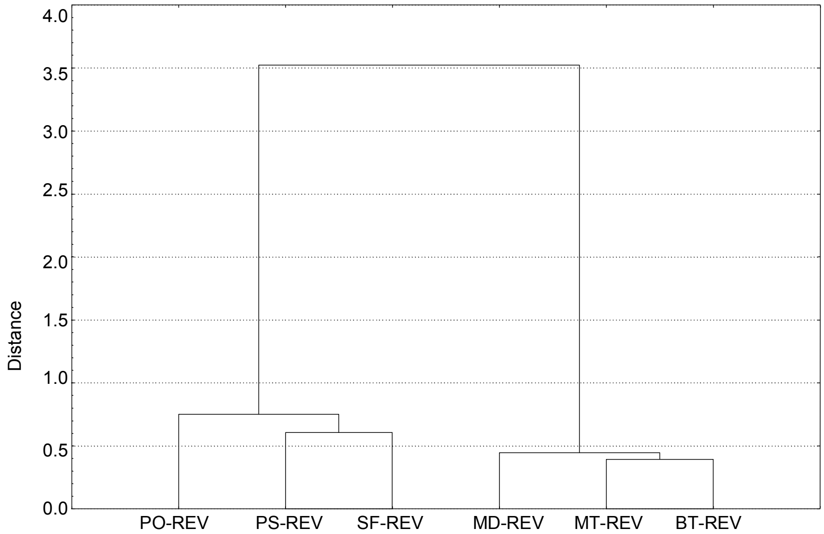


*Figure 3. Dendrogram for the data in Table 5. Euclidean distance – Single linkage.*

But which components of the temporal dynamics of the phenomena analysed give rise to these results? In order to answer this question, a comparison between the groups resulting from analysis of the Euclidean distance matrices was performed, using the parametric formulation of the respective ARIMA models.

The element that differentiates the models built on the revenues series by tonnage in group 1 from those of group 2 is the seasonal component. This is of the (1 0 0) type for bottom trawl (BT), mid water pair trawl (MT) and mechanised dredge (MD), and of the (0 1 1) type for all the other systems. This difference suggests the more marked seasonality of the systems in the second group, as can be easily demonstrated analytically. Let us consider, for example, the models of bottom trawl (BT) and small-scale fishery (SF) (Table 2):

BT-REV ~ ARIMA(1 0 0)(1 0 0):

$$X_t = \phi_{1x} X_{t-1} + \phi_{12x}(X_{t-12} - \phi_{1x} X_{t-13}) + a_t$$

SF-REV ~ ARIMA(1 0 0)(0 1 1):

$$Y_t = \phi_{1y} Y_{t-1} + \phi_{12y}(Y_{t-12} - \phi_{1y} Y_{t-13}) + (a_t - \theta_{12y} a_{t-12}).$$

In both cases, the seasonal component is represented by an autoregressive coefficient of the first order, although for bottom trawl (BT) $-1 < \phi_{12x} < 1$, while for small fishery (SF) $\phi_{12y} = 1$. This means that the seasonal link characterising the systems in the first group is weaker than that characterising the systems in the second.

The models built on the series of the captures by tonnage are all of the *Airline*, ARIMA(0 1 1)(0 1 1), type. Consequently, information to explain the grouping obtained through the application of the AR metric cannot be extracted from their formulation. However, one can infer from the analysis of the estimates of the regular and seasonal MA parameters (Table 3) that the models in group 1 have higher values of the regular component and lower values of the seasonal one, whereas the models in group 2 behave in exactly opposite manner, with purse seine (PS) occupying an intermediate position. This accurately reflects the pattern emerging from the dendrogram in Figure 2 and confirms the lesser seasonal dependency of the bottom trawl (BT), mid water pair trawl (MT) and mechanised dredge (MD) systems.

## *5. Activity regulation*

The results obtained by applying the Autoregressive metric can be interpreted by taking into account the different levels of regulation to which the fishery systems are subject.

Bottom trawl (BT) and mid water pair trawl (MT) exert greatest pressure on fishery resources and are therefore subject to mandatory biological and technical closures. The period of biological closure usually coincides with the phase immediately following reproduction of the economically and biologically most important species, while the technical closure restricts the number of fishing days per week permitted during the eight weeks following biological closure. This regulation, which is applied to both systems, inevitably gives rise to a similarity in the temporal dynamics of their production levels.

Restrictions are imposed on fishing activity as regards dredges (MD) as well, but for different reasons. The local managing consortiums, as officially provided for by the Third Triennial Plan and activated in 1997 by specific ministerial decree (DM 10.04.1997), determine, amongst other things, the nature and timing of the biological closure, which may last up to a maximum of two months between May and September. The presence of the biological closure period for dredges (MD) explains its similarity to bottom trawl (BT) and mid water pair trawl (MT).

Conversely, there is no specific regulation of fishing activity as regards the other three systems, i.e., small-scale fishery (SF), polyvalent (PO), and purse seine (PS). It is consequently continuous throughout the year. A partial exception is purse seine fishing (PS), for which a biological closure was required in the past, albeit at the discretion of the operators. This explains the system's intermediate position in the tree diagram shown in Figure 2.

## *6. Conclusions*

In general, the results obtained by applying the Autoregressive metric closely match the following subdivision of fishing systems based on the level of activity regulation to which they are subject.

<table>
<tr><th>Regulation level</th><th>Regulation type</th><th>System</th></tr>
<tr><td rowspan="2">High</td><td>Regulated</td><td>Bottom trawl<br>Mid water pair trawl</td></tr>
<tr><td>Self-regulated</td><td>Mechanised dredge</td></tr>
<tr><td>Low</td><td>No specifically regulated</td><td>Small scale fishery<br>Purse seine<br>Polyvalent</td></tr>
</table>

It can be explained by taking into account the level of seasonal dependence shown in the ARIMA formulations of the corresponding models. Seasonality is undoubtedly present in the fishery sector, and it is due mainly to meteorological conditions and to the biological cycles of the species fished. Its importance obviously depends on the fishing system considered. Restrictions of activity levels linked not to seasonal factors, but to interventions by the Public Administration authority to increase fish stocks, inevitably interfere with the temporal aspects, and especially the seasonal characteristics, of the phenomenon's evolution. As a consequence, the systems subject to regulation display less seasonality in production levels when compared to other systems.

The results obtained could be verified by applying the same methodology to regional time series. This would also enable classification of the various regions according to the regional production level dynamics of the sector, thereby providing the Public Administration with useful information for the formulation of management programmes intended to rationalise and develop the fishery sector. Moreover, regional-level analysis would introduce a richer deterministic component into the regARIMA models by taking local deterministic effects into account. As the foremost among these effects, the biological and technical closures would make the most difference by being regulated in different forms and with different activation timetables along the Italian coastline. Further research is planned on this.

## ***References***


Box G.E.P. and Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, Holden-Day, S. Francisco CA.

Corduas M. (1996), Uno studio sulla distribuzione asintotica della metrica autoregressiva, *Statistica*, LVI, 3, 321-332.

Corduas M. and Piccolo D. (1995), Mutamenti strutturali della natalità e metodi statistici, *Atti del Convegno SIS: Continuità e discontinuità nei processi demografici*, Arcavacata di Rende, 315-322.

Corduas M. and Piccolo D. (1999), On the use of the AR metric for seasonal adjustments, *Atti del CLADAG*, Rome, 1-4.

Costanzo G.D. and Sarno E. (2000), La metrica autoregressiva per la valutazione e ottimizzazione di sistemi di monitoraggio ambientale, *Quaderni di Statistica*, 2, 205-220.

Findley D.F., Monsell B.C., Bell W.R., Otto M.C. and Chen B. (1998), New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program, *Journal of Business and Economic Statistics*, 16, 127-177.

Gomez V. and Maravall A. (1994), Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman Filter, *Journal of the American Statistical Association*, 89, 611-624.

Gomez V. and Maravall A. (1997), *Programs TRAMO and SEATS: Istructions for the User* (Beta version: November 1997), Bank of Spain.

Irepa (1999), Osservatorio economico sulle strutture produttive della pesca marittima in Italia 1999, *Ricerche*, n. 9.

Istat (1987), La destagionalizzazione degli indici della produzione industriale – anni 1977-1985, ISTAT, *Note e relazioni*, n. 2.

Istat (2000), Seasonal Adjustment Procedures – Experiences and Perspectives, Proceedings of International Conference – 9-10 June , 1998 – Rome, Italy, ISTAT, *Annali di Statistica*, n. 20.

Maharaj E.A. (1996), A significance test for classifying ARMA models, *Journal of Computation and Simulation*, 54, 305-331.

Piccolo D. (1984), Una topologia per i modelli ARIMA, *Statistica*, XLIV, 47-59.

Piccolo D. (1989), On a measure of dissimilarity between ARIMA models, *Proceedings of the Business and Economics Statistics Section of the ASA Meeting*, Washington, 231-236.

Piccolo D. (1990), A Distance Measure for Classifying ARIMA Models, *Journal of Time Series Analysis*, 11, 2, 153-164.

Planas C. (1997), Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter, *Statistical document*, Eurostat.

Sarno E. (2000), The behaviour of the AR metric for MA model comparisons, *Atti della XL Riunione Scientifica SIS*, Firenze, 149-152.

Sarno E. (2001), Further results on the asymptotic distribution of the Euclidean distance between MA models, *Quaderni di Statistica*, 3, 165-175.