Project no. 513754

INDECO

Development of Indicators of Environmental Performance of the Common Fisheries Policy

Specific Targeted Research Project of the Sixth Research Framework Programme of the EU on 'Modernisation and sustainability of fisheries, including aquaculture-based production systems', under 'Sustainable Management of Europe's Natural Resources'.

The application of modelling methods to ecosystem indicators in well-studied European marine fisheries ecosystems.

Dr. Murdoch McAllister, Dr. Pia Orr, Tom Carruthers & Marine Pomarede
(Imperial College, London)

Project Deliverable Number 17

Dissemination Level: Public

Due date: May 2006

Submission date: September 2006

Start date of project: 1 December 2004 Duration: 24 months

Lead name and organisation: Indrani Lutchman and James Brown, Institute for European Economic Policy (IEEP)

Table of Contents

1. Background to the INDECO time series analysis workshop held January 2006 1
1.1 Key Logistical Features of the Workshop 3
1.3 Aims of the Workshop 4
2. Results 5
2.1 Time Series analysis 5
2.2 Statistical Power analysis 6
3. Discussion 10
3.1 Time Series analysis 10
3.2 Statistical Power analysis 10
4. Summary 12
4.1 Time series analysis 12
4.2 Statistical power analysis 13
5. References 15
6. Appendices 16
Appendix 1 Workshop Announcement 16
Appendix 2 Workshop overhead slides 20
Appendix 3 Workshop reports 81
A3.1 Power analysis of time series analysis in log abundance indices from the EVHOE survey 81
A3.2 Time series analysis of trawl fleet data from the Aegean and Ionian Seas 86
A3.3 Basque Country Time Series Analysis 95

1. Background to the INDECO time series analysis workshop held January 2006

To give a brief recap of Work Package 5 (WP5), the main goals and tasks of WP5 are as follows:

- *Review existing models that analyse and apply fisheries ecosystem indicators*
- *Apply some of these models to specific indicators proposed by WP 2 to 4*

INDECO is a coordinated action project It therefore set out to review and possibly reformulate and apply some relatively simple and generic analytical models that integrate and analyse the indicators developed in other WPs for the purpose of monitoring and evaluating the environmental performance of the CFP. The "modelling methods" of concern to WP5 are either mathematical (e.g., ecosystem dynamics or mass balance models) or statistical (e.g., time series analysis or statistical power analysis).

The first deliverable of WP 5 reviewed modelling methods that perform the following tasks:

- Synthesize from existing data indicators suitable for EBFM
- Analyse the properties of indicators for screening purposes
- Utilize indicators to guide management decisions

Most of the review/compilation was done by Dr. Pia Orr who compiled and reviewed several hundred articles on modelling fisheries ecosystem indicators. The revised draft was completed in November 2005 after reviews of earlier drafts by INDECO partners and the INDECO advisory panel. The main conclusion of the review is that there are several models available and potentially suitable for formulating, evaluating and applying indicators in European fisheries. Some of the most common of these modelling methods include the following:

Time series analysis models – These include univariate and multivariate models; one of the main purposes is to separate the signal from the noise, i.e. to estimate and represent the underlying temporal pattern of the process of interest.

Multivariate analysis models -- These include for example, principle components analysis which synthesizes indicators with common properties.

Statistical power analysis models – These can evaluate ability to correctly detect time trends in indicators.

Mass balance models – These evaluate the impacts of fishing on the trophic flow in ecosystems; yet they tend to have poor predictive power, and it is difficult to assess uncertainty in the model outputs.

Simulation evaluation modelling frameworks – These complex models simulate ecosystem dynamics and indicators based on plausible hypotheses for system dynamics and data production.

In January 2006, a three-day workshop was held to explore the application of time series analysis methods to candidate indicators from a few different case studies. The questions addressed in the workshop and in this report include the following:

- What useful things can the various time series analysis modelling methods tell us about the various indicators?
- What are the input requirements, assumptions, key outputs of various methods of time series analysis?
- How should the time series analysis outputs be interpreted?
- How can the alternative methods help us to explore the properties of, and screen, candidate indicators?
- How can the methods help us to synthesize information contained in sets of candidate indicators from a single ecosystem?
- How can the methods provide guidance on minimal data quality requirements for indicators?
- What are some additional methods that deserve further attention?
- What is the minimum detectable annual rate of change in indicators with acceptably high power?

This document provides a brief summary of methods explored and conclusions from their application to a few of the case studies in the three-day INDECO workshop in January of 2006. Supporting material for these conclusions is provided in the following attachments:

(1) The flier sent out to prospective workshop participants prior to the workshop.

(2) The set of overhead instructional and introductory slides on time series analysis and statistical power analysis that were presented during the three day workshop.

(3) Some summary results provided by those workshop participants that attended the workshop and provided results of their analyses. Note that only IFREMER, AZTI and HCMR have provided output from their workshop analyses so far.

The attachments are quite detailed in their outline of methodology, assumptions and results. The summary documentation provided is therefore quite brief and makes reference to the attached documentation where appropriate to avoid unnecessary duplication of text.

1.1 Key Logistical Features of the Workshop

The workshop was hosted at IEEP London Office. The workshop was organized and lead by Dr. Murdoch McAllister of Imperial College London, Senior Lecturer in Statistical Risk Assessment. Dr. Pia Orr, an Imperial College Post doctoral research fellow, and Mr. Tom Carruthers, at PhD student, assisted in administering the workshop. There were twelve INDECO participants:

Gerjan Piet, IMARES, the Netherlands,

Michele Gristina, IAMC- CNR, Italy,

Inigo Muxika, AZTI, Spain,

Marie-Joelle.Rochet, IFREMER, France

Sasa Raicevich, ICRAM, Italy,

Angela Granzotto, IAMC- CNR, Italy,

Fabio Pravoni, University of Venice, Italy,

Magnus Appelberg , Swedish Board of Fisheries, Sweden,

Robert Aps, Estonian Marine Institute, Estonia,

John, Haralabous, HCMR, Greece,

Ingeborg Deboois, IMARES, the Netherlands,

Simone Libralato, University of Venice, Italy.

The three-day workshop consisted of a series of formal presentations of key concepts and methodologies, interspersed with methodological exercises that were applied to the candidate indicator datasets that the participants brought to work on in the workshop. Indicator datasets were brought from the Mediterranean (Adriatic and Aegean Seas), North Sea, Baltic Sea, and Bay of Biscay.

The workshop focused on Time Series Analysis (TSA) methods to evaluate the properties of the candidate indicators that were compiled prior to the workshop. TSA is a potentially useful method for analysing candidate indicator series in the following applications:

1. *formulate and test hypotheses about time trends and other temporal patterns in candidate indicators*
2. *formulate appropriate methods to compute the statistical power of tests for trends in candidate indicators*
3. *formulate recommendations about which sets of indicators may have suitable statistical properties*
4. *formulate recommendations about possible modifications to survey design to improve indicators*

1.3 Aims of the Workshop

To :

- *Apply time series analysis (TSA) methods to candidate indicator time series*
- *Evaluate statistical power in tests for future time trends in candidate indicators*
- *Apply multivariate statistical methods for TSA and statistical power analysis*
- *Provide training in some basic principles and methods of TSA*
- *Provide inputs to INDECO deliverables in WP2-5*

A variety of time series models were applied. These included the following:

- *Polynomial, ARIMA, ANCOVA*
- *Univariate (single dependent variable) and multivariate (more than one dependent variable) models*

These methods were applied to estimate time trends in the candidate indicators and the error variance associated with the estimated trends. Both the estimated trends and the associated error variance are key inputs to statistical power analysis (see below).

A variety of different statistical software were applied, including the following:

- *R – applied for simple time series analysis (excluding ARIMA modelling), univariate and multivariate statistical modelling, statistical power analysis.*
- *SAS - applied for simple time series analysis (excluding ARIMA modelling), univariate statistical modelling, and statistical power analysis.*
- *SPSS - applied for simple time series analysis and ARIMA TSA modelling, univariate and multivariate statistical modelling,*
- *Excel - applied for simple time series analysis, univariate statistical modelling and statistical power analysis*

Following the workshop sessions on time series analysis modelling, there were a few sessions on statistical power analysis. Power is the probability of correctly detecting an effect or trend in an index at a given sample size (number of years of data), level of significance (value for alpha or the acceptable agreed chance of a type I error or false positive), error variance and magnitude for the true underlying trend or slope in the time series. Power analysis can be applied to evaluate the statistical power to detect trends in indicator time series in the future, the number of years that it might take to detect a trend with acceptably high power, and the detectable effect size or trend under a given preset statistical power. We evaluated sensitivity of the calculated statistical power to the following:

- the value set for alpha,
- whether the test is one- or two-tailed test,
- the value for the slope in the time trend,
- the number of years of indicator data,

- residual error variance in the data relative to the presumed statistical model that was fitted to the data.
- whether only a univariate statistical model with only one dependent variable (indicator) was fitted to the data or a multivariate model with more than one indicator was fitted to the data.

2. Results

2.1 Time Series analysis

Generally speaking, the ARIMA family of time series models can be quite highly parameterized statistical time series models with a large variety of potential combinations of the auto regressive (AR) and integrated moving average (IMA) ARIMA components. Reliable ARIMA model selection can only take place if there are at least 50 years of data available. Most of the candidate indicators evaluated at the workshop had only up to about 20 or 30 years of indicator values in a time series. Thus, no confident and conclusive results could be obtained from TSA regarding the selection of ARIMA models for particular indicator time series. Relatively simple ARIMA models however were explored and some of these appeared to fit moderately better than others according to AIC criteria for model selection (see Appendix 2 for overhead presentation on ARIMA models for further details and example applications).

Alternatively, due to the relatively short time series of indicator data available, some simpler statistical time series modelling approaches were applied. These included polynomial statistical models and Loess smoothers. In a number of instances, polynomial models could be found to fit the historical data well. Because polynomial models of increasing order form a set of nested models, likelihood ratio tests or F-tests can be applied to test the null hypothesis that a polynomial one order higher does not fit the data better than the simpler polynomial model. Error variances are also easily computed using polynomial models. Thus for time series that are less than about 20 years in length seemed to lend themselves well to polynomial statistical modelling (see overhead presentation on Time Series Analysis which among other things gives an example of polynomial modelling of size based indicators from the IBTS).

Loess smoothing was also applied to estimate trends in indicators. Because Loess smoothers do not conform to parametric statistical models, it is not possible to estimate the error variance in a Loess smooth. However, it is possible to obtain a plausible value for the maximum rate of change in the indicator in the historic time series. Such approximations can be used in statistical power analyses to evaluate the power to detect potential future trends in time series (Nicholson and Jennings 2004).

In indicator datasets from several different regions, e.g., North Sea, Adriatic Sea, Aegean Sea, and Bay of Biscay, the annual fluctuations showed pronounced covariation. Take for example the size-based indicators from the IBTS survey in the North Sea (Fig. 1). The mean maximum length and mean maximum weight indicators both show in some years a strong upward fluctuation and then in the next year a downward fluctuation and this co varying pattern of fluctuation tends to be consistent across years. When indicators were standardized so that the mean value of all co varying time series was equal to the same constant, statistical analyses enabled the rejection of the null hypothesis that there is no common annual deviation from trend lines fitted separately to the different time series

(Segurado 2006). The statistical time series model can thus model more than one indicator series at a time and include a common time effect to take into account the covariation in deviates across years. The degrees of freedom for the N x 2 data points for the two time series are reduced by N-1 estimated common time effect terms. The residual variance however is considerably reduced and this may help to increase the statistical power in tests for trends in the indicator time series (see below). Similarly, multivariate analysis of abundance indices for species which covary across years could be modelled with a common year effect (see A3.1).

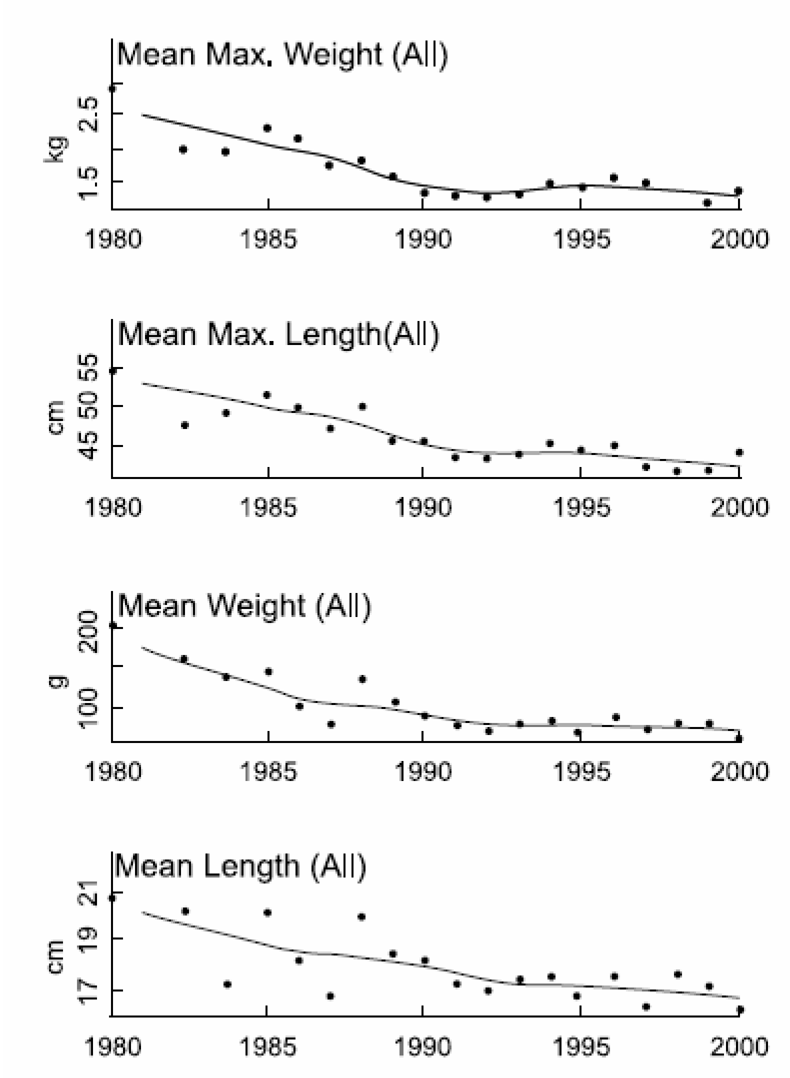


Figure 1. Loess Smooth trajectories plotted against IBTS size based indicators reported in Nicholson and Jennings (2004).

2.2 Statistical Power analysis

Nicholson and Jennings (2004) found that statistical power can be expected to be low for commonly available fisheries ecosystem indicators. For example, in size based ecosystem indicators from the IBTS, it would take about 12 years for statistical power of a two-tailed test of the hypothesis that the slope is equal to zero to exceed 80% (Fig. 2). This was computed presuming that the slope was at the maximum observed rate of change from the analysis of historical data. In abundance indices in the Bay of Biscay, it would take 10 – 50 years for statistical power to exceed 80% for most species (A3.1). From the point of view of basing management actions on detected changes in indicators, this was judged to be too many years to have to wait to be able to detect statistically significant changes in ecosystem indicators. Preferably, no more than about 3-5 years should go by before a trend can be statistically detected.

To evaluate the sensitivity of power to the inputs to the statistical power analysis, alternative inputs for whether it is a two-tailed test, alpha, and the residual standard error were inputted. In most statistical power analyses of fisheries ecosystem indicators to date (e.g., Nicholson and Jennings 2004; Trenkel and Rochet 2004), it has been assumed that the test is a two-tail test. However, in many management decision making contexts, decision rules about whether to take a management action depend on whether the indicator has gone over some threshold limit reference point. If there is a concern for example that a decline in a size based indicator suggested a further decline in the acceptable state of the ecosystem, then it may be appropriate to set up a decision rule requiring remedial action, if a decline in the indicator can be detected. This would then set up a one-tailed test since the null hypotheses would be no change or an increase in the indicator is occurring, while the alternative hypothesis of concern is that there is a decline in the indicator. When a one-tailed test was presumed in the statistical power analysis, power increased noticeably. For example in the IBTS, size based indicators, power increased from about 60% to about 75% at 10 years (Fig. 2). However, power is still low for management purposes. In the Bay of Biscay, the power of abundance indices increased by 20 to 50% depending on species (see A3.1)

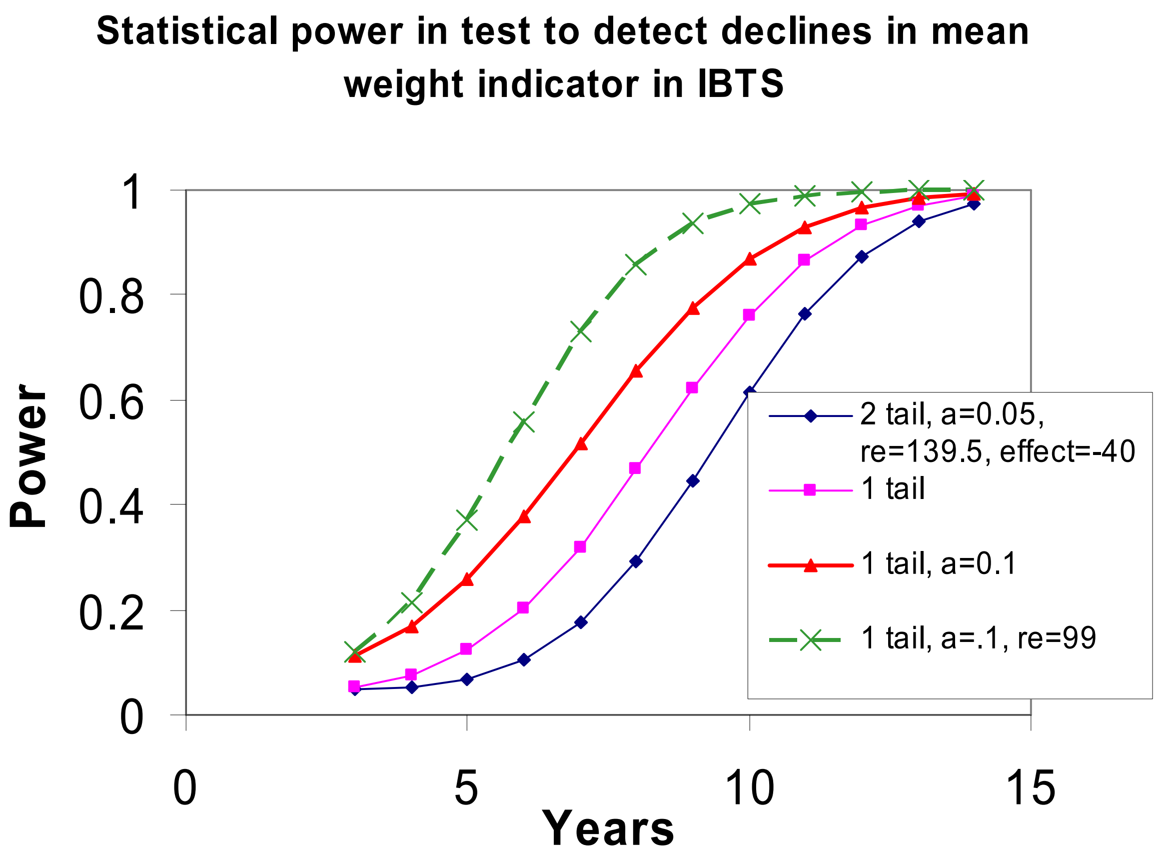


Figure 2. Statistical Power versus number of years in a test to correctly detect declines in mean weight indicators in the IBTS. "2 tail" refers to a one tailed test where the direction of the trend, either negative or positive is not of interest. "1 tail" refers to a two tailed test where only a decline in abundance is of interest and the statistical test applied is a one-tailed test. Where "a" refers to alpha, or the pre-agreed acceptable level of a type I error and “re" refers to the residual standard error assumed in the observed values. The term “effect” refers to the assumed true value for the slope of the indicator.

The alpha value has typically been set at 0.05 in previous statistical power analyses of fisheries ecosystem indicators (e.g., Nicholson and Jennings 2004). Reviews of the use of statistical power in environmental studies, e.g., Toft and Shea (1983) and Peterman (1990) have suggested that where the costs of a false negative, i.e. failing to detect a real change, are very high, and it is difficult to improve the precision and number of observations, it may be acceptable to increase the value for alpha from 0.05 to as high as about 0.1. When alpha was increased from 0.05 to 0.1, power increased noticeably (e.g., from about 75 to about 87% for the IBTS size based indicator, Fig. 2).

Typically, only one methodology and particular model has been applied to compute the potential residual error variance that is inputted into statistical power analyses of indicators. For example, Nicholson and Jennings (2004) applied a non-parametric estimator of residual variance that has been found to be "asymptotically inefficient" but provides "a good balance between bias and precision with small values of T" (Dette et al. 1998). Nicholson and Jennings (2004) applied only this single estimate of residual variance obtained for each candidate indicator in their statistical power analyses. However, it is never possible to know precisely the actual residual error variance, unless the various processes (i.e. the underlying dynamics) generating the data are known with considerable scientific certainty and there are very large amounts of data available with which to estimate the residual error variance.

Reviews of the application of statistical power analysis in environmental science have indicated that the assumed residual error variance is typically an uncertain input into statistical power analysis (e.g., Toft and Shea 1983; Peterman 1990). This is because of the sparseness of data with which residual error variance is often computed and the typical high uncertainty over the processes generating the data in the first place. These authors have thus suggested that power analyses should take into account uncertainty in the inputted residual error variances by evaluating the sensitivity of power to alternative plausible values for error variance. Alternative plausible values for error variance can be obtained by fitting structurally different time series models to historical data or applying different parametric and non-parametric approaches to estimating the residual error variance from available historic data or data obtained from the literature or pilot studies. These alternative methods can indicate a range of plausible values for the residual error variance.

When a set of alternative polynomial models were fitted to the size based indicator data from the IBTS and AIC was applied to choose the best model, the residual standard error dropped from about 140 to about 99. This drop in standard error increased the statistical power markedly (Figure 1). When a one –tailed test is applied, alpha is set at 0.1, and a lower residual variance is applied, power can be increased up to nearly five fold, e.g., from about 18% at seven years for the conventional case to about 85% (Figure 3).

When the covariance in a set of related indicators was taken into account, the residual variance was considerably less than when the indicators were analysed separately. When statistical power analysis was evaluated to detect trends in sets of covarying indicators where the common time effects were modelled, statistical power increased markedly. For example, high statistical power (i.e. about 80%) could be obtained in 6 years for the four IBTS sized based indicators and in 8 years for four correlated abundance indices in the Bay of Biscay (see A3.1).

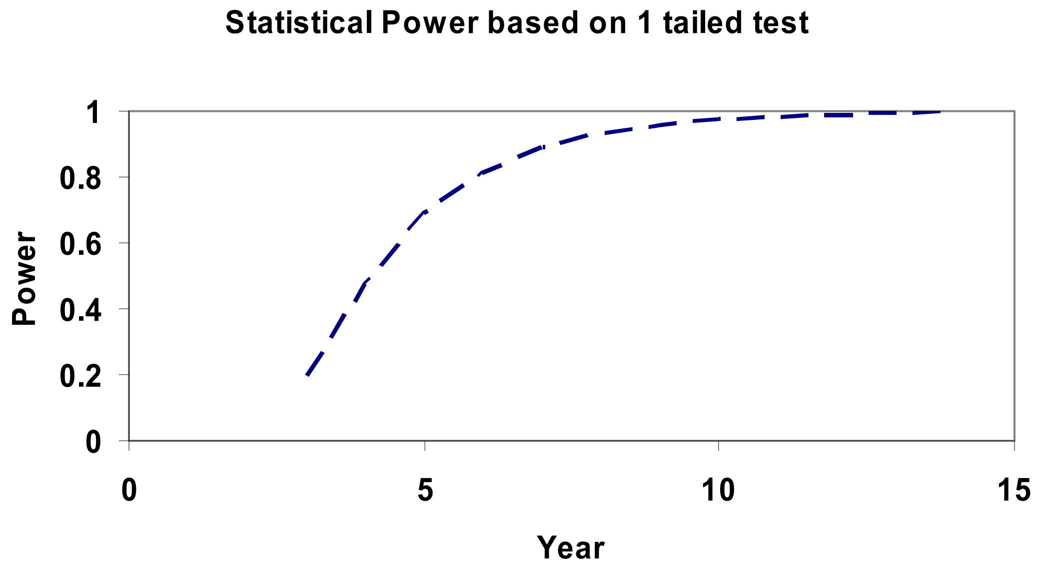


Figure 3. Statistical power obtained for 1-tailed tests of the null hypothesis that the indicators are not decreasing when the mean weight, mean length, mean maximum length and mean maximum weight indicators are modelled together with a common time effect. The same relative decline in indicators and value for alpha was presumed as those in the previous analyses (Fig. 2).

3. Discussion

3.1 Time Series analysis

A variety of time series analysis methods were applied to the candidate indicator datasets that were brought to the INDECO time series analysis workshop. The time series methods made it easier to visualize and analyse the time trends in the indicators by providing fitted time trends to each time series. The statistical TSA methods generally enabled computation of the uncertainty in the estimated trends and also the error variance associated with the estimated trends. As there were a variety of different TSA methods available, uncertainty in the estimated trends could also be accounted for by applying different TSA methods to the same time series. The TSA methods provided key inputs, i.e. plausible estimates of error variance and credible estimates of potential future rates of change, for statistical power analyses to detect trends in the indicators.

3.2 Statistical Power analysis

With conventional inputs for the level of significance (i.e. alpha), a two-tailed test, the statistical power analyses generally showed that power was low for detecting trends occurring within about 10 years or less. However, when uncertainty in the estimate of the residual error variance was taken into account with alternative methods to estimate error variance, alpha was increased from 0.05 to 0.1, and a one-tail test was applied to detect decreasing trends, the power increased substantially. However, it still took more than about 5 years to achieve high power tests. When the covariance in sets of related indicators was taken into account and statistical power analyses were conducted for sets of covarying indicators with the common time effects modelled, the power increased considerably and it was possible to obtain high power in about 5-6 years for some of the time series.

It appeared that at the workshop all of the participants became familiar with the time series modelling approaches if they had not been already. The computation of statistical power was new to most but within the workshop, all attendees were able to begin to appreciate the concept of statistical power and to compute power in tests for time trends in their indicators.

In workshop discussions, a few issues were raised with the estimation of statistical power for candidate indicators.

1. Detectable effect size. This is a key input for statistical power analysis. It reflects the magnitude of e.g. the trend that would be of interest to be able to detect in future tests for trends of concern in ecosystem indicators. For some quantities such as fish stock abundance, rates of decline that may be of concern to fisheries managers can be defined by historic rates of decline when the stock experienced serious declines in the past, or by population dynamics modelling. Many papers exist where the authors were able to define rates of decline in fish stocks that were of concern that were then applied in statistical power analyses (e.g., Gerrodette (1987); Maxwell and Jennings 2005). However, for many of the ecosystem indicators, for example, those for aggregate mean maximum length; it is difficult to formulate methods with which to identify rates of decline in the indicators that could be considered to be of concern. It would for example be quite difficult to undertake mathematical ecosystem modelling to identify rates of future decline that might be considered to be of concern when the indicator is composed of data from multiple species from different trophic levels in the ecosystem. Also, taking the maximum observed historic rates of decline from time series analysis (e.g., as in Nicholson and Jennings 2004) as the rate of decline of concern for power analysis might be either under estimating or over-estimating rates of decline that could signal undesirable ecosystem changes. The rate of decline of concern for example could vary with the absolute value of the indicator; a 10% rate of decline might not be a concern when the indicator was at moderate levels five years ago but could be of much greater concern if the indicator is at 75% of its value five years ago. It was urged at the workshop that further research effort be directed at formulating methods and protocols for identifying rates of decline of concern in the various candidate ecosystem indicators. But at present, it appeared to be acceptable to look for maximum rates of change in historic data and to formulate from that a range of plausible rates of change that could be utilized in statistical power analysis. This resulted in rather small effect sizes which were the main cause for a low statistical power in most instances.

2. One-tailed versus two-tailed tests. Some participants questioned whether it was acceptable to switch from a two-tailed test to a one-tailed test in evaluations of statistical power for candidate indicators. It was stated that in a number of instances, decisions on what to do in response to changes in indicators, could depend on whether either positive or negative trends were detected statistically. As such, two-tailed tests should be applied in statistical power analyses. However, it was argued by others that decision rules in a given time frame in a given ecosystem might be contingent upon detection of a change in one direction. Thus, if the ecosystem is already in a compromised state, a decision rule might be set up such that remedial action will be required if a further decrease in some ecosystem indicators is

statistically detected. Under such instances, where there is one particular direction of change that is of concern that will trigger a particular management action, a one – tailed test might be justified. The choice of the test thus depends on the null hypothesis to be tested and the null hypothesis to be tested in turn depends on the management question to be evaluated. It was agreed by all that clear scientific justifications will be required for the choice of a one-tailed test or a two-tailed test in evaluations of statistical power and that a two-tailed test should not necessarily be the default for statistical power tests.

3. Increasing the value for alpha from 0.05 to 0.1 in tests of statistical power. Within the discussions, it appeared that most could accept that it may be appropriate to modify the value assumed for alpha depending on the merits of the situation. For example, when the number of data that could be obtained is low, the error variance high relative to the effect or change that would be of interest to detect, and the cost of a type II error was very high (e.g., an undesirable change in the state of the ecosystem), an increase in the value for alpha might be justifiable. As this suggestion has long been in the literature and has seen acceptance in applied environmental science, no one at the workshop raised objections to considering using alpha of 0.1 in statistical power analyses of ecosystem indicators.

4. Choice of Software for statistically analysing indicators. Regarding the choice of software for evaluation of indicators, this seemed to be a matter of personal choice. However, it was generally agreed that the use of some statistical software package was essential. A wide variety of types of statistical analysis could be considered and simple non-statistical spreadsheet software such as Excel, simply did not have the capabilities to carry out the sorts of statistical analyses that could be appropriate such as ARIMA TSA modelling and multivariate statistical modelling. Overall, it appeared that all of the analyses that were applied at the workshop could be carried out in R and that batch files could be set up to evaluate large sets of data on ecosystem indicators. SAS also seemed to have most if not all of the desired capabilities. However, the help files for SAS did not seem to be so readily accessible and easy to interpret for applying statistical power analyses and multivariate tests. SPSS seemed to be set up to permit easy application of ARIMA models and univariate and multivariate time series modelling. However, it did not seem so accessible for setting up statistical power analyses based on the outputs from TSA.

5. Other analyses not covered in the workshop. Mulitvariate statistical methods for classifying sets of variables with similar properties such as principle components analysis were not touched upon in the workshop. Such methods have already been applied in evaluations of ecosystem indicators. However, the time available did not permit exploration of additional such statistical methodologies.

4. Summary

In summary the key conclusions of the workshop are as follows:

4.1 Time series analysis

1. Only simple ARIMA models with no more than 1 year lag effects in the AR and/or moving average components could be applied to most of the candidate indicator time series brought to the workshop.

- The time series for most candidate indicators are short, i.e. 15-30 years, and yet 50+ years are typically recommended for conventional ARIMA modelling.

2. Polynomial models described short time series (< 20 years) quite well.

   - The models provided an alternative method to characterize historical trends in indicators and compute residual error variance.
   - The risk of choosing a polynomial model that is over-parameterized (i.e. over fits the data) and under estimating residual variance can be reduced because the degrees of freedom (df) used to estimate the residual variance is reduced by the number of estimated parameters in the polynomial.

$$\text{resid var} = \frac{1}{N-p}\sum_{i=1}^{N}\left(y_i^{obs} - y_i^{pred}\right)^2$$

where y_i^{obs} is the observed value for the indicator and y_i^{pred} is the value for the indicator predicted by the polynomial, and p is the number of estimated parameters in the polynomial.

   - Furthermore, likelihood ratio tests or F-tests and AIC model choice criteria can be applied to choose the best fitting model taking into account the criterion of parsimony.

3. For most case studies, significant common time effects were found for two or more indicators. This has two important consequences.

   - A large “noise” component can be removed from the estimated time trend in each indicator and the estimated residual error variance can be reduced considerably.
   - This can make the underlying trend in the different indicators easier to detect and can give statistical tests for trends in indicators higher statistical power.

4. Loess smoothing is useful for characterizing time trends in indicators but not for estimating error variance, because Loess smoothing is not obtained by traditional statistical estimation.

4.2 Statistical power analysis

1. As in previous studies, we found low power in univariate analyses for 5-10 year trends in candidate ecosystem indicators.
2. Increasing alpha (chance of false positive) from 0.05 to 0.10 increased power noticeably.

- A value for alpha higher than 0.05 may be justifiable in statistical power analyses of candidate indicators due to high costs of failure to detect trends in the direction of further ecosystem deterioration.

3. Changing from a two-tailed test to a one-tailed test might be appropriate when it is of interest to test whether further deterioration has occurred. Doing so also increased power noticeably.

- It was agreed that there is need for clear justifications for the choice of a one-tailed or two-tailed test in a statistical analysis.
- e.g. there is only one reference direction that that will trigger remedial action.

4. We found considerably higher power could be obtained with multivariate time effect models and these were potentially applicable when sets of related indicators show covariance in temporal variation.

- Therefore, ability to detect trends can be improved by combining covarying indicators into the same statistical analysis.

5. Defining the magnitude of trends in candidate ecosystem indicators that would be of interest to detect for statistical power analysis is not straightforward.

- The maximum observed trend in historical time series might either smaller or larger than a future trend that would occur under future serious changes to the ecosystem. In the data available to the workshop, maximum observed trend was generally low, leading to a very low power of most tests.
- Rate of decline that might be of serious concern in indicators that are further reduced below current levels might be impossible to reliably determine even through ecosystem modelling.

5. References

Dette, H., A. Munk, and T. Wagner. 1998. Estimating the variance in non-parametric regression- what is a reasonable choice? Journal of the Royal Statistical Society. Series B. **60**:387-392.

Gerrodette, T. 1987. A power analysis for detecting trends. Ecology **68**:1364-1372.

Maxwell, D., and S. Jennings. 2005. Power of monitoring programmes to detect decline and recovery of rate and valuable fish. Journal of Applied Ecology **42**:25-37.

Nicholson, M. D., and S. Jennings. 2004. Testing candidate indicators to support ecosystem-based management: the power of monitoring surveys to detect temporal trends in fish community metrics. ICES Journal of Marine Science **61**:35-42.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Science **47**:2-15.

Trenkel, V.M. & M.J. Rochet (2003). Performance of indicators derived from abundance estimates for detecting the impact of fishing on a fish community. Canadian Journal of Fisheries and Aquatic Sciences 60, 67-85.

Segurado, S.P.B. 2006. Statistical power analysis of candidate system indicators. Final Project Report- University of the Algarve. Imperial College London 29 pp.

Toft, C. A., and P. J. Shea. 1983. Detecting community-wide patterns: estimating power strengthens statistical inference. American Naturalist **122**:618-625.

6. Appendices

Appendix 1 Workshop Announcement

INDECO workshop on time series analysis of candidate indicators

To be held January 24-26, 2006 at the following address:

Institute for European Environmental Policy (IEEP)
28 Queen Anne's Gate
London SW1H 9AB, UK

tel: 44-(0)20-77 99 22 44
Near to the St. James's Park Underground station on the District and Circle Lines

The workshop is to be lead by Drs. Murdoch McAllister and Pia Orr.

Workshop Goals and Activities

The candidate indicators to be analysed are those agreed to be compiled by region at the September 6-8, 2005 INDECO meeting in Gdynia. Please see the Gdynia meeting report or check with Indriani Lutchman or Gerjan Piet if you are still uncertain about precisely which datasets to prepare and have analysed for this workshop.

A variety of alternative time series statistical models will be identified to which the indicator time series can be fitted. Some of these methods will be univariate statistical models; others will be multi-variate time series models in which more than one candidate indicator time series can be statistically analysed at once. One goal of the analyses will be to characterize the temporal patterns in the time series, the error variability, and possible correlations with historic changes in fishing effort. The statistical power to detect trends in the indices will be evaluated using a few different methods to compute statistical power. Imperial College will prepare general statistical modelling routines for either R or S+, SAS, a few routines in SPSS, and WINBUGS. A few statistical power protocols will also be prepared in EXCEL. Participants can also bring their own routines. An effort will be made to reach a consensus among attendees on the formulation of appropriate protocols to compute the statistical power in candidate indices using either univariate or multivariate methods.

The workshop will be held from 9AM - 5pm on 24-26 January. There is capacity for about 12 attendees. If you wish to attend this workshop, please reply by e-mail at your earliest convenience.

<u>Notes for Participants</u>:
1) Please bring your own lap top computer
2) Please make sure your laptop has on it Excel with the data analysis tool pack and solver
3) Please bring text or Excel files containing your candidate indicator time series
4) Please ensure that you have on your laptop either R, S+, SAS or SPSS
5) Please make sure you have administrative privileges on your laptop computer

6) Please download WinBUGS version 1.4 with software key incorporated (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml)

DRAFT Time Table

Day 1

9 AM-1030AM

Welcome by INDECO Coordinator, Indriani Lutchman
Introduction to workshop by M. McAllister
Introductions of participants and their datasets (each participant be prepared to introduce themselves and their indicator time series for no more than 5 minutes each)

1030-1045 AM Coffee Break

1045-1230AM

Session on fitting univariate time series models to data

- Linear and log-linear models
- Polynomial models
- ARIMA models
- Lowess smoothing

Software: EXCEL, SPSS, SAS, R, WinBUGS

1230-1330 PM Lunch

1330-1500

Session on diagnostics and model selection for fitting TS models to data
Residual Analysis
Auto correlation plots and tests
R^2
AIC
Likelihood Ratio Tests
F-tests

1500-1515 Coffee Break

1515-1700

Estimation of slope and rate of change in indicators
Methods for estimating the residual variance from univariate time series models

Day 2: Statistical Power Analysis

9AM-1030AM

Seminar: Statistical Power Methods for INDECO: M. McAllister
Group discussion of conventions for computing statistical power

1030-1045AM Coffee Break

1045-1230

Statistical Power computations for a simple univariate linear model in Excel
One-tailed tests versus two-tailed tests
The effects of alpha, error variance, number of years, effect size

1230-1330 Lunch

1330-1500

Statistical Power computations for a simple univariate log linear model in Excel
Statistical Power computations using Monte Carlo Simulation in Visual Basic

1500-1515 Coffee Break

1515-1700

Univariate statistical power analyses of various candidate indicators

Day 3: Multivariate time series analysis

9AM-1030AM

Seminar: Multivariate TS Methods for INDECO: M. McAllister
Group discussion of conventions for Multivariate TS Methods

1030-1045AM Coffee Break

1045-1230

Fitting Multivariate statistical models to candidate indicator series
Computing statistical power of multivariate time series models

1230-1330 Lunch

1330-1500

Evaluation of correlations between indicators and measures of fishing intensity

1500-1515 Coffee Break

1515-1700

Discussion and conclusions regarding time series analysis of candidate indicators

Appendix 2 Workshop overhead slides

INDECO workshop on time series analysis of candidate indicators

Murdoch McAllister, Pia Orr, Susana Segurado & Tom Carruthers
Division of Biology
Imperial College London

Jan. 24-26, 2006

IEEP
London

Why is it appropriate to apply time series analysis to candidate indicator time series?

1. *to be able to formulate and test hypotheses about time trends and other temporal patterns in candidate indicators*
2. *to obtain an improved understanding of the statistical properties of candidate indicators*
3. *to formulate appropriate methods to compute the statistical power of tests for trends in candidate indicators*
4. *to formulate recommendations about which sets of indicators may have suitable statistical properties*
5. *to formulate recommendations about possible modifications to survey design to improve indicators*

Aims of workshop

- formulate, discuss and apply candidate time series analysis (TSA) methods to candidate indicator time series
- agree on suitable general procedures for TSA of candidate indicators
- agree on and apply suitable protocols for statistical power analysis in tests for time trends in candidate indicators
- agree on and apply suitable protocols for multivariate statistical methods for TSA and statistical power analysis
- to provide training in some basic principles and methods of TSA
- to provide inputs to INDECO deliverables (e.g., 5.2, 5.3)

Workshop outline

Tuesday 24 January: Univariate time series analysis

1. **Review of alternative univariate TSA methods**
2. **Diagnostics**
3. **Model selection**
4. **Estimation of temporal patterns in candidate indicators**

Wednesday 25 January: Statistical power analysis

1. **Introduction to statistical power analysis**
2. **Statistical power computations**
   - **Excel**
   - **Visual basic**
3. **Univariate statistical power analysis of candidate indicators**

Thursday 26 January: Multivariate time series analysis

1. **Review of Multivariate TS Methods for INDECO**
2. **Fitting Multivariate statistical models to candidate indicator series**
3. **Computing statistical power of multivariate time series models**
4. **Evaluation of correlations between indicators and measures of fishing intensity**
5. **Discussion and conclusions regarding time series analysis of candidate indicators**

Univariate time series analysis (TSA) models

- **There is only a single independent variable in the model,**
- **e.g. Univariate linear model**

$$Y_i = m X_i + b$$

where

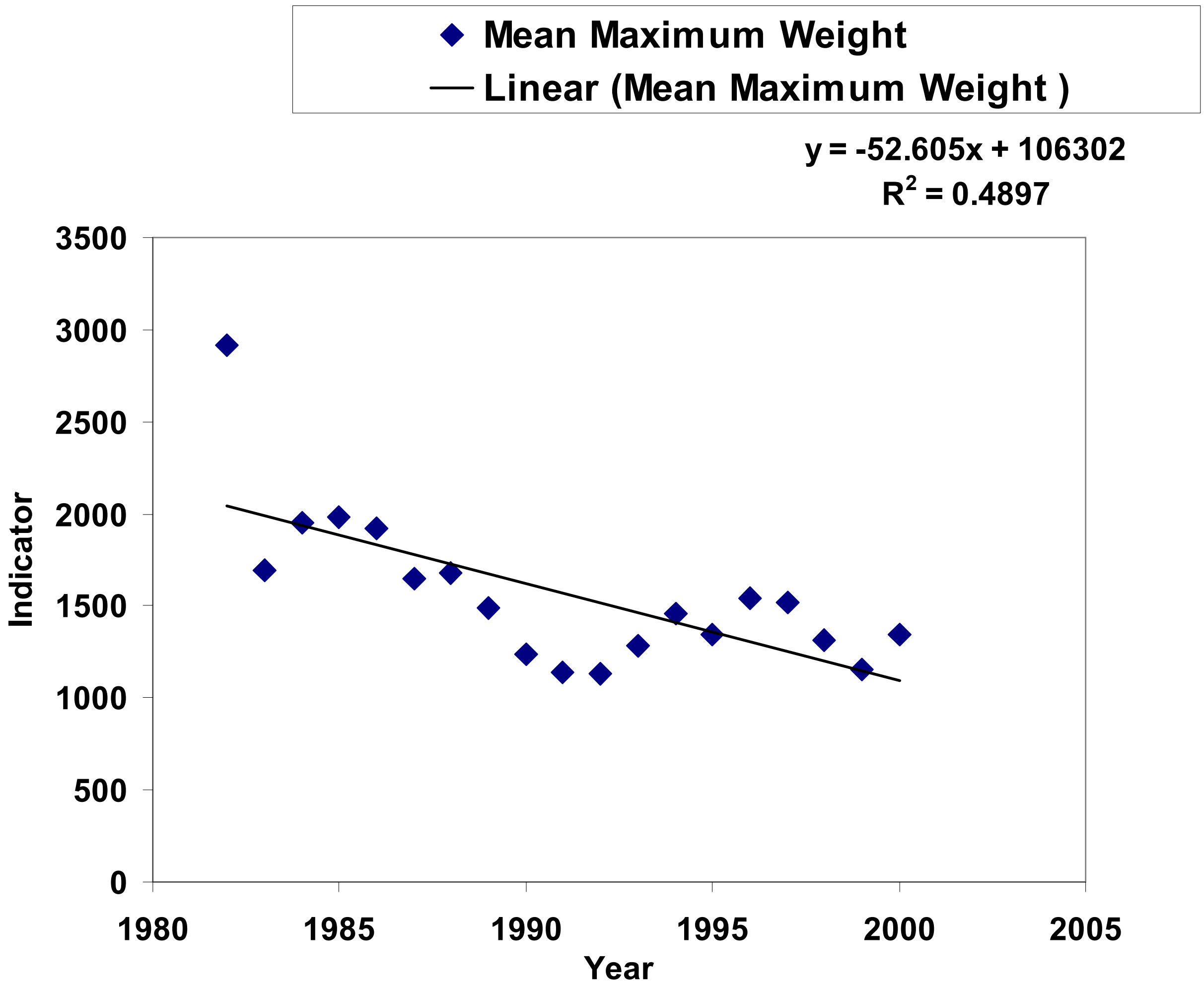
Y_i is the dependent variable (candidate indicator) at time step i

m is the slope parameter

X_i is the time elapsed

b is the intercept parameter

Mean Maximum Weight (Nicholson & Jennings 2004) on time



Univariate log linear model

Y_i = a * exp((X_i – X_0) * d)

where

Y_i is the independent variable (candidate indicator) at time step i

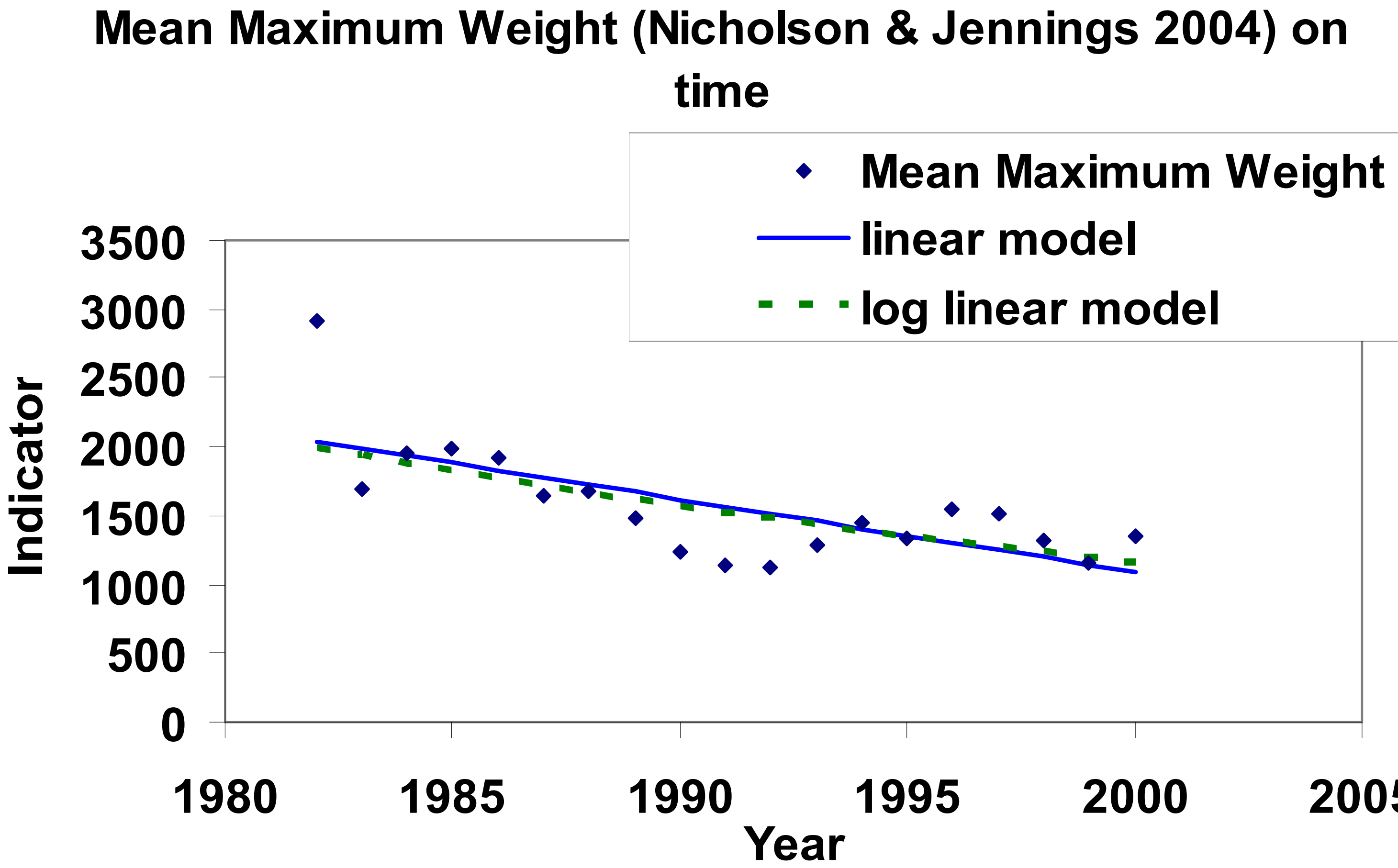
a is the log linear slope parameter

X_i is the time elapsed since initial time X_0

d is the estimated value of Y at the start of the series

Can do a linear regression analysis of log transform:

ln(Y_i) = ln(a) + (X_i – X_0) * d


Mean Maximum Weight (Nicholson & Jennings 2004) on time
Mean Maximum Weight
linear model
log linear model
Indicator
3500
3000
2500
2000
1500
1000
500
0
1980
1985
1990
1995
2000
Year

Key features of the linear and log linear TSA models

	Linear Model	Log linear model
Temporal pattern	Follows constant linear change	Follows constant exponential or proportional change
Interpretation	$Y_{n+1} – Y_n$ = constant	Y_{n+1} / Y_n = constant
Negative values of Y?	Indicator can be negative	Indicator can only be positive
Distribution for deviates	Normal distribution	Log normal distribution
Application	Short – moderate time series Statistical power analysis	Short-moderate time series Statistical Power analysis

Polynomial model

- **Has general form:**

$Y_i = a_n X_i^n + a_{n-1} X_i^{n-1} + a_{n-2} X_i^{n-2} + \ldots + a_1 X_i + a_0$

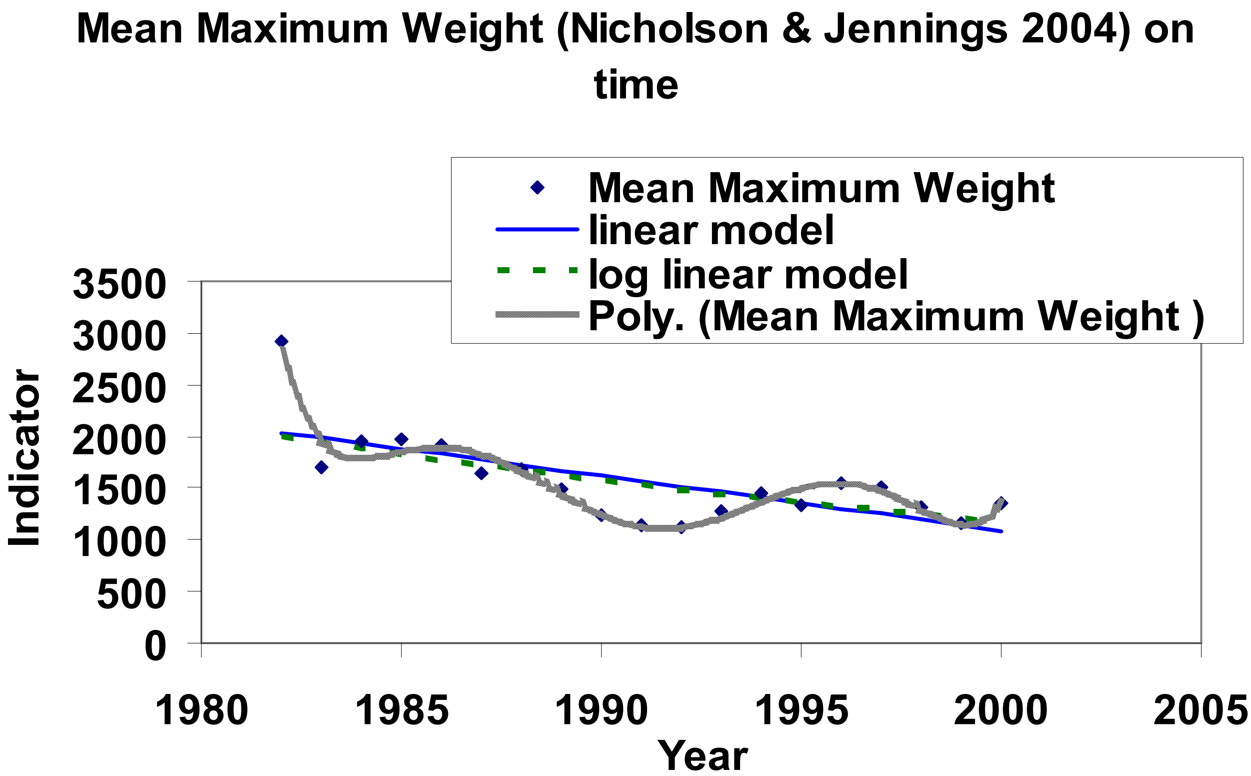
where a are the parameters of the n^{th} order polynomial model.

e.g. 2^{nd} order polynomial is a parabolic model:

$Y_i = a_2 X_i^2 + a_1 X_i + a_0$

- **Common to fit 3^{rd} and larger order polynomials to relatively short time series**

Fit of a 6th order polynomial to a 20-year time series



Polynomial Models

- ***provide a purely descriptive or explanatory time series model***
- ***can obtain a much better fit to the data than a linear model***
- ***risk over-fitting of the model to the data to obtain artificial temporal patterns***
- ***have little or no predictive power, even less than linear models***
- ***can be used to quantify error variance for future indicator values***

Exercise 1.1 Fitting linear and polynomial models to data

Aims

1. Fit linear, log-linear and a variety of polynomial models to your time series
2. Evaluate whether these models can adequately describe any of your time series

Methods

1. Use software of your choice to do this analysis
2. Or try using Excel trendline chart & data analysis

Outputs (on one or two candidate time series)

1. Plot of fit of each model tried to time series
2. Show models obtained with parameter estimates
3. Show R^2, p-values for parameter estimates

4. Take about forty minutes to do analysis

Conclusions from linear model and polynomial model TSA

1. Do any of the models adequately describe any of your time series?
2. What criteria did you use to decide whether a model fitted the data?
3. What time trends if any are indicated by the various models?
4. Do different models suggest different temporal patterns?

Auto Regressive Models

1. This year's observation Y_t is a linear function of

- ***the long term average value (u) plus***
- ***some of the most recent observations***

  (e.g., Y_{t-1}, Y_{t-2}, Y_{t-3}, …)

2. the number of recent observations included is denoted by p

- ***p is usually 1, 2, or 3***

3. the error deviates in observations

- ***have mean zero,***
- ***are independent of each other,***
- ***the error variance is stationary over time***

Auto Regressive Model for p = 1 (AR(1)):

$Y_t = (u + (g_1 (Y_{t-1} – u))) + a_t$

where u is the long term average value for the set of observations Y

g_1 is the slope coefficient for observation Y at lag of 1 year.

a_t is the error deviation between the observation (Y_t) and its predicted value $(u + (g_1 (Y_{t-1}-u)))$ for year t

Auto Regressive Model for p = 2 (AR(2)):

$Y_t = (u + g_1 (Y_{t-1} - u) + g_2 (Y_{t-2} - u)) + a_t$

where u is the long term average value for the set of observations Y

a_t is the error deviation between the observation Y_t and its predicted value $(u + g_1 (Y_{t-1}-u) + g_2 (Y_{t-2}-u))$ for year t

g_i is the slope coefficient for the observation Y at lag of i years.

Moving Average Models (MA(q))

1. ***This year's observation Y_t is a linear function of***
   - ***the long-term average value (u) plus***
   - ***some of the most recent deviations from predicted values for recent observations (e.g., a_{t-1}, a_{t-2}, a_{t-3}, …)***
2. ***the number of recent observations included is denoted by q***
   - ***q is usually 1, 2, or 3***
3. ***the error deviates in observations***
   - ***have mean zero,***
   - ***are independent t of each other and***
   - ***the error variance is stationary over time***

Moving Average Model for q = 1 (MA(1)):

$$Y_t = (u + h_1 a_{t-1}) + a_t$$

where u is the long term average value for the set of observations Y

h_1 is the slope coefficient for annual deviate at lag of 1 year (a_1).

a_t is the error deviation between the observation (Y_t) and its predicted value ($u + h_1 a_{t-1}$) for year t

Moving Average Model for q = 2 (MA(2)):

$$Y_t = (u + h_1 a_{t-1} + h_2 a_{t-2}) + a_t$$

where u is the long term average value for the set of observations Y

a_t is the error deviation between the observation Y_t and its predicted value ($u + h_1 a_{t-1} + h_2 a_{t-2}$) for year t

h_i is the slope coefficient for annual deviate at lag of i years.

Differencing (d) in AR and MA models

- **appropriate when the process is non-stationary, i.e., there appear to be trends in the time series**
- **e.g. differencing at d = 1**

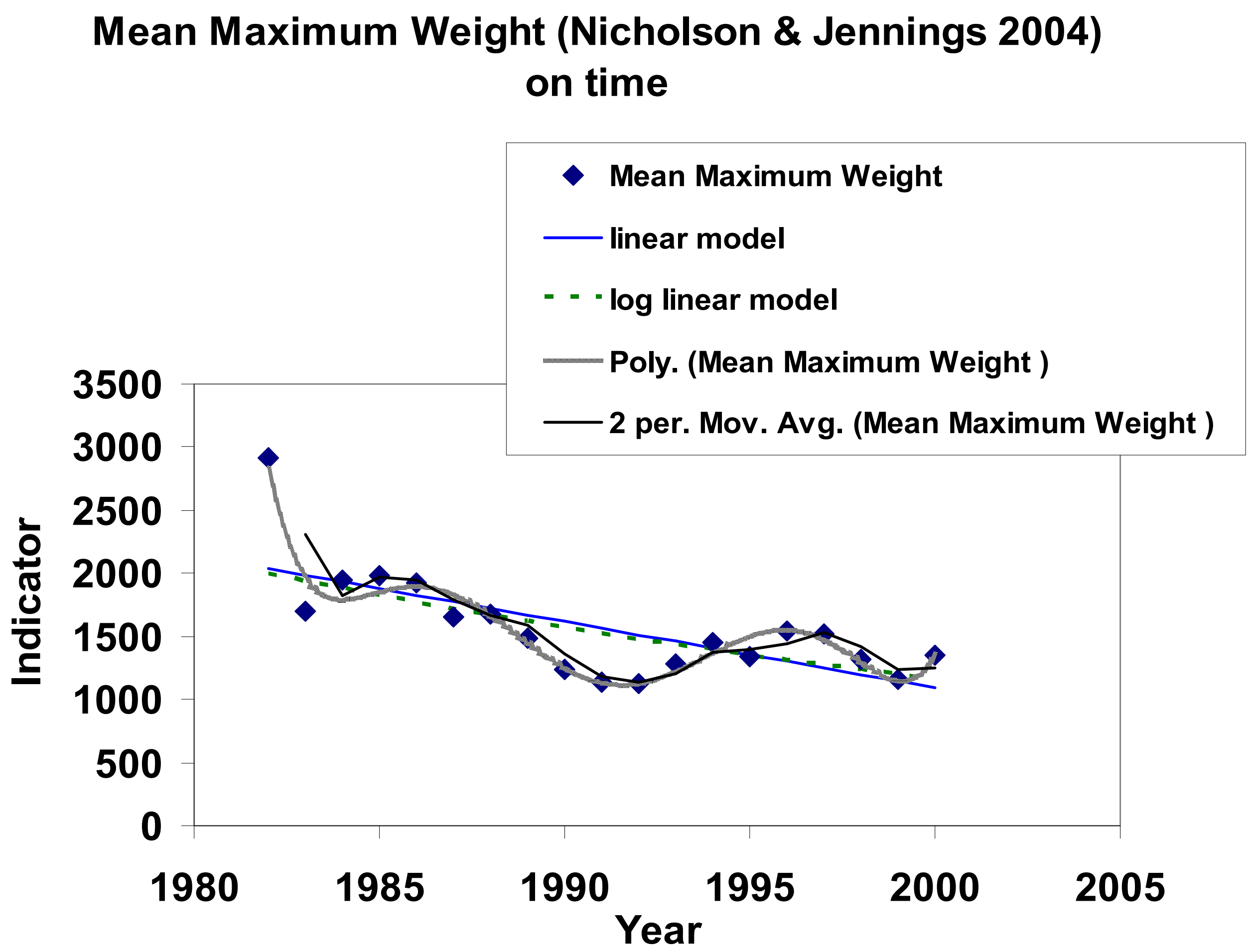
  $Z_t(1) = Y_t – Y_{t-1}$
- **most common to have d = 1, i.e., differencing at lag of 1, but could try other values for d.**
- **http://www.eng.usf.edu/~argangul/TSSP%20Lecture%204_files/frame.htm**

ARIMA(p, d, q) models

- **combination of AR(p) and MA(d) models with differencing at lag d.**
- **d = 0 means no differencing in the observations**
- **p = 0 means no AR model**
- **q = 0 means no MA model**
- **also an option for a constant (u) to be included or not**
- **ARIMA(1, 1, 1) would mean difference the observations Y at a lag of 1, and apply in the same time series model both the AR(1) and MA(1) models:**

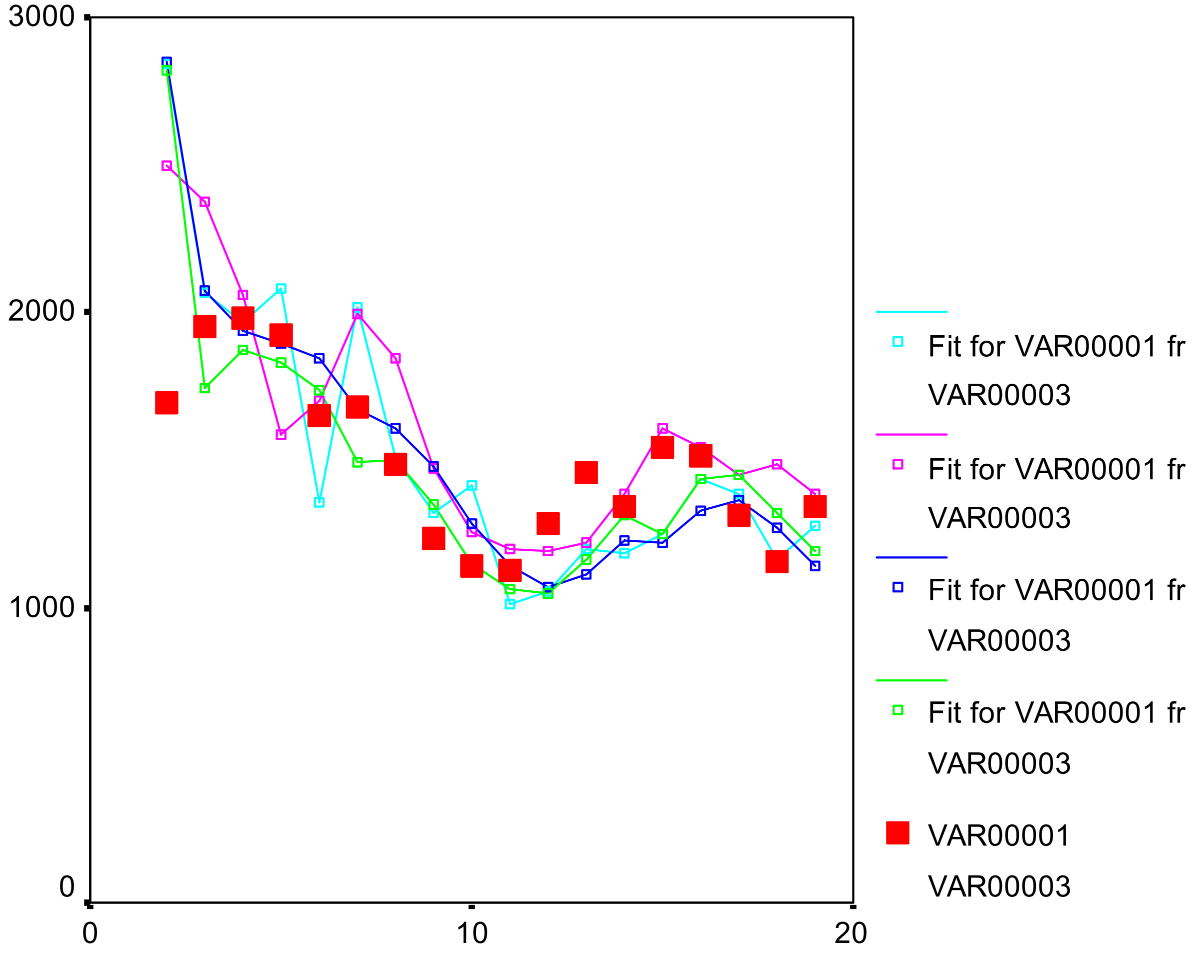
$$Z_t = (u + g_1 (Z_{t-1} - u) + h_1 a_{t-1}) + a_t$$

- **Most statistical software allows TSA with ARIMA(p, d, q) models**



Mean Maximum Weight (Nicholson & Jennings 2004)
on time
Mean Maximum Weight
linear model
log linear model
Poly. (Mean Maximum Weight)
2 per. Mov. Avg. (Mean Maximum Weight)
Indicator
3500
3000
2500
2000
1500
1000
500
0
1980
1985
1990
1995
2000
2005
Year

Fits of various ARIMA(p,d,q) models to Nicholsen and Jennings (2004) Mean Maximum Weight Data using SPSS



Lowess Smooth Functions

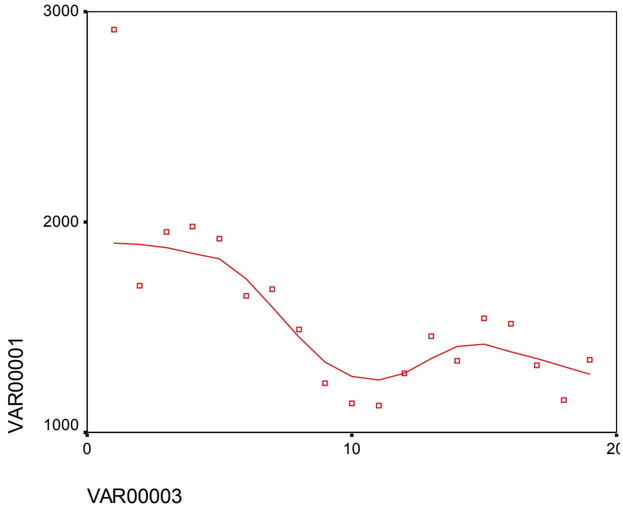
- **“Locally weighted least squares”**
- **Carries out locally weighted time series and scatter plot smoothing**
- **For both equispaced and non-equispaced data**
- **Analyst can vary the size of the smoothing window (0 to 1, with default usually set at 0.5)**
- **Available in several stats packages.**
- **Utilized e.g. in Nicholson & Jennings (2004) to identify slope values at various points in time for various indicators series.**

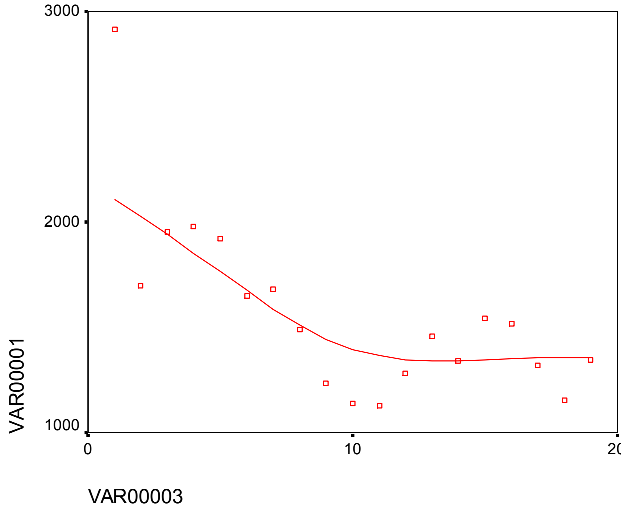
Lowess smooth fitted to N&J (2004) MMW data

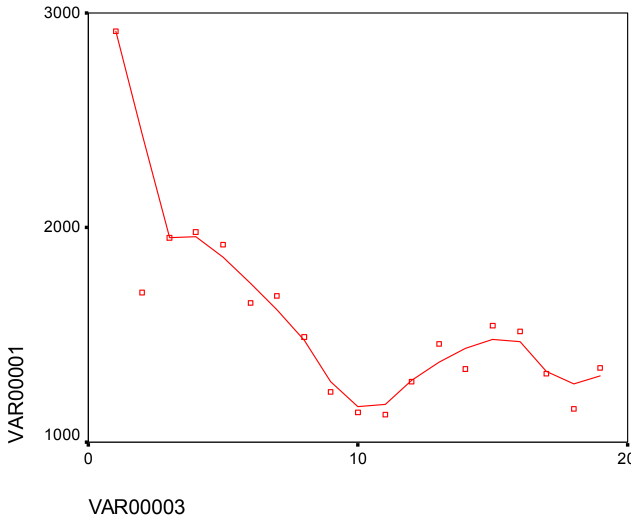
Percentage of points to fit varied

default = 50% 75%

25%







Exercise 1.2 Fitting ARIMA and Lowess models to data

Aims

1. **Fit ARIMA and Lowess smooth models to your time series**
2. **Evaluate whether these models can adequately describe any of your time series**

Methods

1. **Use software of your choice to do this analysis**
2. **Or try using SPSS time series analysis ARIMA option**
3. **For ARIMA models try at least the following:**
   - **(1,0,0), (0,0,1), (1,0,1), (0,1,0), (1,1,0), (0,1,1), (1,1,1)**
   - **try these models with and without constant also**
4. **For Lowess smooth models try varying the smoothing parameter**

(percentage of points to fit) from 0.5 to 0.75 to 0.25.

Outputs (on one or two candidate time series)

1. **Plot of fit of each model tried to data time series**
2. **Show models obtained with parameter estimates**
3. **Show where available**
   - **AIC, p-values for parameter estimates**
   - **Correlations between parameter estimates**
4. **Take about one hour to do analysis**

Conclusions from previous and this ARIMA/ Lowess TSA

1. Do any of the ARIMA models adequately describe any of your time series?
2. What criteria did you use to decide whether a model fitted the data?
3. What time trends if any are indicated by the various models?
4. Do different ARIMA, Lowess, linear, and polynomial models suggest markedly different temporal patterns?
5. What are potential advantages of ARIMA models over Lowess, linear or polynomial models?

Diagnostics & Model Selection of TSA Models

Problems addressed:

- ***Infinite variety of alternative TSA models that could be fitted***
- ***Different models may suggest different temporal patterns and lead to different decisions***
- ***The goodness of fit of different models is difficult to assess just by eye-balling the data***

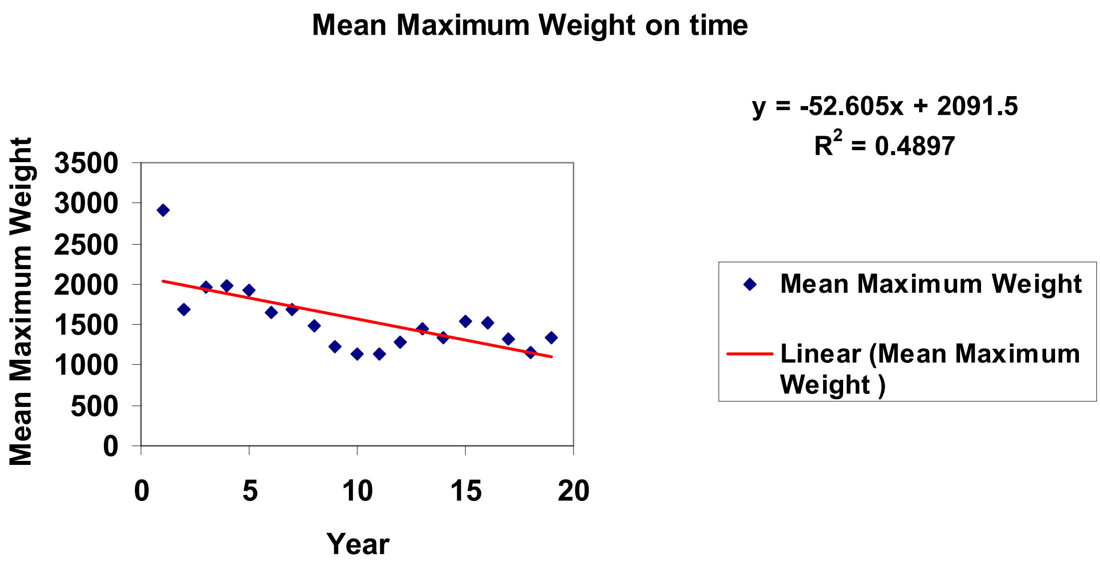
Questions

- ***What criteria should be used to assess whether to accept or reject a particular TSA model?***
- ***Are there some formal methods to test hypotheses about whether one model fits the data better than another model?***

Residual Analysis

Residual at time t:

$e_t = Y_t - \text{prediction}(Y_t)$



where

Y_t is the observation at time t

prediction(Y_t) is the TSA model prediction of Y_t

Statistical assumptions; residuals

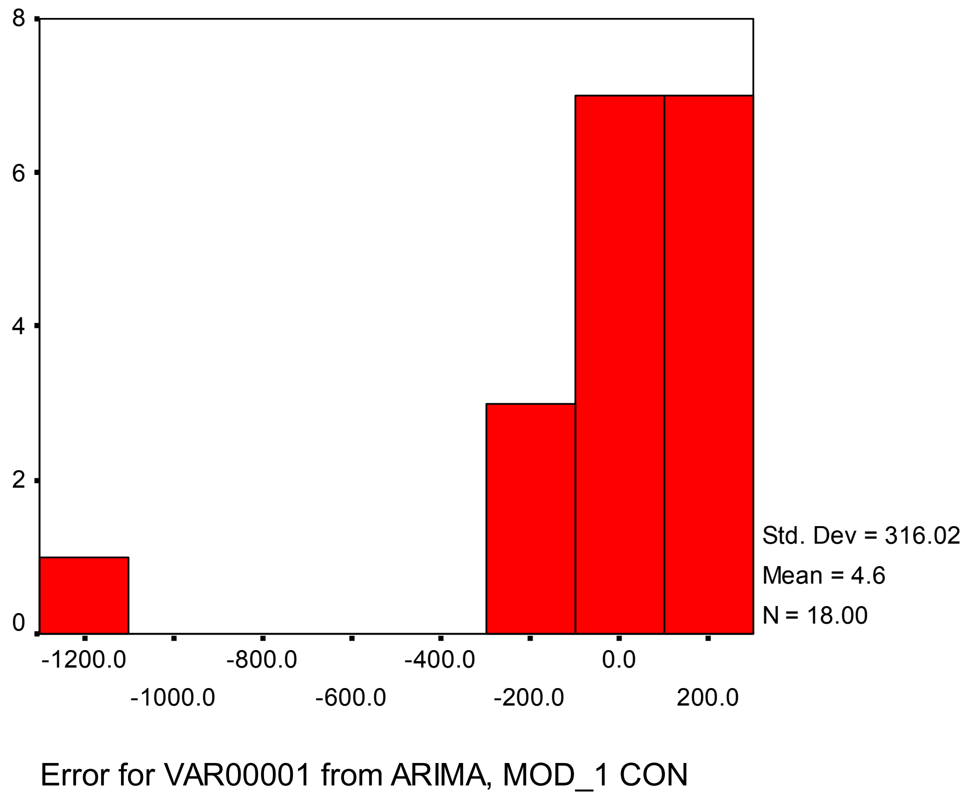
- ***are not auto-correlated, i.e., are independent***
- ***are normally distributed***

- ***variance in residuals is constant over time***

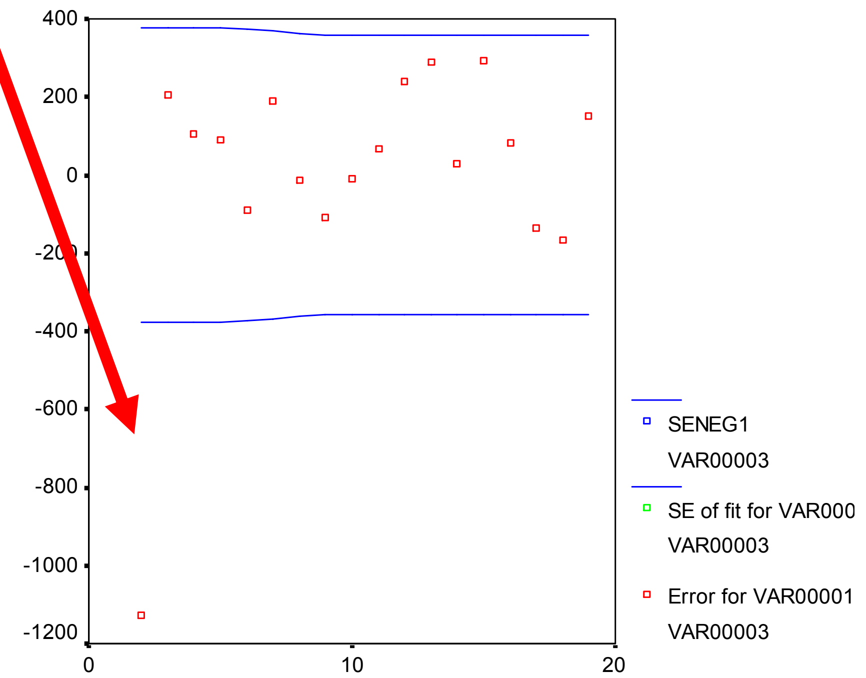
Procedures to analyse residuals

1. Produce a histogram of the residuals

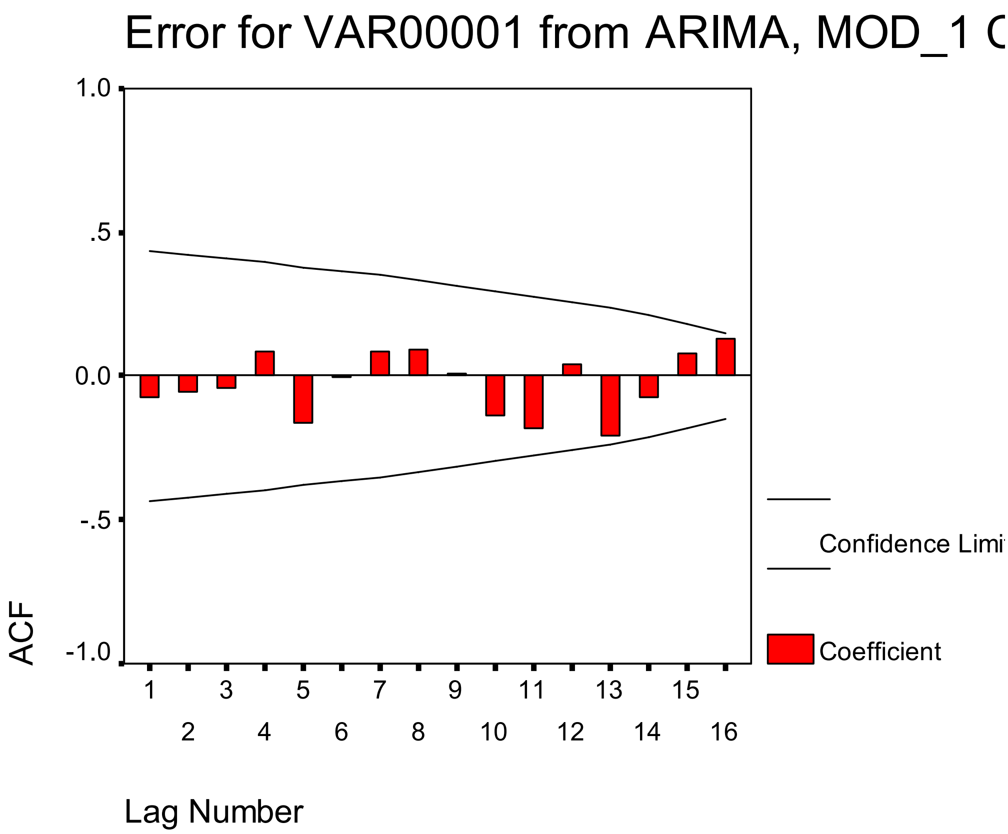
- ***is the distribution symmetric?***
- ***does it appear normally distributed?***
- ***are there apparent outliers?***



2. *Plot standardized residuals over time (i.e. divide residuals by the standard error in residuals) and compare with plots of ±1 standard error*
3. *If an outlier exists (> 2-3SE units from zero), e.g., at the very beginning of the time series, it may be prudent to conduct the TSA both with and then without it*



3. Test whether autocorrelation exists in the residuals from a given model fit



A first-order autoregression, AR(1), has equation

$$X_t = \mu + \alpha (X_{t-1} - \mu) + e_t ,$$

where $\{e_t\}$, the innovations, are zero-mean, uncorrelated, variance σ^2_e.

A first-order moving average, MA(1), is

$$X_t = \mu + e_t + \beta e_{t-1} .$$

For AR(1) with parameter α the ACF is $\rho_k = \alpha^k$.

For MA(1) with parameter β we have

> **$\rho_k = 0$ except for $\rho_0 = 1$ and $\rho_1 = \beta/(1+\beta^2)$.**

This distinguishes the MA(1) from the AR(1):

> **for AR(1) only the effects of the early observations continue to be felt.**

http://www.staff.city.ac.uk/r.j.gerrard/courses/3ts/ts1.htm

Partial ACF

This suggests a technique for identifying a first-order MA:

- **See if the ACF is close to 0 except at lag 1.**

Checking whether a decrease is 'close to geometric' (AR(1)) is much harder.

- **The *partial ACF* (PACF) was introduced to combat this.**
- **A partial *auto*correlation is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all *lower-order*-lags.**

A few rules to determine ARIMA model structure

"Rule 6:

- **If the PACF of the differenced series displays a sharp cutoff**
- **and/or the lag-1 autocorrelation is positive**
- **i.e., if the series appears slightly "underdifferenced"**
- **then consider adding an AR term to the model.**
- **The lag at which the PACF cuts off is the indicated number of AR terms".**

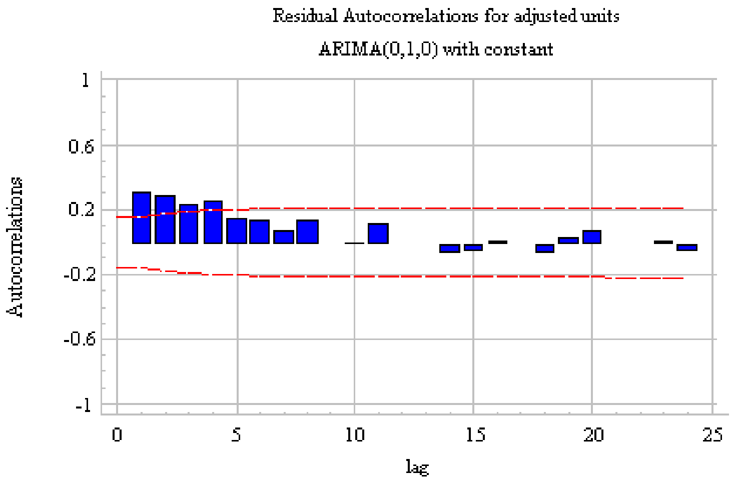
  http://www.duke.edu/~rnau/411arim3.htm

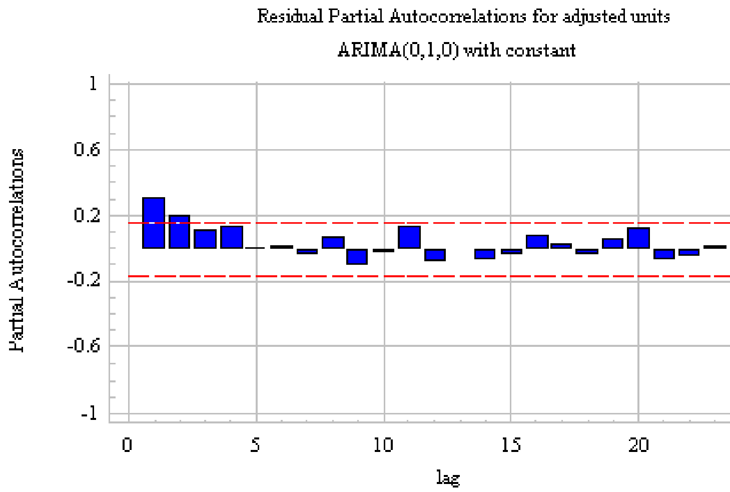
"Rule 7:

- **If the <u>ACF</u> of the differenced series displays a sharp cutoff**
- **and/or the lag-1 autocorrelation is <u>negative</u>**
- **i.e., if the series appears slightly "overdifferenced"**
- **then consider adding an <u>MA</u> term to the model.**
- **The lag at which the ACF cuts off is the indicated number of MA terms".**

  http://www.duke.edu/~rnau/411arim3.htm

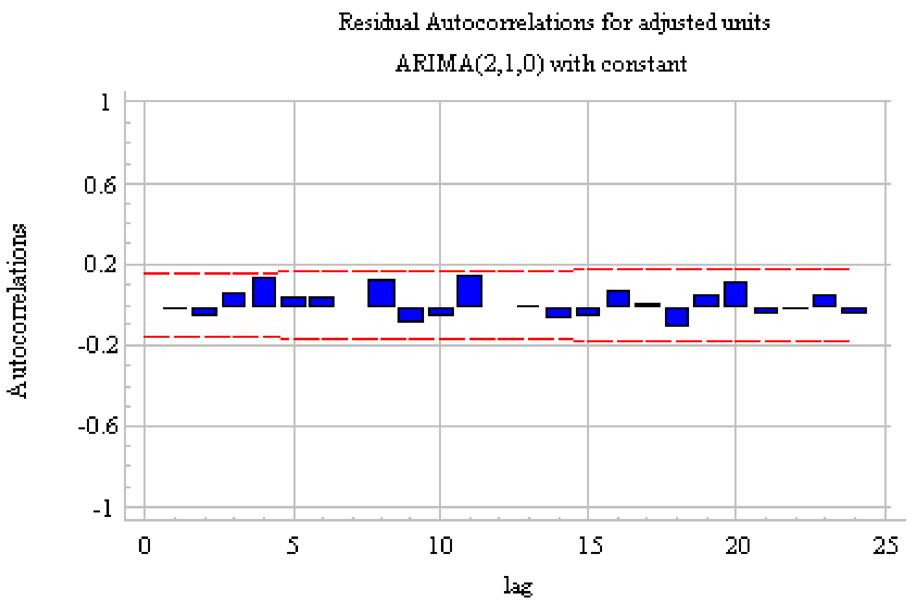
A model for the UNITS series--ARIMA(2,1,0): Previously we determined that the UNITS series needed (at least) one order of nonseasonal differencing to be stationarized. After taking one nonseasonal difference--i.e., fitting an ARIMA(0,1,0) model with constant--the ACF and PACF plots look like this:

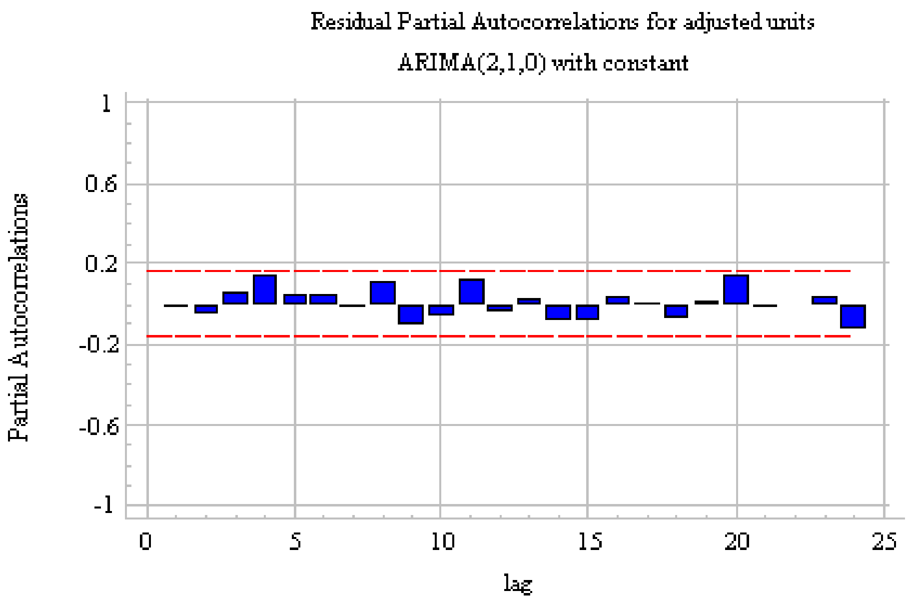




Notice that (a) the correlation at lag 1 is significant and positive, and (b) the PACF shows a sharper "cutoff" than the ACF. In particular, the PACF has only two significant spikes, while the ACF has four. Thus, according to Rule 7 (6?) above, the differenced series displays an AR(2) signature.

If we therefore set the order of the AR term to 2--i.e., fit an ARIMA(2,1,0) model--we obtain the following ACF and PACF plots for the residuals:



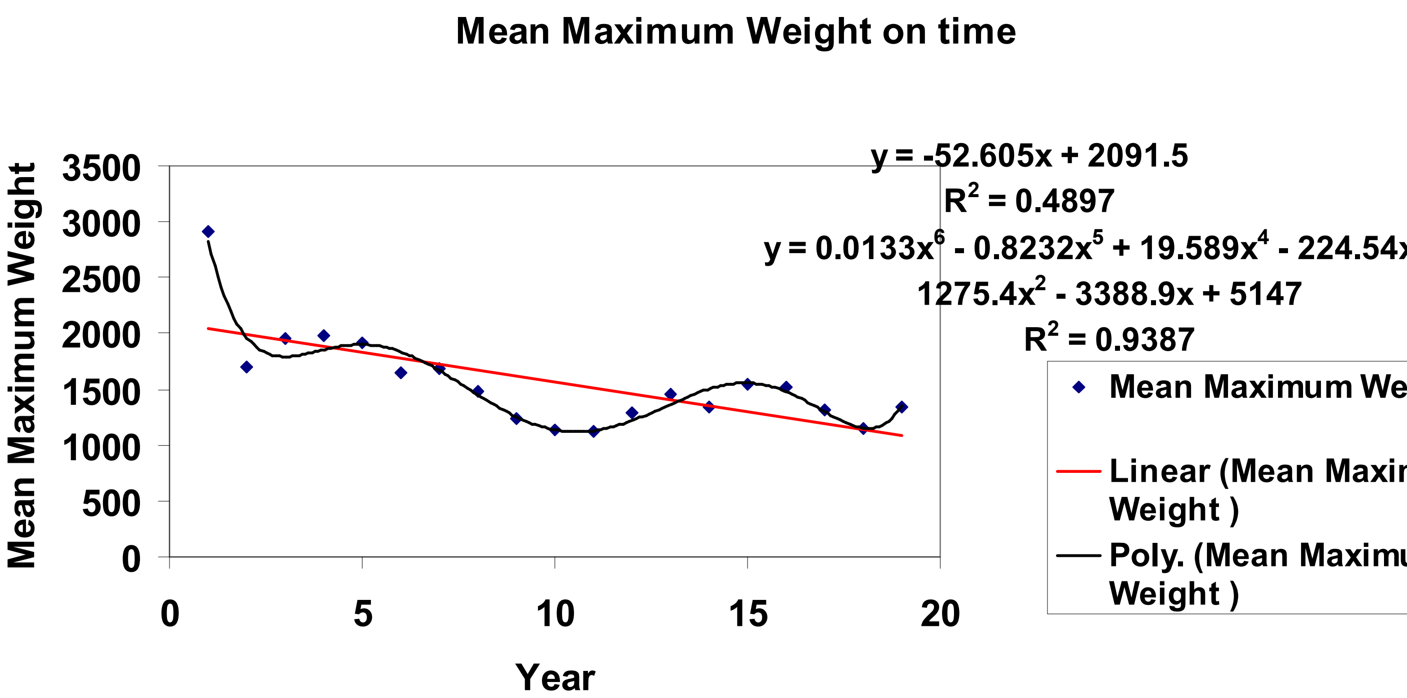


The autocorrelation at the crucial lags--namely lags 1 and 2--has been eliminated, and there is no discernible pattern in higher-order lags.

Other diagnostics

1. R^2 - indicates the fraction of variance explained

- **not always provided by software**
- **ranges between 0 and 1, with higher values better**
- **high values can be misleading if the model is overparameterized (has estimated too many parameters)**



2. P-values for estimated parameters

- **The p-value provided can be used to evaluate whether the model is overparameterized**
- **if p-value > alpha for a given parameter this suggests that the parameter could be eliminated from the model**

```
Variables in the Model:

                         B           SEB
T-RATIO   APPROX. PROB.

AR1             1.513240      .439493
3.4431516          .00436513
AR2             -.688075      .404861    -
1.6995345          .11300520
MA1             1.781771    34.764794
.0512522          .95990374
MA2             -.999009    39.059125     -
.0255768          .97998331
CONSTANT      -93.525493    90.099210    -
1.0380279          .31817983
```

- **Here, with alpha = 0.05, AR2, MA1, MA2, and possibly constant could be eliminated**
- **Since p-value for AR2 is close to alpha, AR2 might be kept**

2. Correlations between estimated parameters

- **if high (e.g., r> 0.9 or r < -0.9), the model may be overparameterized and might be reduced**

ARIMA(2,1,2)

Correlation Matrix:

	AR1	AR2	MA1	MA2
AR1	1.0000000	-.7521010	.2938497	-.2884104
AR2	-.7521010	1.0000000	-.5814676	.5805937
MA1	.2938497	-.5814676	1.0000000	-.9999693
MA2	-.2884104	.5805937	-.9999693	1.0000000

AIC – Akaike Information Criterion

- **Provides a measure of goodness of fit of a model to the data that also penalizes models with more parameters**
- **AIC(M_1) = -2 * (Log Likelihood (m_1))
  + 2 (number of estimated parameters)**

where

Log Likelihood (m_i) is the log of the statistical likelihood of model i (m_i), and

Interpreting AIC

- **Two different models must be fitted to the same data**
- **The two models do not need to be nested versions**
- **The model with lowest AIC is considered best**
- **Differences in AIC units of about 3 or more are considered by convention to be meaningful**
- **There is no probabilistic interpretation of AIC**
- **AIC is computed by most statistical softwares, e.g., SPSS for ARIMA and other mod**

ARIMA(2,1,2)

Standard error	335.00556
Log likelihood	-128.44782
AIC	266.89563

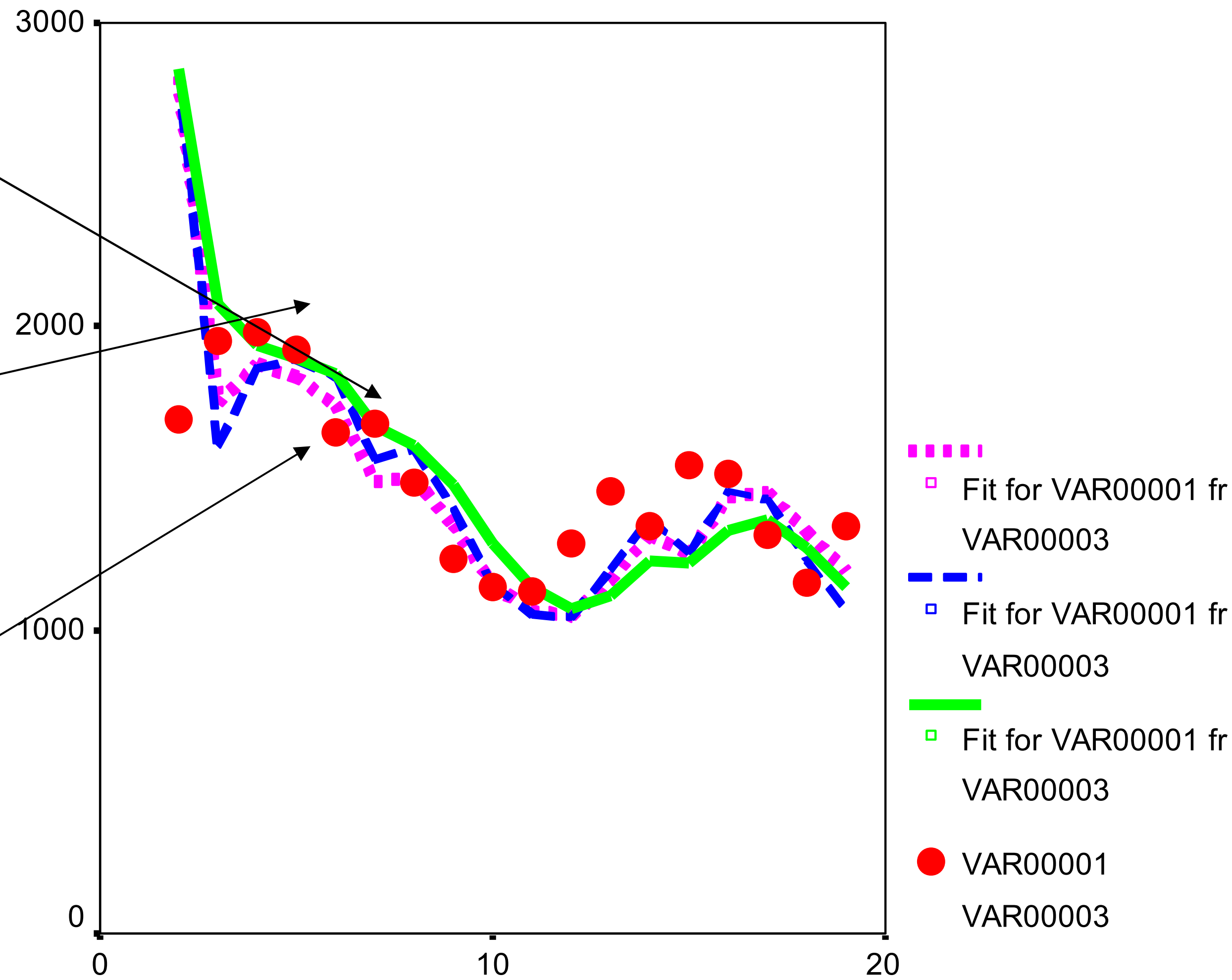
ARIMA(0,1,1)

Standard error	317.46885
Log likelihood	-128.34074
AIC	260.68149

ARIMA(0,1,0)

Standard error	327.27535
Log likelihood	-129.27533
AIC	260.55065

(constant fitted in all models)



Likelihood Ratio Tests

- **Can be used to test whether the less complex model in a pair of two nested models can be rejected in favour of the more complex model**
- **Takes the log likelihood of each model and formulates a chi square statistic based on the difference:**

- **ChiSq(df) = 2 * (Log Likelihood (m_1) – log Likelihood (m_2))**

where

Log Likelihood (m_i) is the log of the statistical likelihood of model i (m_i), and

model 1 has more parameters than model 2.

df = degrees of freedom = dif in number of estimated pars

e.g.., can we reject ARIMA(0,1,1) in favour of ARIMA(0,1,2)?

H_0: accept ARIMA(0,1,1) over ARIMA(0,1,2)

H_A: reject ARIMA(0,1,1) in favour of ARIMA(0,1,2)

ARIMA Model	0, 1, 2	0,1,1
Standard error	324.1131	317.46885
Log likelihood	-128.26786	-128.34074
AIC	262.53572	260.68149

Test Statistic:

Chi sq(1) = 2*(-128.26786-(-128.34074)) = 0.14576

Critical Value:

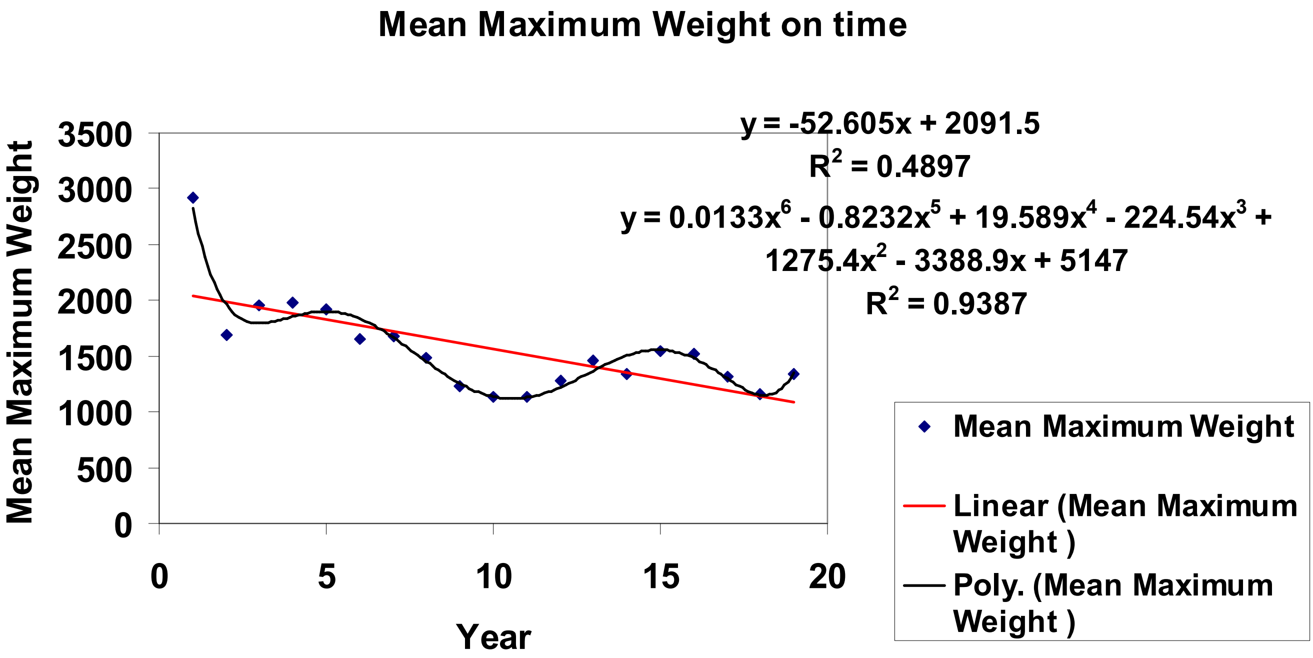
Chi sq(1, 0.95) = 3.84

Conclusion:

Test statistic is less than critical value

Therefore we fail to reject the null hypothesis and accept the ARIMA(0,1,1) model

Can we reject a linear model in favour of a sixth order polynomial?


Mean Maximum Weight on time
y = -52.605x + 2091.5
R2 = 0.4897
y = 0.0133x6 - 0.8232x5 + 19.589x4 - 224.54x3 + 1275.4x2 - 3388.9x + 5147
R2 = 0.9387
Mean Maximum Weight
3500
3000
2500
2000
1500
1000
500
0
0
5
10
15
20
Year
Mean Maximum Weight
Linear (Mean Maximum Weight)
Poly. (Mean Maximum Weight)

Likelihood Ratio Test

H_0: accept linear model over 6th order polynomial model

H_A: reject linear model and accept 6th order polynomial model

Log likelihood 6th order polynomial: -98.23

Log likelihood of linear model: -117.55

Test statistic

Chi-square statistic = 2 * (-98.23 –(-117.55)) = 38.64

Critical Value

Chi-square(df = 5, 0.05) = 11.07

Conclusion:

reject Null hypothesis

Accept the polynomial model

Exercise 1.3 Diagnostics and Model Selection

Aims

1. To apply conventional diagnostics to evaluate TSA model assumptions
2. To apply model selection criteria to alternative TSA methods

Methods

1. Use software of your choice to do this analysis
2. Or try using Excel to compute some of the diagnostics

Evaluations and Outputs

1. Evaluate one or two candidate time series and try a few alternative time series models,
   - e.g., linear, polynomial, ARIMA(1,1,0), ARIMA(2,1,0)

2. **Plot a histogram of residuals**
   - **are there outliers?**
   - **are the residuals symmetrically distributed?**
3. **Plot the time series of standardized residuals**
   - **if there are outliers, where are they located?**
4. **Evaluate whether there are high correlations among parameters of a given model**
5. **Evaluate whether autocorrelation (ACF and PACF) in residuals is significant**
   - **try starting with ARIMA(0,1, 0) with constant**
   - **try to use rules 6 and 7 to help in model formulation**

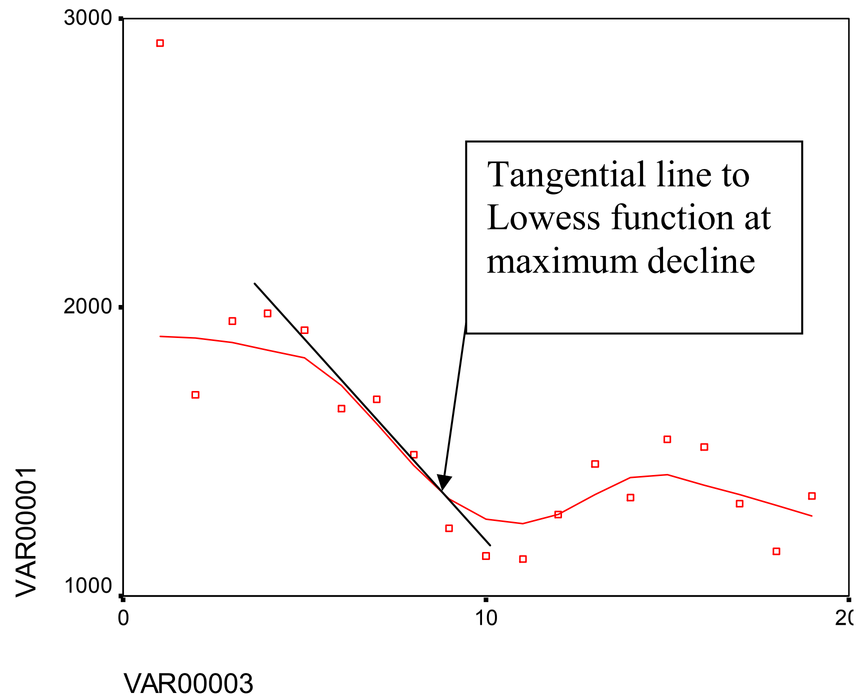
6. **Compare the AIC of alternative TSA models fitted to the same time series**
   - **Is the different larger than about 3?**
   - **If the models are nested models, try a likelihood ratio test to see whether the more complex model can be accepted.**
   - **Are the likelihood ratio test conclusions different from those based on the AIC results?**
7. **Try a likelihood ratio test on a linear model versus a 5th or 6th order polynomial model**

8. **Based on the above analyses, try to select a few plausible alternative TSA models for each of the alternative indicators**
   - **Characterize the temporal patterns in the candidate indicators based on the selected TSA models**
9. **Take about two hours to do these analyses**

Estimation of slope and rate of change in indicators

- Can use TSA model to provide estimate of slope over various historic periods or at certain points in time
- slope = dy / dx or $\Delta y / \Delta x$
- Thus the slope is the first derivative of the TSA function evaluated at a particular point in time
- It may be of interest to estimate maximum rate of change or rate of decrease over the time series within a given time frame
- Slope can be determined by

1. Taking the 1st derivative of the TSA model wrt to time and then evaluating the result at various points in time
2. Numerical approximation



Methods of estimating residual variance from univariate time series models

1. **Software estimates of residual standard error**
2. **Sums of squares of residuals divided by T less the number of estimated parameters**

$$\hat{\sigma}^2 = \frac{1}{(T-P)}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

where T is the number of years, P is the number of estimated parameters, and $\hat{y}_t$ is the TSA model estimate of y_t

Examples

- Linear model has two estimated parameters
- 6^{th} order polynomial has seven estimated parameters
- ARIMA(1,1,1) with constant has three estimated parameters
- Not possible to determine number of estimated parameters from a Lowess smooth

3. $\hat{\sigma}^2 = \frac{2}{3(T-2)} \sum_{t=3}^{T} \left(0.5 y_{t-2} - y_{t-1} + 0.5 y_t\right)^2$

(Nicholsen & Jennings 2004)

- **"asymptotically inefficient"**
- **simulation tested and found to provide "a good balance between bias and precision with small values of T"**

Exercise 1.4: Estimating slope and rate of change of indicators

Aims

1. For each candidate indicator, fit a TSA model to it
2. Summarize apparent temporal trends in candidate indicators
3. Obtain approximations of maximum slope over periods of about five years.

Methods

1. Utilize software TSA models of your own choice
2. Apply some diagnostic methods to justify your choice of particular TSA models
3. Apply either analytical or numerical methods to estimate the maximum slope for candidate indicators
4. Use software output or Excel to provide estimates of SE in the residual error for each candidate indicator

Outputs

1. For each candidate indicator provide summary plots of the fit of your choice of TSA models to the data
2. Provide approximations of maximum slope in each time series
3. Where possible provide estimates of SE in the residual error for each candidate indicator

Appendix 3 Workshop reports

A3.1 Power analysis of time series analysis in log abundance indices from the EVHOE survey

INDECO Workshop London January 2006

by Marie-Joëlle Rochet, IFREMER Nantes, France

Purpose:

Explore statistical properties of a population indicator, log abundance, analysed by TSA.
Data: EVHOE annual bottom trawl survey, 1987-2004 with gaps in 1991, 1993 and 1996.
Abundances estimated by swept area method (with depth stratification). 58 species.

Methods:

1. fit full series of ARIMA models with p,d,q=0 or 1 (larger lags were not tried due to short time series).
2. select best model based on AIC criterion
3. estimate size effect as maximum slope from series of four data points (*i.e.*, not necessarily four years due to missing data) for models based on raw data, and average difference for models based on differences.
4. calculate power for 1-2 tails tests with α=0.05-0.1, with effect size of interest = maximum short-term slope for the species in question
5. fit multivariate time-series models to groups of species with similar time patterns, models:
   glm(Nombre ~ Espece +time - 1, family = Gamma(link = log))
   glm(Nombre ~ Espece +time +Annee- 1, family = Gamma(link = log))
   glm(Nombre ~ Espece +time/Espece +Annee- 1, family = Gamma(link = log))

   (not done on differences because of missing values).
6. select best model and calculate power, with effect size of interest = common slope.

Results:

ARIMA models fitting
The models selected had generally 1 (15 species) or 2 components (42 species). A model with 3 components was selected for only one species (SPONCAN). 49 models were fitted on differenced observations; 11 had an autoregressive component; 42 had a moving average component. Selected models and aic values are reported in Table 1. Generally the autocorrelation in the data was weak (all data and autocorrelation plots shown in Appendix 1). Only 13 species had significant autocorrelation at lag 1; most of these had no further partial autocorrelation significant (except for large lags, which are probably not meaningful given the short data series, Appendix 2).

Table 1: ARIMA models results. p=lag in auto-regressive component, d=differencing (1) or not (0), q=lag in moving average component, aic =Akaike Information Criterion, y=year at the beginning of period with maximum slope (maxslope), sigma2=residual variance, meandiff=average difference between two successive years.

Species	p	d	q	aic	y	maxslope	sigma2	meandiff
ALLO	0	1	1	51.80	1988	0.240	1.564	0.352
AMMOTOB	0	1	1	63.71	1995	0.457	3.434	0.400
ARGESPH	0	1	1	46.29	1988	0.147	0.989	0.209
ARNOIMP	0	1	1	41.57	1992	0.374	0.788	0.122
ARNOLAT	0	1	0	44.08	1988	0.860	1.019	0.194
BOOPBOO	0	1	1	60.11	1988	-0.265	2.978	0.196
CALMMAC	0	1	0	37.91	1989	0.454	0.656	0.252
CANCPAG	1	0	0	31.74	1999	-0.144	0.315	0.042
CAPOAPE	0	1	0	52.48	1995	2.206	1.858	-0.369
CEPOMAC	1	1	0	39.30	1999	0.416	0.643	0.047
CHELCUC	0	0	1	35.37	2000	-0.335	0.286	0.072
CHELGUR	0	1	1	44.42	1998	0.346	0.971	0.099
CONGCON	1	1	0	33.53	1995	0.626	0.436	0.034
DICELAB	1	1	0	34.91	1990	0.356	0.473	0.185
DICOCUN	0	1	1	49.13	1999	0.418	1.357	0.038
ECITVIP	1	1	0	45.95	1987	0.458	1.028	0.238
ELEDCIR	0	1	1	45.53	1988	-0.296	1.033	-0.204
ENCHCIM	0	1	0	43.57	1997	0.717	0.983	0.211
ENGRENC	0	1	1	59.19	1988	1.125	2.789	0.029
GADIARG	0	1	1	45.90	2000	-1.817	0.614	-0.009
GALUMEL	1	1	0	33.93	1995	0.811	0.451	0.130
HELIDAC	0	1	1	43.11	1999	-0.152	0.789	0.074
ILLECOI	0	1	1	44.23	1988	0.236	0.913	0.226
LEPIBOS	0	1	1	38.54	1987	0.180	0.629	0.064
LEPIWHI	0	0	1	25.07	1988	0.098	0.204	0.081
LESUFRI	0	1	1	58.92	1995	0.367	2.640	0.314
LEUCNAE	0	1	1	36.30	1995	0.263	0.541	0.066
LIZARAM	0	0	1	47.48	1995	-0.714	0.640	-0.062
LOLIFOR	1	1	0	34.45	1988	0.424	0.436	0.318
LOLIVUL	1	1	0	35.79	1988	0.322	0.479	-0.074
LOPHPIS	0	1	0	40.57	1999	-0.835	0.794	0.089
MERLMER	0	0	1	32.84	1995	-0.166	0.347	0.099
MERNMER	0	1	1	48.99	1987	-0.373	1.309	0.017
MICMPOU	0	0	1	34.51	1988	-0.263	0.382	-0.158
MICUVAR	0	1	1	32.08	1995	0.531	0.399	0.057
MULLSUR	0	1	1	49.73	1987	0.214	1.265	-0.203
NEPHNOR	0	1	1	41.93	1988	0.138	0.724	0.093
PHYIBLE	0	0	1	40.47	1995	0.188	0.566	0.168
POMOMIN	0	1	1	70.53	1999	0.726	6.114	0.673
RAJACLA	0	1	1	49.48	1987	0.151	1.243	-0.034
SARDPIL	0	1	1	42.07	1995	0.755	0.821	0.111
SCOMJAP	0	1	1	57.95	1995	0.586	2.502	-0.390
SCOMSCO	1	1	0	48.07	1988	0.466	1.221	0.339
SCYOCAN	0	1	1	32.29	1988	0.413	0.409	0.232
SEPIELE	0	1	1	41.79	1999	0.320	0.801	0.010
SEPIOFF	0	1	1	50.52	1995	0.467	1.489	-0.083
SEPIORB	1	1	0	35.63	1992	0.424	0.497	0.035
SEPO	0	1	0	49.04	1999	-0.749	1.453	0.497
SOLESOL	0	1	1	49.74	1995	0.268	1.384	-0.040

SPONCAN	1	1	1	41.17	1997	0.319	0.632	0.180
SPRASPR	0	1	1	58.72	1987	-0.833	2.404	0.185
TODASAG	0	1	1	60.42	1988	-0.524	2.717	-0.499
TODIEBL	0	1	1	38.65	1998	0.365	0.642	0.221
TRAC	0	1	1	38.47	1995	0.445	0.631	0.141
TRAHDRA	0	1	1	49.17	1988	0.706	1.360	0.333
TRISLUS	0	0	1	40.90	1995	0.448	0.568	0.077
TRISMIN	0	0	1	27.56	1995	0.098	0.244	0.017
ZEUSFAB	0	1	1	45.01	1999	0.509	1.013	0.104

Power analysis of ARIMA models

Power of tests for time trends highly varied among species (table 2, Figure 1). This was related both to differences in expected effect size, of which magnitude varied owing to the method used to determine it: slope can vary significantly between 4-year periods among 15, and average difference was generally small. Generally difference model tests had a low power, because the effect size was small, variance was higher and degrees of freedom were lower than for other model tests. Increasing α and using one-tailed rather than two-tailed tests increased power with similar magnitudes, but these effects were low compared to the effect of expected effect size.

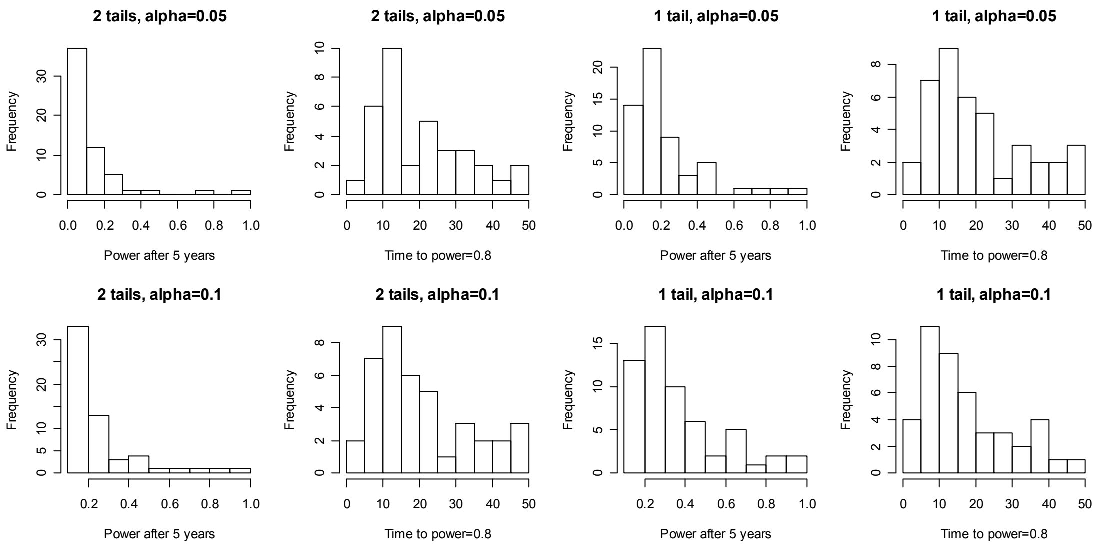


Figure 1: Results of power analysis of ARIMA models with various α risk and tails settings.

Table 2: Results of power analysis of ARIMA models for log abundance of Bay of Biscay species. p5years = power with 5 data points, timeto0.8 = number of years required for power to be at least 0.8.

	α=0.05, 2 tails		α=0.1, 2 tails		α=0.05, 1 tail		α=0.1, 1 tail	
Species	p5years	timeto0.8	p5years	timeto0.8	p5years	timeto0.8	p5years	timeto0.8
ALLO	0.056	>50	0.113	>50	0.082	>50	0.166	>50
AMMOTOB	0.060	>50	0.122	>50	0.095	>50	0.191	>50
ARGESPH	0.053	>50	0.108	>50	0.073	>50	0.148	>50
ARNOIMP	0.080	47	0.168	37	0.150	37	0.293	27
ARNOLAT	0.221	13	0.424	11	0.416	11	0.636	8
BOOPBOO	0.054	>50	0.108	>50	0.074	>50	0.150	>50
CALMMAC	0.109	28	0.228	22	0.214	22	0.396	16
CANCPAG	0.063	13	0.136	12	0.110	12	0.234	10
CAPOAPE	0.778	6	0.897	5	0.894	5	0.947	4
CEPOMAC	0.099	32	0.208	25	0.193	25	0.363	18

CHELCUC	0.165	8	0.379	7	0.368	7	0.623	6
CHELGUR	0.070	>50	0.146	>50	0.125	>50	0.248	38
CONGCON	0.278	11	0.502	9	0.496	9	0.706	7
DICELAB	0.099	32	0.207	25	0.192	25	0.362	19
DICOCUN	0.071	>50	0.148	50	0.127	50	0.253	37
ECITVIP	0.086	41	0.180	32	0.162	32	0.314	24
ELEDCIR	0.064	>50	0.131	>50	0.107	>50	0.214	>50
ENCHCIM	0.161	18	0.327	14	0.317	14	0.532	10
ENGRENC	0.143	20	0.294	16	0.283	16	0.490	12
GADIARG	0.964	4	0.982	3	0.981	3	0.989	3
GALUMEL	0.475	8	0.705	6	0.701	6	0.846	5
HELIDAC	0.055	>50	0.111	>50	0.078	>50	0.157	>50
ILLECOI	0.060	>50	0.122	>50	0.095	>50	0.191	>50
LEPIBOS	0.058	>50	0.119	>50	0.090	>50	0.181	>50
LEPIWHI	0.059	14	0.125	13	0.097	13	0.206	11
LESUFRI	0.058	>50	0.118	>50	0.090	>50	0.181	>50
LEUCNAE	0.071	>50	0.148	>50	0.127	>50	0.252	37
LIZARAM	0.375	6	0.671	6	0.664	6	0.839	5
LOLIFOR	0.133	21	0.275	17	0.263	17	0.464	13
LOLIVUL	0.088	39	0.185	31	0.168	31	0.324	22
LOPHPIS	0.271	11	0.493	9	0.486	9	0.698	7
MERLMER	0.066	12	0.144	11	0.120	11	0.255	10
MERNMER	0.067	>50	0.140	>50	0.117	>50	0.233	44
MICMPOU	0.092	9	0.211	9	0.194	9	0.395	8
MICUVAR	0.215	14	0.415	11	0.407	11	0.627	8
MULLSUR	0.056	>50	0.113	>50	0.081	>50	0.165	>50
NEPHNOR	0.054	>50	0.109	>50	0.076	>50	0.153	>50
PHYIBLE	0.063	13	0.134	12	0.108	12	0.230	11
POMOMIN	0.064	>50	0.132	>50	0.107	>50	0.215	>50
RAJACLA	0.053	>50	0.107	>50	0.071	>50	0.143	>50
SARDPIL	0.212	14	0.409	11	0.401	11	0.621	8
SCOMJAP	0.073	>50	0.152	47	0.131	47	0.260	34
SCOMSCO	0.081	47	0.169	37	0.150	37	0.293	27
SCYOCAN	0.134	21	0.277	17	0.265	17	0.467	12
SEPIELE	0.071	>50	0.148	50	0.127	50	0.252	37
SEPIOFF	0.075	>50	0.156	44	0.135	44	0.268	32
SEPIORB	0.120	24	0.251	19	0.238	19	0.430	14
SEPO	0.126	23	0.262	18	0.250	18	0.446	13
SOLESOL	0.058	>50	0.119	>50	0.090	>50	0.182	>50
SPONCAN	0.077	>50	0.161	41	0.142	41	0.279	30
SPRASPR	0.103	30	0.217	23	0.202	23	0.378	17
TODASAG	0.066	>50	0.138	>50	0.114	>50	0.229	46
TODIEBL	0.086	40	0.181	32	0.164	32	0.317	23
TRAC	0.109	28	0.228	22	0.214	22	0.396	16
TRAHDRA	0.121	24	0.253	19	0.240	19	0.433	14
TRISLUS	0.149	8	0.346	7	0.334	7	0.588	6
TRISMIN	0.058	15	0.121	13	0.092	13	0.194	12
ZEUSFAB	0.096	33	0.202	26	0.187	26	0.354	19

Multivariate time series power

A group of species with similar fluctuations between years was selected on purpose: SCYOCAN (small spotted dogfish), TRAC (horse mackerel), TRISMIN (poor cod), and ZEUSFAB (John Dorry).

The model with a common year effect and separate slopes for each species was selected by analysis of deviance (Table 3). Power to detect common slope was much higher than the power to detect maximum slope or average difference for separate species (Table 4).

Table 3: Results of analysis of deviance for multivariate time series analysis.

```
Analysis of Deviance Table
Model 1: Nombre ~ Espece + time - 1
Model 2: Nombre ~ Espece + time + Annee - 1
Model 3: Nombre ~ Espece + time/Espece + Annee - 1
  Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
1        55    25.5761
2        42    17.5783 13   7.9978 2.1286 0.0345357 *
3        39    11.7220  3   5.8563 6.7541 0.0008879 ***
---
Model 1: Nombre ~ Espece + time - 1
Model 2: Nombre ~ Espece + time/Espece - 1
Model 3: Nombre ~ Espece + time/Espece + Annee - 1
  Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
1        55    25.5761
2        52    19.5986  3   5.9775 6.8938 0.0007784 ***
3        39    11.7220 13   7.8766 2.0963 0.0374395 *
```

Table 4: Power analysis for multivariate time series analysis.

Species	p	d	q	aic	Effect size	sigma2	Power 5 year	Time to Power=0.8
SCYOCAN	0	1	1	32.29	0.232	0.409	0.134	21
TRAC	0	1	1	38.47	0.141	0.631	0.109	28
TRISMIN	0	0	1	27.56	0.098	0.244	0.058	15
ZEUSFAB	0	1	1	45.01	0.104	1.013	0.096	33
Multivariate				2256	0.12	0.188	0.171	8

A3.2 Time series analysis of trawl fleet data from the Aegean and Ionian Seas

by John Harabalous

HCMR, Greece

2.3 Average size (length and weight) in the community

The Mediterranean fisheries are highly diverse in terms of species and fishing gears used. Bottom trawling fisheries are essentially multi-species, and they are carried out in a wide range of depths and affect different bottoms and communities. Bottom trawl fleets predominate in many Mediterranean fisheries, being responsible for a high share of total catches and, in many cases, yielding the highest earnings among all the fishing sub-sectors. The high profitability of this fishing practice is largely due to its low selectivity with respect to sizes and species caught, and to the high harvests generated.

Trawlers have dramatic effects on the ecosystem including physical damage to the seabed and the degradation of associated communities, the over-fishing of demersal resources, and the changes in the structure and functioning of marine ecosystems derived from the depletion of populations and the huge amount of by-catches and associated discards. From a total of 300 species in the eastern Mediterranean about 60% are always discarded and mean discarded proportions reach 45% of the total catches (Machias *et al*., 2001). The latter underlines the necessity to gather relevant information and develop indicators contributing to track the impact of trawling on stocks, communities and finally the ecosystem. On the other hand, due to their multispecific nature and the large number of landing harbours involved, it has been traditionally difficult to gather long and reliable series of trawl fisheries data in Mediterranean countries.

For the Mediterranean two case studies are provided, one for the Aegean and the other for the Ionian Sea, on the basis of a monitoring program gathering data on a broad number of both target and non-target species on-board commercial trawlers during an eleven years period from 1995 till 2005. Both mean annual length and weight values were calculated to quantify size indicators of the Mediterranean demersal fish assemblages.

2.3.1 Material & methods

From 1995 to 2005, on a seasonal basis i.e. in October for autumn, in February for winter, and in May for spring, (summer trawling is prohibited in Greek waters) observers on board commercial trawlers followed fishing operations and recorded data from hauls stratified in three different depth strata in the central Aegean and in the Ionian Sea, considered to be among the most important fishing grounds for trawl fisheries in Greece.

In a representative sample from each haul, the various species were sorted out; the number of individuals per species and their total weight were taken down, while total length of each individual was also recorded. From this, mean length and mean weight of the Aegean and Ionian demersal assemblages were calculated. Mean length was calculated as

$$\overline{L} = \frac{\sum_{i=1}^{N} L_i}{N}$$

where L is the length of an individual and N is the total number of individuals.

Mean individual weight in the catch was calculated as the sum of the catch weights divided by the total number of fish caught.

Due to unbalances in sampling stations among seasons/depth zones in the various years of the surveys, the effects created by the aforementioned factors were tested using General Linear Model Analysis of Variance (GLM ANOVA) using SPSS v.11 for Windows. Since those effects were found to be significant ($p<0.05$) it was decided to proceed in adjustments by producing marginal means of the indicators.

Possible trends in the indicators' time series were extracted through GLM ANOVA, using the time t (year) as the covariate and the slope of the linear model was used as the descriptor of the series' trend. The *partial eta squared*, the *noncentrality parameter* and the *observed power* were also estimated. Partial eta squared is the ratio of the variation accounted for by an individual independent variable to the sum of the variation accounted for by the independent variable and the variation unaccounted for by the model as a whole. The estimated non-centrality parameter is used in determining the observed power under the alternative hypothesis for the two tailed test F test. The observed power gives the probability that the F test will detect the differences and it was calculated at a=0.05 significance level.

For comparison of the performance of this indicator with that of the other three (i.e. mean max length, mean trophic level and Shannon-Wiener diversity index) that were produced for Mediterranean waters, the following analyses have been conducted:

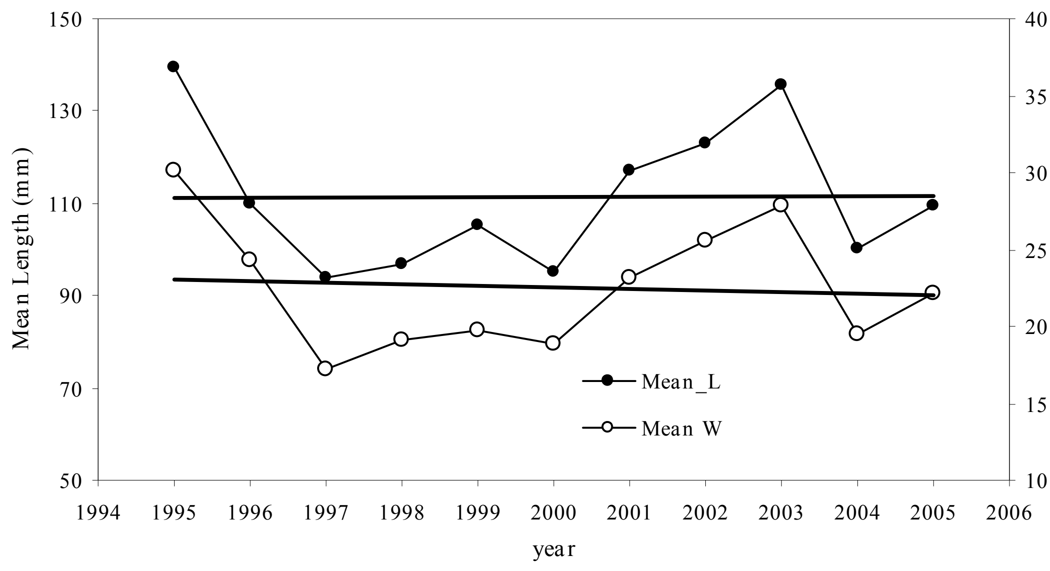
1. Zero lag correlation among time series using Pearson r correlation coefficient.
2. After standardisation of each series (by subtraction of mean and division by standard deviation) a GLM ANOVA was applied using time (year) as covariate and the various indicator time series as a random factor ('metric') to evaluate the metric's effect.

Results for these analyses are given within the section 2.6 (i.e. in the biodiversity indicator section).

2.3.2. Results

Aegean Sea

Values of mean length and weight per haul were calculated and time series graphs of the two indicators are shown in Figure 2.3.2. Although both series appeared to have a negative trend, this trend was not significant (Table 2.3.2) and its power was found to be very low (about 0.055).



Parameter Estimates

Dependent Variable: Mean Length

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared	Noncent. Parameter	Observed Power[a]
					Lower Bound	Upper Bound			
Intercept	113.491	10.425	10.887	.000	89.908	137.074	.929	10.887	1.000
T	-.315	1.537	-.205	.842	-3.792	3.162	.005	.205	.054

a. Computed using alpha = .05

Parameter Estimates

Dependent Variable: Mean Weight

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared	Noncent. Parameter	Observed Power[a]
					Lower Bound	Upper Bound			
Intercept	23.110	2.779	8.315	.000	16.822	29.397	.885	8.315	1.000
T	-.099975	.410	-.244	.813	-1.027	.827	.007	.244	.056

a. Computed using alpha = .05

2.3.3 Evaluation & Interpretation

The fact that our results show a non significant linear trend with a very low power suggests that the effects of trawl fishing on the demersal assemblages of the Mediterranean are not detectable over the 11-year study period. The latter does not imply that fishing has no impact on resources, but could be linked with the small time period of our observation and /or to errors of our monitoring program taking into account that our data were derived from commercial fishing observations, with no spatial repetition. Although such effects were tackled through GLM techniques they could still cause the observed great interannual variations in the indicator values.

2.3.4 Recommendations

The fact that mean length and weight are considered to be influenced by the status of the assemblage as determined by the monitoring programme, it is suggested that data derived through experimental surveys with spatio-temporal repetition should be also used. Such data for the Mediterranean area are those from the MEDITS surveys conducted since 1994, and would be quite interesting to quantify indicator values from those, which at the moment are restricted to particular Institutes of the various Mediterranean EU countries. It should be pointed out, however, that at present there’s an effort to produce indicator values based on MEDITS at a pan-Mediterranean level from the working groups involved in the project, and results will be published in a report that will come out in the next few months. Thus, since results are not finalised yet, they could not be used in the present study.

Furthermore, the fact that the Mediterranean monitoring programs run for a rather small time period (about 10 years) it is suggested that gathering a greater time series would possibly clarify patterns and provide more clear trends.

2.4 Mean trophic level

2.4.1 Materials and methods

Time-series of the "Trophic level" indicator for the Mediterranean is based on the same monitoring program running in the Aegean and the Ionian and following the operation of commercial trawlers. These surveys are described in section 2.2.

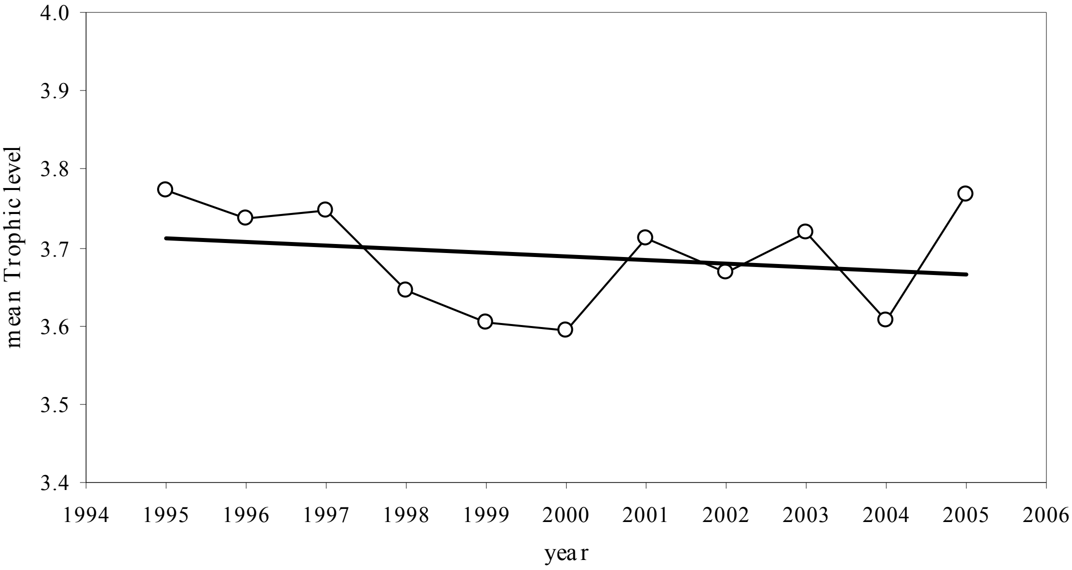
The trophic levels of individuals were estimated from their length, using relationships between length and trophic level as determined by Stergiou and Karpouzi (2002) using TrophLab (Pauly et al., 2000d) applied to data from Mediterranean fish species. Trophic level versus length relationships were available for 110 out of the 166 species appearing in our original data set. The mean trophic level (*TL*) was calculated per haul according to the formula used for the North Sea

Possible trend of the time series as well as its power were estimated following the procedure presented in 2.3.1. For comparison of the performance of this indicator with the other, the same analyses appearing in 2.3.1. were applied and results are provided within section 2.6.

2.4.2 Results

Aegean Sea

Values of mean trophic levels per haul for the Mediterranean fish assemblages were calculated and the respective time series is given in Figure 2.4.2. Mean trophic level indicator values ranged between 3.59 and 3.77, appearing to follow more or less the same pattern as the mean size indicators. Again a negative trend appears, and this trend is also non significant (Table 2.4.2) with a very low power (0.05).



Parameter Estimates

Dependent Variable: Mean Trophic level

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	3.690	.044	83.505	.000	3.590	3.790	.999	83.505	1.000
T	-.00154	.007	-.236	.819	-1.628E-02	1.320E-02	.006	.236	.055

a. Computed using alpha = .05

2.4.3 Evaluation and interpretation

Trophic level values calculated for the Mediterranean demersal fish assemblages were generally lower than in other areas studied in the framework of this project. The same observation as that made for the mean size indicators could be also made herein regarding the small power to detect effects of trawl fishing based on our time series. As before time constraints as well as sampling errors could impact our data set and hence estimation of trends.

2.3.4 Recommendations

For the Mediterranean it is recommended to compare the present findings with those that will be derived in the near future from the MEDITS project, which could clarify if the lack of spatial repetition of the samplings in the present study influenced the lack of significance in the observed trends for the indicator's time series, as well as their power.

2.5 Mean maximum length

2.5.1 Material & methods

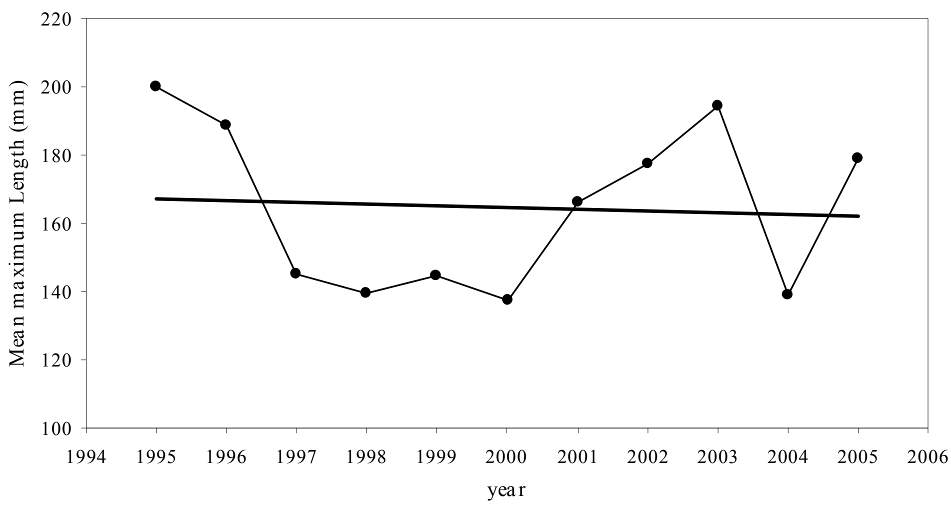
Mean maximum length was calculated per haul according to 2.5.1 Time-series for this indicator were based at the same surveys as the indicator "average size in the community" (section 2.3). Performance was also assessed in the same manner as that indicator. Possible trend of the time series as well as its power were estimated following the procedure presented in 2.3.1.

For comparison of the performance of this indicator with the other, the same analyses appearing in 2.3.1. were applied and results are provided within section 2.6.

2.5.2 Results

Aegean Sea

Values of mean maximum length per haul for the Mediterranean fish assemblages were calculated and the respective time series is given in Figure 2.5.2. This indicator's time series appeared to follow the exact same pattern as the mean length indicators. The negative trend that also appears, is again non significant (Table 2.5.2) and has a very low power (0.05).\



Parameter Estimates

Dependent Variable: Mean maximum Length

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared	Noncent. Parameter	Observed Power[a]
					Lower Bound	Upper Bound			
Intercept	167.895	16.489	10.182	.000	130.594	205.196	.920	10.182	1.000
T	-.522	2.431	-.215	.835	-6.021	4.978	.005	.215	.054

a. Computed using alpha = .05

2.5.3 Evaluation and interpretation

Values of the mean max indicator exhibited a non significant linear trend with a very low power, observations made also for the mean size ant mean trophic level indicators. Once again the rather small time series of data as well as the fact that they come from a monitoring program, which has no spatial repetition could impact our results.

2.5.4 Recommendations

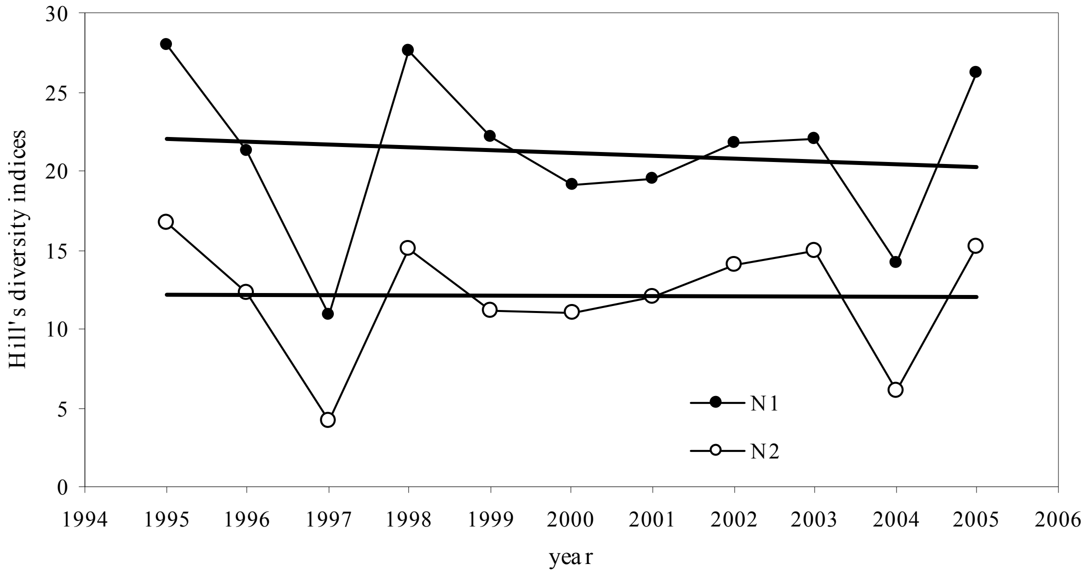
Our recommendation would be once again to compare values from the present study with the respective ones from the MEDITS program, when they will be available.

2.6 Biodiversity

For the Mediterranean data from the same surveys as described in section 2.3 were used to calculate values of Hill's N_1 and N_2. Performance was also assessed in the same manner as that indicator. Possible trend of the time series as well as its power were estimated following the procedure presented in 2.3.1.

For comparison of the performance of this indicator with the other, the same analyses appearing in 2.3.1. were applied and results are provided within section 2.6.

2.6.2 Results
Values of Hill's N_1 and N_2 per haul for the Mediterranean fish assemblages were calculated and the respective time series are given in Figure 2.6.2. Although these two indicators appear to have a similar pattern, that pattern deviates from that of the other indicators. A negative trend appears in both indicator time series (slope for N_1: –0.165 per year and for N_2: –0.008), which in both cases is insignificant (Table 2.6.1: p>0.75) and has a very low power (0.059 and 0.050 respectively).



Parameter Estimates

Dependent Variable: Hill' s N1

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	22.134	3.559	6.219	.000	14.082	30.185	.811	6.219	1.000
T	-.165	.525	-.315	.760	-1.352	1.022	.011	.315	.059

a. Computed using alpha = .05

Dependent Variable: Hill' s N2

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	12.102	2.669	4.535	.001	6.066	18.139	.696	4.535	.980
T	-.00816	.393	-.021	.984	-.898	.882	.000	.021	.050

a. Computed using alpha = .05

2.6.3 Evaluation and interpretation

Both diversity indices show the same pattern.

2.6.6 Relationships among indicators

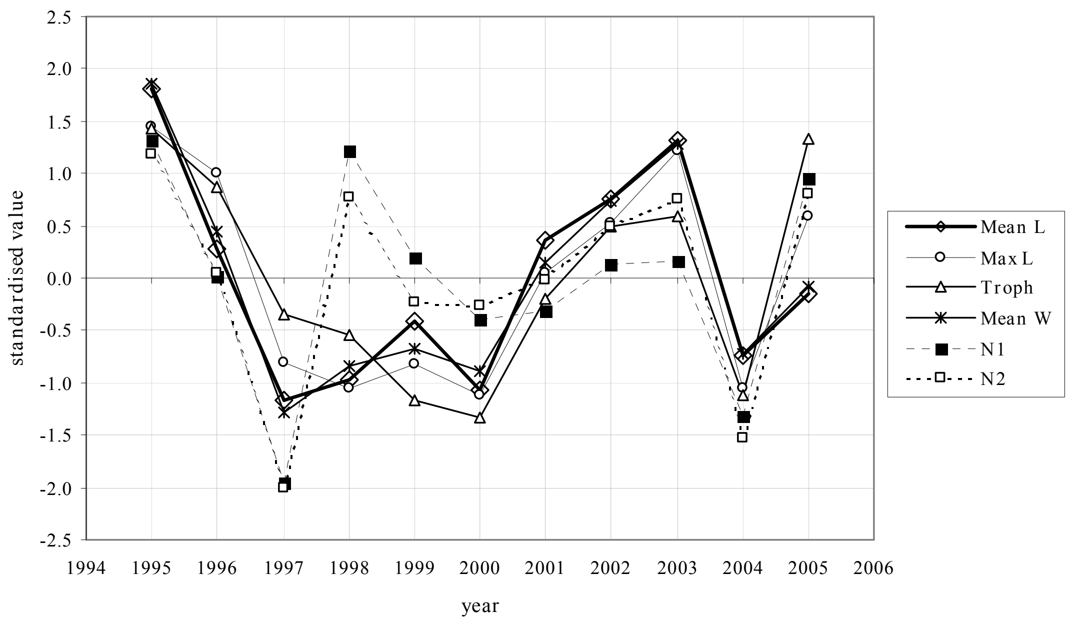
From Table 2.6.2 it is obvious that all indicators are positively correlated to each other. As expected, the highest correlation was found to be between mean length and mean weight, since they provide the same type of information. Then this also holds for the relationship between N_1 and N_2, while N_1 appears to have a non significant correlation with all the rest. Moreover, a high correlation existed between mean trophic level and all the rest indicators except for N_1. A more detailed representation of these correlations appears in Figure 2.6.2

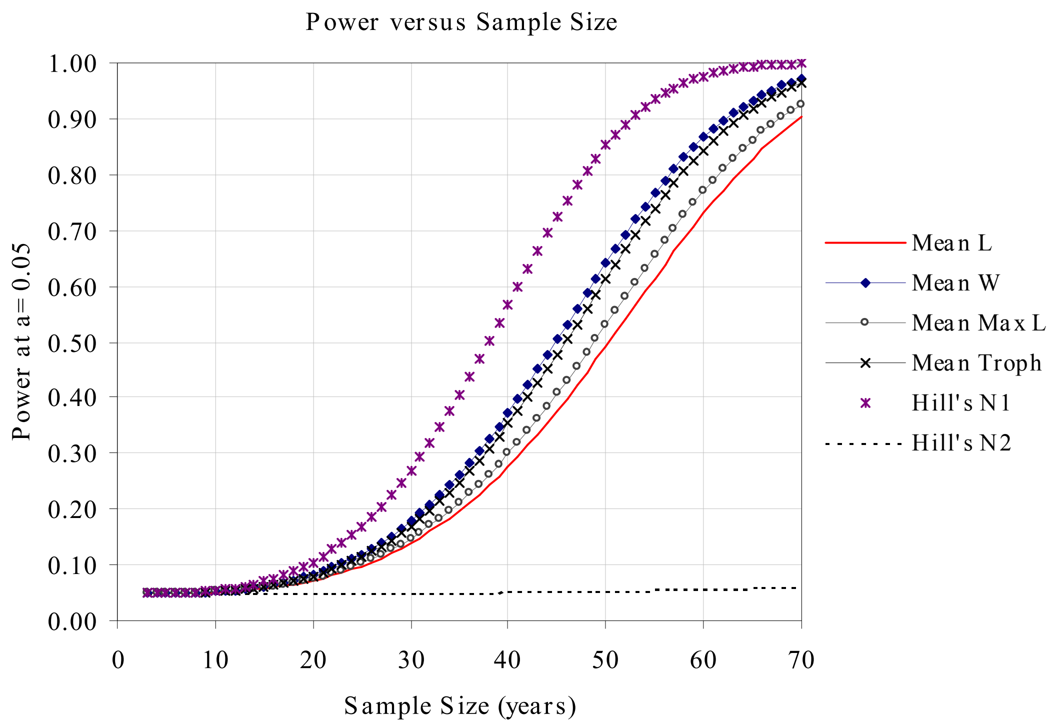
Correlations

		Mean L	Mean W	Mean Max L	Mean Troph	N1	N2
Pearson Correlation	Mean L						
	Mean W	.989**					
	Mean Max L	.912**	.928**				
	Mean Troph	.742**	.773**	.918**			
	N1	.486	.528	.469	.518		
	N2	.652*	.693*	.633*	.613*	.947**	
Sig. (2-tailed)	Mean L						
	Mean W	.000					
	Mean Max L	.000	.000				
	Mean Troph	.009	.005	.000			
	N1	.130	.095	.145	.103		
	N2	.030	.018	.036	.045	.000	

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).





Power versus Sample Size
Power at a= 0.05
Sample Size (years)
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
10
20
30
40
50
60
70
Mean L
Mean W
Mean Max L
Mean Troph
Hill's N1
Hill's N2

A3.3 Basque Country Time Series Analysis

By Iñigo Muxika and Angel Borja (AZTI-Tecnalia)

Two locations from the Basque Country (SE Bay of Biscay) were selected. One of them (named Station E) is located at about 7 m water depth, in the inner part of the Nervión Estuary (Bilbao), which has been monitored since 1989. The second sampling station (named Station M) is located at about 35 m water depth, in the coastal area near Deba, and has been monitored since 1995. A study on the pressures over the entire Basque Country is available in Borja *et al.* (2006a).

None of the two sampling stations is affected by fishery pressures. However, Station E was considered azoic until 1990 due to high levels of pollutants (heavy metals and organic matter) in sediments and low (even anoxic) dissolved oxygen concentration in bottom water layers (Borja *et al.,* 2006b). The estuary has been cleaned up in the past two decades, and an important recovery of the quality has been detected, especially in the inner part of the system (Borja *et al.*, 2006b).

Conversely, no important pressure is known in the surroundings of Station M and trends are not expected for the benthic communities.

For this exercise AMBI was selected as benthic community 'health' indicator (Borja *et al.*, 2000, 2003; Borja and Muxika, 2005; Muxika *et al.*, 2005). High values of AMBI (close to 7) indicate highly disturbed sediments; low values of AMBI (close to 0) correspond to undisturbed sediments.

For power analysis, changes of 0.5 units·y^{-1} in AMBI were arbitrarily considered as ecologically meaningful.

Station E:
Selection of model:

Both the linear regression model (Figure 1) and the 2nd order polynomial regression model (Figure 2) between AMBI and the sampling year fitted quite good to the data ($p<0.000$ for both of them, and $r^2= 0.768$ for the linear regression and $r^2= 0.80$ for the polynomial regression).

However, a comparison of the likelihood between both models (χ^2) shows that the addition of one more parameter to the model does not improve it enough. So the simplest one (linear regression) was chosen for this exercise.

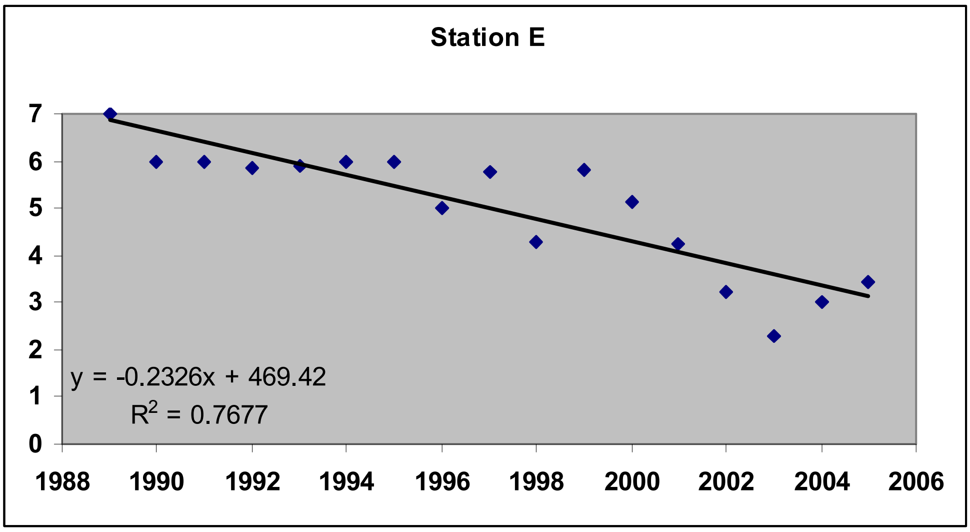


Figure 1: Linear regression between AMBI and sampling years for Station E.

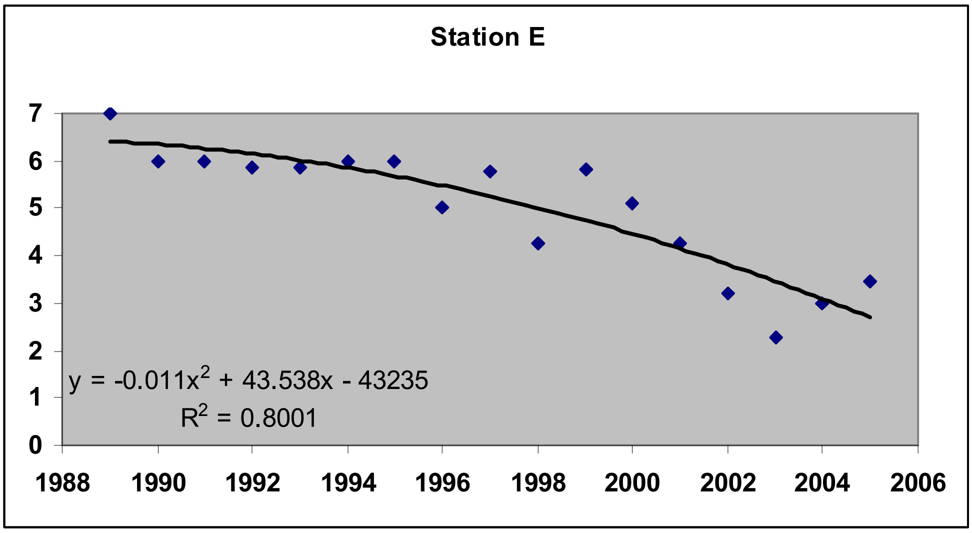


Figure 2: 2nd order polynomial regression between AMBI and sampling years for Station E.

At Station E, a recovery of the benthic community (negative trend of AMBI indicator) was detected due to the sewerage plan, so a one tailed power analysis was carried out for the linear regression model abovementioned.

Power Analysis:

A power analysis was carried out for four different significance levels (α= 0.100, α= 0.050, α= 0.010 and α= 0.001) to study the effect of each significance level on power (Figure 3). In this way, it can be seen that in almost five years there is a power of 0.80 to detect a significant trend of 0.5 AMBI units for α= 0.100; almost six years are needed to detect the same trend for a significance level of α= 0.050; for α= 0.010, about eight years are necessary; and for α= 0.001, almost nine years.

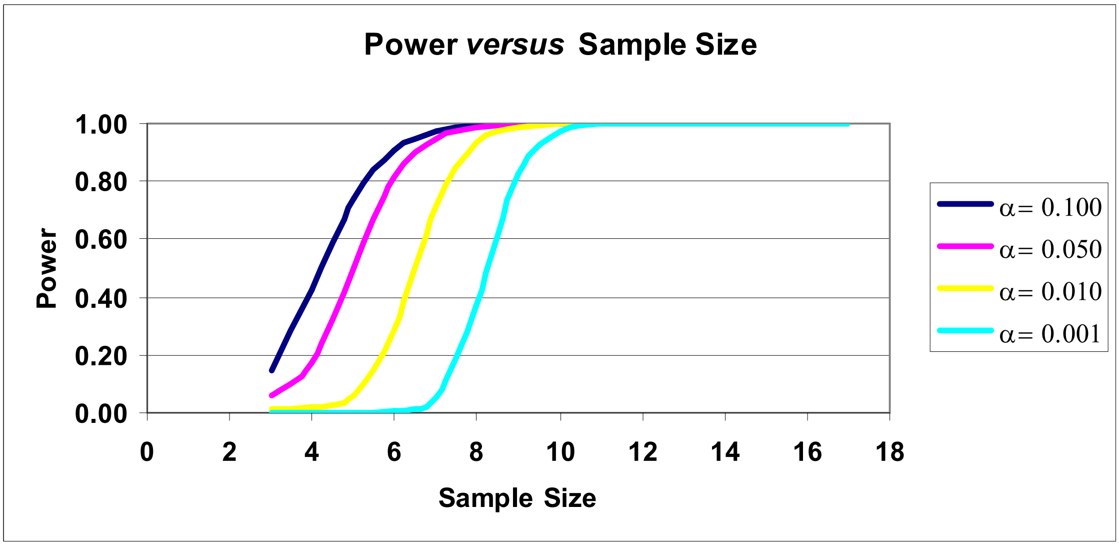


Figure 3: Power of the linear regression model for Station E *versus* sample size for different significance levels.

Station M:
Selection of model:

As it was expected (for the absence of significant pressures in the area (Borja *et al.* (2006a)), the regression between AMBI and sampling years was not significant for any of the regression models (p> 0.1). The linear regression is presented in Figure 4 as an example.

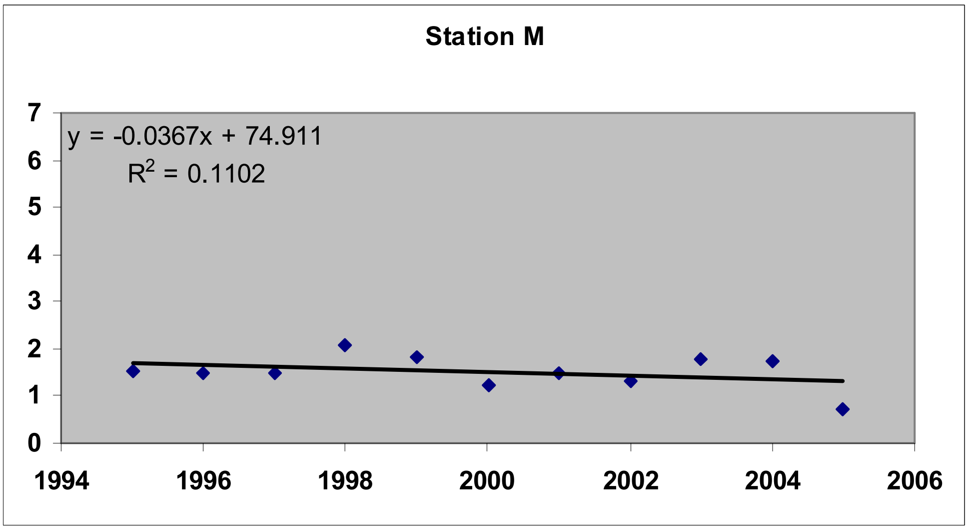


Figure 4: Linear regression between AMBI and sampling years for Station M.

Power Analysis:

At Station M, any trend in AMBI would be indicative of something: a positive trend would indicate an unknown pressure which has disappeared; conversely, a negative trend would indicate a new pressure on benthos.

So, it is interesting to analyze the power of the linear regression model with two tails. In Figure 5 the results for different significant levels are shown. For α= 0.100; a little more than four years would be enough to obtain a 0.80 power of detecting a trend of 0.5 AMBI units; almost five years are needed to detect the same trend for a significance level of α= 0.050; for α= 0.010, almost six years are needed; and for α= 0.001, about seven years.

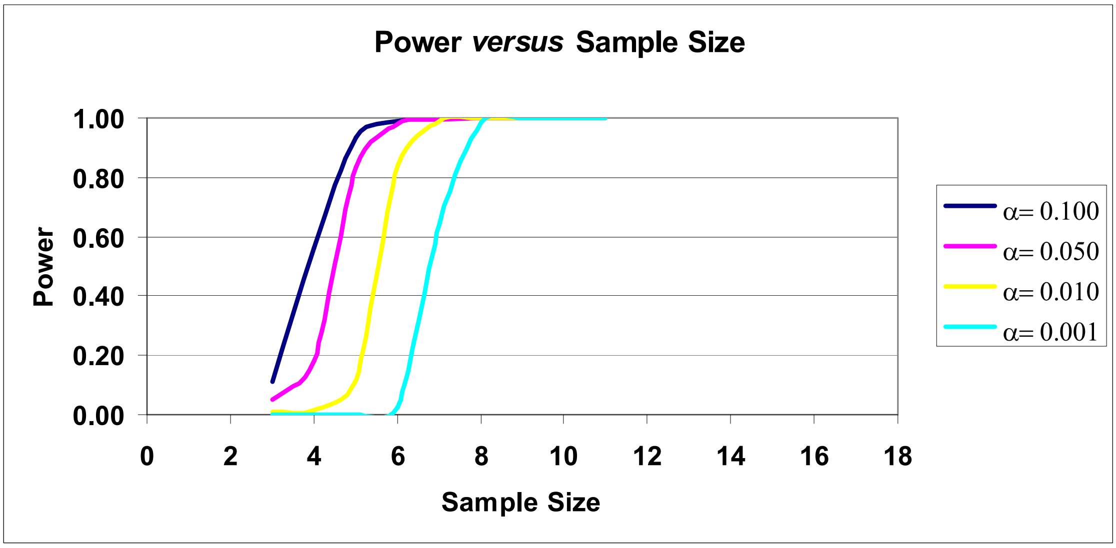


Figure 5: Power of the linear regression model for Station M *versus* sample size for different significance levels.

REFERENCES

BORJA, Á.; MUXIKA, I. 2005. Guidelines for the use of AMBI (AZTI's marine biotic index) in the assessment of the benthic ecological quality. *Marine Pollution Bulletin*, 50: 787-789.

BORJA, Á.; FRANCO, J; PÉREZ, V. 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Marine Pollution Bulletin*, 40 (12): 1100-1114.

BORJA, Á.; MUXIKA, I.; FRANCO, J. 2003. The application of a Marine Biotic Index to different impact sources affecting soft-bottom benthic communities along European coasts. *Marine Pollution Bulletin*, 46: 835-845.

BORJA, A., I. GALPARSORO, O. SOLAUN, I. MUXIKA, E.M. TELLO, A. URIARTE AND V. VALENCIA, 2006a. The European Water Framework Directive and the DPSIR, a methodological approach on assessing the risk of failing to achieve the good ecological status. *Estuarine, Coastal and Shelf Science*, 66: 84-96.

BORJA, A., I. MUXIKA, J. FRANCO, 2006b. Long-term soft-bottom benthos recovery, following urban and industrial sewage treatment in the Nervión estuary (southern Bay of Biscay). *Marine Ecology Progress Series*, 313: 43-55.

MUXIKA, I.; BORJA, Á.; BONNE. W. 2005. The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. *Ecological Indicators*, 5: 19-31.