

8. STOCHASTIC MODELS FOR PREDICTION OF THE CATCH DYNAMICS

The modelled time series are represented by catch variations of the 12 main commercial species. The sampling interval of the time series is 1 year, and all series finish in 1996 and cover the following periods:

- 1900 (1 series - Pacific herring, 97 observations),
- 1920 (3 series - Californian and Japanese sardine, and salmon, 77 observations),
- 1930 (2 series - Atlantic cod and herring, 67 samples),
- 1950 (4 series - Peruvian anchovy and sardine, Alaska pollock, South African sardine, 47 observations),
- 1957 (1 series - European sardine, 40 observations),
- 1969 (1 series - Chilean jack mackerel, 29 observations)

All considered time series undergo well-manifested cyclic variation. The time series can be represented as the following sum:

$$\mathbf{x}(t) = \xi(t) + \mathbf{F}(t), \quad \mathbf{x}(t), t=t_0, \dots, t_0+N-1, \quad (1)$$

where $\mathbf{x}(t)$ is the catch volume in the year t , t_0 is the starting year of the series, $\mathbf{F}(t)$ is deterministic function (trend), and $\xi(t)$ is a stationary component (including random noise) of the time series. The stationary component has no trends in mean or variation. The deterministic part of the model ($\mathbf{F}(t)$) is taken as a multiple cyclic trend with m periods T_i , $i=1, \dots, m$ (Anderson 1971; Kashyap and Rao 1976):

$$\mathbf{F}(t) = \sum_{i=1}^m (\mathbf{B}_i \sin(\theta_i(t)) + \mathbf{D}_i \cos(\theta_i(t))) + \mathbf{G}, \quad \text{where } \theta_i(t) = 2\pi(t-t_0)/T_i, \quad (2)$$

where \mathbf{B}_i , \mathbf{D}_i are unknown amplitudes, and \mathbf{G} is an unknown static shift constant. Equation (2) can also be expressed as:

$$\mathbf{F}(t) = \sum_{i=1}^m \mathbf{A}_i \cos(\theta_i(t) - \varphi_i) + \mathbf{G}, \quad \text{where } \mathbf{A}_i^2 = \mathbf{B}_i^2 + \mathbf{D}_i^2, \quad \tan(\varphi_i) = \mathbf{B}_i/\mathbf{D}_i \quad (3)$$

The stochastic part, $\xi(t)$, may be represented as an autoregressive Box-Jenkins model (Box and Jenkins 1970; Anderson 1971; Kashyap and Rao 1976) of the given order p (or AR(p)-model):

$$\xi(t) = -\sum_{k=1}^p \mathbf{a}_k \mathbf{x}(t-k) + \varepsilon(t) \quad (4)$$

where p is the number of autoregressive parameters, \mathbf{a}_k , and $\varepsilon(t)$ is a residual signal, which is considered to be Gaussian³ white noise with zero mean and unknown variance s^2 . Combining formulae (1) and (4), the general model is presented as follows:

³ The Gaussian probability distribution, also known as the Normal distribution, is often appropriate when errors show constant variation.

$$\mathbf{x}(t) = \sum_{k=1}^p \mathbf{a}_k \mathbf{x}(t-k) + \sum_{i=1}^m (\mathbf{B}_i \sin(\theta_i(t)) + \mathbf{D}_i \cos(\theta_i(t))) + \mathbf{G} + \boldsymbol{\varepsilon}(t) \quad (5)$$

Thus, unknown parameters of model (5) to be estimated from the data, are: \mathbf{p} autoregressive parameters \mathbf{a}_k ; $2\mathbf{m}$ harmonic amplitudes \mathbf{B}_i and \mathbf{D}_i ; a static shift constant \mathbf{G} ; and variance \mathbf{s}^2 of the residual signal $\boldsymbol{\varepsilon}(t)$. These parameters are estimated using the maximum likelihood method (Kashyap and Rao 1976), which in case of Gaussian noise $\boldsymbol{\varepsilon}(t)$ coincides with the least squares method. Let us define $\mathbf{Y}(t)$ as the following $(\mathbf{p}+2\mathbf{m}+1)$ -dimensional column vector:

$$\mathbf{Y}(t) = (-\mathbf{x}(t-1), \dots, -\mathbf{x}(t-p), \sin(\theta_1(t)), \cos(\theta_1(t)), \dots, \sin(\theta_m(t)), \cos(\theta_m(t)), 1)^T \quad (6)$$

where the index “ T ” indicates a transposed (i.e. column) vector, and the parameters’ column vector \mathbf{c} of the same dimensionality is:

$$\mathbf{c} = (\mathbf{a}_1, \dots, \mathbf{a}_p, \mathbf{B}_1, \mathbf{D}_1, \dots, \mathbf{B}_m, \mathbf{D}_m, \mathbf{G})^T \quad (7)$$

Then the model (5) in matrix form could be described as follows:

$$\mathbf{x}(t) = \mathbf{c}^T \cdot \mathbf{Y}(t) + \boldsymbol{\varepsilon}(t) \quad (8)$$

where \mathbf{c}^T = row vector of parameters.

The following gives a detailed account of the method used to fit the parameters. It is not necessary to understand the fitting method to understand the results of the model. But it is useful to follow the principles used here for robust regression to understand how the particular results are derived. The parameters, \mathbf{c} , can be estimated by minimising the sum:

$$\sum_{t=t_0+p}^{t_0+N-1} \boldsymbol{\varepsilon}^2(t) = \sum_{t=t_0+p}^{t_0+N-1} (\mathbf{x}(t) - \mathbf{c}^T \cdot \mathbf{Y}(t))^2 \rightarrow \min_{\mathbf{c}} \quad (9)$$

A solution of (9) can be obtained as the solution to the Normal equations (i.e. standard least-squares regression):

$$\mathbf{c} = \mathbf{A}_0^{-1} \cdot \mathbf{R}_0, \text{ where matrix } \mathbf{A}_0 = \sum_{t=t_0+p}^{t_0+N-1} \mathbf{Y}(t) \cdot \mathbf{Y}^T(t), \text{ vector } \mathbf{R}_0 = \sum_{t=t_0+p}^{t_0+N-1} \mathbf{x}(t) \mathbf{Y}(t) \quad (10)$$

and the unbiased maximum likelihood estimate of the variance of $\boldsymbol{\varepsilon}(t)$ is:

$$\mathbf{s}^2 = \frac{\sum_{t=t_0+p}^{t_0+N-1} \boldsymbol{\varepsilon}^2(t)}{(N-p)} \quad (11)$$

Estimates defined by equations (10) and (11) are the least squares estimates and maximum likelihood where the errors are genuinely normally distributed. The least squares estimates can be very sensitive to outliers, so that only a few outliers can lead to parameter estimates, which are considerably different from the “true” values (Huber 1981). The outliers may occur in fishery data for various reasons, which cause a temporary violation of the assumptions of the model: increases in illegal activities or the influences of rare but powerful events (war, economic crisis, etc.). One way to make estimates more resistant to outliers is to apply maximum likelihood estimates, but with the assumption that the distribution of the residual signal $\boldsymbol{\varepsilon}(t)$ is Gaussian for small values of $\boldsymbol{\varepsilon}$ and Laplacian for large values of $\boldsymbol{\varepsilon}$. This kind of distribution has the following density (Huber 1981):

$$\mathbf{p}(\boldsymbol{\varepsilon}) = \frac{\boldsymbol{\beta}}{\mathbf{s}} \exp\left(-\frac{\boldsymbol{\varepsilon}^2}{2\mathbf{s}^2}\right) \text{ for } |\boldsymbol{\varepsilon}| \leq \mathbf{s} \cdot \mathbf{a} \quad (12a)$$

$$\mathbf{p}(\boldsymbol{\varepsilon}) = \frac{\boldsymbol{\beta}}{\mathbf{s}} \exp\left(-\frac{\mathbf{a}}{\mathbf{s}} \left(|\boldsymbol{\varepsilon}| - \frac{\mathbf{a}\mathbf{s}}{2}\right)\right) \text{ for } |\boldsymbol{\varepsilon}| > \mathbf{s} \cdot \mathbf{a} \quad (12b)$$

where \mathbf{a} is a robustness parameter, which usually takes values 1-3 (we use $\mathbf{a}=2$), \mathbf{s} is a scale parameter (analogous to standard deviation in Gaussian law), and $\boldsymbol{\beta}$ is normalising constant:

$$\beta = \beta(\mathbf{a}) = 1 / (2(\exp(-\mathbf{a}^2 / 2) / \mathbf{a} + \int_0^{\mathbf{a}} \exp(-z^2 / 2) dz)) \quad (12c)$$

Maximum log-likelihood method for the model (12) leads to the following problem (with accuracy to an additive constant, which depends on \mathbf{a} only):

$$\mathbf{J}(\mathbf{c}, \mathbf{s}) = \sum_{t=t_0+p}^{t_0+N-1} \log(p(\boldsymbol{\varepsilon}(t))) = -(N-p) \cdot \log(\mathbf{s}) - \frac{1}{2\mathbf{s}^2} \sum_{|\boldsymbol{\varepsilon}(t)| \leq \mathbf{s}\mathbf{a}} \boldsymbol{\varepsilon}^2(t) - \frac{\mathbf{a}}{\mathbf{s}} \sum_{|\boldsymbol{\varepsilon}(t)| > \mathbf{s}\mathbf{a}} |\boldsymbol{\varepsilon}(t)| \rightarrow \max_{\mathbf{c}, \mathbf{s}} \quad (13)$$

Solution of equation (13) can be obtained using an iterative method, which is a combination of a generalised Newton-Raphson method for vector \mathbf{c} and a simple iteration method for \mathbf{s} :

$$\mathbf{c}^{(j+1)} = \mathbf{c}^{(j)} + \mathbf{A}^{-1}(\mathbf{c}^{(j)}, \mathbf{s}^{(j)}) \cdot \mathbf{R}(\mathbf{c}^{(j)}, \mathbf{s}^{(j)}) \quad , \quad \mathbf{s}^{(j+1)} = 1 / \chi(\mathbf{c}^{(j)}, \mathbf{s}^{(j)}) \quad (14)$$

where $\mathbf{j}=0, 1, \dots$ is the iteration index. Matrix $\mathbf{A}(\mathbf{c}, \mathbf{s})$ and vector $\mathbf{R}(\mathbf{c}, \mathbf{s})$ are computed as follows:

$$\mathbf{A}(\mathbf{c}, \mathbf{s}) = \sum_{t=t_0+p}^{t_0+N-1} \mathbf{F}''(\boldsymbol{\varepsilon}(t), \mathbf{s}) \cdot \mathbf{Y}(t) \cdot \mathbf{Y}^T(t) \quad , \quad \mathbf{R}(\mathbf{c}, \mathbf{s}) = \sum_{t=t_0+p}^{t_0+N-1} \mathbf{F}'(\boldsymbol{\varepsilon}(t), \mathbf{s}) \cdot \mathbf{Y}(t) \quad (15)$$

where

$$\mathbf{F}'(\boldsymbol{\varepsilon}, \mathbf{s}) = \begin{cases} \boldsymbol{\varepsilon}, & \text{if } |\boldsymbol{\varepsilon}| \leq \mathbf{a}\mathbf{s} \\ \mathbf{s}\mathbf{a} \cdot \text{sign}(\boldsymbol{\varepsilon}), & \text{if } |\boldsymbol{\varepsilon}| > \mathbf{a}\mathbf{s} \end{cases} \quad , \quad \mathbf{F}''(\boldsymbol{\varepsilon}, \mathbf{s}) = \begin{cases} 1, & \text{if } |\boldsymbol{\varepsilon}| \leq \mathbf{a}\mathbf{s} \\ 0, & \text{if } |\boldsymbol{\varepsilon}| > \mathbf{a}\mathbf{s} \end{cases} \quad (16)$$

Function $\chi(\mathbf{c}, \mathbf{s})$ is defined as:

$$\chi(\mathbf{c}, \mathbf{s}) = (\sqrt{\gamma^2 + 4\alpha} - \gamma) / (2\alpha) \quad (17)$$

where

$$\alpha = \sum_{|\boldsymbol{\varepsilon}| \leq \mathbf{a}\mathbf{s}} \boldsymbol{\varepsilon}^2(t) / (N-p) \quad , \quad \gamma = \mathbf{a} \cdot \sum_{|\boldsymbol{\varepsilon}| > \mathbf{a}\mathbf{s}} |\boldsymbol{\varepsilon}(t)| / (N-p) \quad , \quad \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(t, \mathbf{c}) = \mathbf{x}(t) - \mathbf{c}^T \cdot \mathbf{Y}(t) \quad (18)$$

Derivation of equation (17) is the following. If we consider the dependence of value (13) on parameter \mathbf{s} , it could be noticed that the major part of this dependence is due to multipliers $1/(2\mathbf{s}^2)$ and \mathbf{a}/\mathbf{s} . If $\delta\mathbf{s}$ is a small variation of \mathbf{s} , then variation of the sums:

$$\sum_{|\boldsymbol{\varepsilon}(t)| \leq \mathbf{s}\mathbf{a}} \boldsymbol{\varepsilon}^2(t) \quad \text{and} \quad \sum_{|\boldsymbol{\varepsilon}(t)| > \mathbf{s}\mathbf{a}} |\boldsymbol{\varepsilon}(t)| \quad (19)$$

multiplied by $1/(2\mathbf{s}^2)$ and \mathbf{a}/\mathbf{s} , are much smaller than the variation of $1/(2\mathbf{s}^2)$ and \mathbf{a}/\mathbf{s} alone. Hence, for a small compute variation of $\mathbf{J}(\mathbf{c}, \mathbf{s})$ with respect to $\delta\mathbf{s}$, values (19) can be considered approximately constant, so:

$$\delta\mathbf{J}(\mathbf{c}, \mathbf{s}) \approx -\frac{(N-p)}{\mathbf{s}} \left(\frac{\alpha}{\mathbf{s}^2} - \frac{\gamma}{\mathbf{s}} \right) \delta\mathbf{s} \quad (20)$$

Applying the condition for the maximum point, $\delta\mathbf{J}(\mathbf{c}, \mathbf{s})=0$, results in one non-trivial solution for \mathbf{s} in the form of a quadratic equation: $\alpha\mathbf{r}^2 + \gamma\mathbf{r} - 1 = 0$, where $\mathbf{r}=1/\mathbf{s}$. This has a unique positive root:

$$\mathbf{r} = \chi = (\sqrt{\gamma^2 + 4\alpha} - \gamma) / (2\alpha) = 1/\mathbf{s}$$

However, the values of α and γ are dependent on \mathbf{s} , so the last equation should be used in an iterative procedure to refine the value of \mathbf{s} (the 2nd equation in (14)). It should be noted that if the value of the robustness parameter \mathbf{a} is large, then $\gamma=0$ and, according to (20), $\mathbf{s}=\sqrt{\alpha}$, which is the same as (11).

The iterative procedure starts from an initial approximation, which is given by the least squares' method (10) and (11). It was found that values converged reasonably rapidly within 5-10 iterations.

The order of the autoregression, \mathbf{p} , and number \mathbf{m} of harmonics in cyclic trend and their periods \mathbf{T}_i need to be set before estimating the parameters of the model (5). The autoregression order (\mathbf{p}) was assumed to be equal to 2 for all cases as the minimum value for describing a wide range of types of stochastic oscillations (Box and Jenkins 1970). As for choice of values of \mathbf{m} and periods \mathbf{T}_i , two possible approaches were used.

8.1 THE FIRST APPROACH – NUMBER OF HARMONICS AND THEIR PERIODS ARE DEFINED FROM FISHERY TIME SERIES THEMSELVES

Let us define the time series of more than 64 samples as “long” and others as “short” time series data sets.

For “long” time series, the number of harmonics \mathbf{m} and their periods \mathbf{T}_i are defined from power spectra estimates of the appropriate series. The value of \mathbf{m} varies from 1 to 6 and is usually equal to 3-4. Values of periods \mathbf{T}_i were taken as periods corresponded to the essential peak values of power spectra estimates. Although the time series are labelled as “long”, their lengths (from 67 up to 97 samples) has rather few observations for applying common procedures of spectra estimates based on the Fourier transform. For such a short series, parametric methods are more advantageous in terms of frequency resolution (Marple 1987). We use the autoregression approximation for the series, which is the procedure estimating parameters of the following model:

$$\mathbf{x}(t) + \sum_{j=1}^q \alpha_j \mathbf{x}(t - j) = \boldsymbol{\eta}(t) \quad (21)$$

where $\alpha_j, j=1, \dots, q$ are autoregressive parameters, $\boldsymbol{\eta}(t)$ is the residual signal, which is assumed to be Gaussian white noise with zero mean and variance σ^2 . Equation (21) applies the designations for parameters of the AR(\mathbf{q})-model different from those for the AR(\mathbf{p})-model in equations (4) and (5), to stress that although these models are the same, they are used for distinctly different purposes. The AR-members in formulae (4) and (5) are introduced to describe essential features of stochastic oscillations of the signal close to the deterministic cyclic trend. That is why we assumed a small order $\mathbf{p}=2$. The AR(\mathbf{q})-model (21) is estimated to describe the spectral structure of the signal and for this purpose needs many more parameters (i.e. larger \mathbf{q}). The larger the value of \mathbf{q} , the more sensitive is the power spectra estimate. At the same time, increase in \mathbf{q} results in the corresponding increase in the statistical fluctuations of the estimate. Thus, the choice of \mathbf{q} is a compromise between sensitivity and stability. Usually, for short series $\mathbf{q}=\mathbf{N}/5-\mathbf{N}/3$, where \mathbf{N} is the length of time series. We used $\mathbf{q}=20$. As soon as parameters of the model (21) are estimated, the power spectra estimate could be computed using the following equation:

$$S_{xx}(\omega) = \frac{\sigma^2}{2\pi \left| 1 + \sum_{j=1}^q \alpha_j \exp(-i\omega j) \right|^2} \quad (22)$$

where ω is a cyclic frequency: $\omega=2\pi/\mathbf{T}$, \mathbf{T} = period, measured in units of sampling time interval (1 year in our case), i = imaginary unit. AR-methods of power spectra estimates differ from each other by procedures of estimating parameters of the model (21). We use Burg’s method of maximum entropy as the most reliable and providing the best frequency resolution for short time series (Marple 1987).

For the “short” series, we used the model (5) with $\mathbf{p}=2$ and single harmonic cyclic trend: $\mathbf{m}=1$. Because of lack of samples in the series, we could not define the period value \mathbf{T}_1 from power spectra estimates. That is why it was defined by minimization of s^2 after estimating the autoregression parameters \mathbf{a}_1 and \mathbf{a}_2 , amplitudes $\mathbf{B}_1, \mathbf{D}_1$ and static shift constant \mathbf{G} for different trial values of the period: $s^2(\mathbf{T}_1) \rightarrow \min$. The last task of one-dimensional minimization is solved by a Golden Section method. This method will be defined as fitting an AR(2)-model with single cyclic trend with unknown period.

After the model had been fitted, it was used to forecast the interval 1997-2056 (the near 60-year cycle). The procedure is described below. For some species, the forecasted time series took negative values during some future time intervals. In such cases, we interpret these as a negligible forecasted catch and substituted all negative values with zero.

8.2 THE SECOND APPROACH – AR(2)-MODEL WITH A SINGLE CYCLIC TREND WITH THE PERIOD LENGTH TAKEN FROM THE CLIMATIC TIME SERIES

Earlier we used those periods in the equation (1), which follow naturally from the power spectra of the fishery time series, or those periods for which the cyclic trend is the best approximation to the data. However, such values of periods may be too sensitive to patterns in the data, which are not useful for forecasting. A long-term prediction based on the short-term behaviour of a signal is risky. The catch depends not only on natural processes in the ocean, but on economic politics of fishing-dependent countries, availability of fuel for fleets, power resources for fish processing, military and economic collisions in various regions of World ocean, etc. It is hardly possible to forecast all these (and other) factors for 10-20 years ahead. It is therefore prudent to confine the model to patterns, which can be established through independent means.

From all the above, it may be concluded that the period lengths derived from the fishery time series and then used in model (5), are estimated and therefore depend on realization and length of the sample. At the same time, there are time series for global climatic processes which, although believed to be somewhat affected by human activities or changes in local measurement environments, are much less "human-dependent" than fish stocks are.

These are:

- mean global temperature - **dT** time series, 138 annual samples for 1861-1998;
- atmospheric circulation index - **ACI** time series, 109 annual samples for 1891-1999;
- length of day (i. e. Earth rotation velocity) - **LOD** time series , 151 annual samples for 1850-2000.

In addition, the following long-term series were analyzed:

- Reconstructed air surface temperature for the period of 552-1975 AD by Greenland Ice Cores – 1420 annual samples.
- Reconstructed summer air surface temperature for the period of 500-1990 AD by Tree Rings of Scots pine –140 samples aggregated by a 10-year moving average.
- Reconstructed sardine and anchovy population by analysis of varved sediment cores located in Californian upwelling for the period of 500-1950 AD – 145 samples aggregated by 10–years averaging.

Climatic processes definitely affect the productivity of major commercial marine species, although in an undefined way. Thus, we applied model (5) to make forecasts, but the periods were derived not only from fishery time series, but also from using the most significant climate cycles. Applying such cyclic trends to model (5) results in more reliable values for the modelled (and forecasted) period lengths.

The values of the period lengths derived from the climatic and geophysical (LOD) time series are the following:

- for **dT** - 55, 9, 5 years (evaluation from power spectrum estimate), 64 years (evaluation from single cyclic trend fitting with unknown period);
- for **ACI** - 50, 19 years (evaluation from power spectrum estimate), 58. 5 years (evaluation from single cyclic trend fitting with unknown period);
- for **LOD** - 64, 23 years (evaluation from power spectrum estimate).

Thus, we examined the AR(2)-model with a single cyclic trend with the following modelled periods close to those derived from the above-mentioned climatic (geophysical) time series: 55, 60 and 65 years.

8.3 FORECASTING

To make the forecast, we applied a “bootstrap” method (Efron and Tibshirani 1986) to generate a probability distribution for the forecasts. The method consisted of generating a large number, \mathbf{M} (we take $\mathbf{M}=1000$), of independent stochastic artificial trajectories of the model (5), starting from the last sample of the series into the future time interval of the given length. These trajectories differ from each other only by independent realizations of the white noise component $\boldsymbol{\varepsilon}(\mathbf{t})$. All other parameters (including variance \mathbf{s}^2) are fixed as the estimates. Each trajectory presents a scenario for a future time series following model (5) drawn from a probability distribution based on the variation in data. Thus we have a “bundle” of \mathbf{M} samples of generated trajectories which fill some “strip” on the plane of (\mathbf{t}, \mathbf{x}) . For each value of \mathbf{t} in the future we can compute the average value among \mathbf{M} values, corresponding to the different realisations and a variance (a width of the projected “strip”). These mean values constitute the forecasting curve, and the variance values reflect standard deviations bars for each future \mathbf{t} value (for more explanation see Chapter 9)

Besides forecasting, we also applied the same bootstrap technique for “forecasting the past” or “hindcasting” (see figures of Chapter 9, dashed line). In this case, \mathbf{M} bootstrap trajectories were generated starting from the third rather than first data point of the series because the AR(2) used the first two values to initiate the autoregression. This procedure can be regarded as a test of the model by visual comparison between the mean values (i.e. expected from the model) and observed time series. It should be noticed, however, that such mean trajectory depends on the first two samples taken as initial values. Other methods can also be used to test model (5), such as the usual statistical criteria to test whether the residual signal $\boldsymbol{\varepsilon}(\mathbf{t})$ is Gaussian white noise or more sophisticated tools such as the AIC (Akaike Informational Criterion) to decide upon which terms to include (Kashyap and Rao 1976).