Northwest Atlantic Fisheries Organization

Serial No. N4053 **NAFO SCR Doc. 99/7**

SCIENTIFIC COUNCIL MEETING – JUNE 1999

Environmental Indices – Predicting Air Temperatures and Water Temperatures in the Northwest Atlantic By Time Series Analysis

by

M. Stein [1]) and J. Lloret [2])

[1]) Institut für Seefischerei, Palmaille 9, D-22767 Hamburg, Germany,
e-mail: stein.ish@bfa-fisch.de

[2]) Institut de Ciències del Mar-CSIC, Pg. Joan de Borbó sn, E-08039 Barcelona, Catalonia, Spain,
e-mail: lloret@icm.csic.es

Abstract

Univariate seasonal ARIMA and intervention models are used to forecast monthly mean air and bottom water temperatures from 3 sites in the Northwest Atlantic region, up to one year in advance. These models explain a reasonable amount of the total variability, with results showing a good agreement between the forecasts and the observed mean air and bottom water temperatures. The structure of the random processes that generated the temperature time series was specified for most cases as ARIMA models with moving-average terms. All fitted interventions appeared during wintertime (December-March), and therefore prediction of temperatures during these months must be taken with caution.

Keywords: ARIMA, Box-Jenkins models, forecasts, intervention analysis, water temperature, air temperature, Greenland, Newfoundland

Introduction

Forecasting of physical parameters is usually performed by using differential equation systems with given boundary conditions. An alternative methodology to forecast physical time series is that of time series analysis. In general terms, much of statistical methodology is concerned with models in which observations are assumed to vary independently. However, a great amount of data in natural sciences but also in business, economics and engineering occur in the form of time series, where observations are dependent and where the nature of this dependence is of interest itself. The body of techniques available for the analysis of such series of dependent observations is called time series analysis. A class of these techniques is the Box-Jenkins methodology (Box and Jenkins, 1976), which deals with the building of linear and stochastic-dynamic models that need minimum data requirements, e.g. monthly mean temperatures. The present study deals with the use of the Box-Jenkins univariate time series methodology, Autoregressive-Integrated-Moving-Average (ARIMA), and intervention models to forecast (up to one year in advance) dynamics of air and bottom water temperatures for 3 stations in the Northwest Atlantic region.

In the following a chapter on Material and Methods is given where a special computer software is presented which suits well for prognostic issues, provided environmental data are available on monthly basis.

Under Results three case studies are presented (two air temperature time series from West Greenland, one ocean temperature time series from Station 27 off St. John's, NFLD, Canada). In a final chapter the authors discuss the usefulness of prognostics in environmental data, and point at the risks when performing an environmental prognosis.

Materials and Methods

Two series of monthly mean air temperatures from Greenland, Egedesminde (68°42.5'N, 51°44.5'W), and Nuuk (64°11'N, 51°44.5'W) were taken for analysis (January 1991-December 1998 for Egedesminde, and January 1960-December 1998 for Nuuk). A time series of monthly mean bottom water temperatures at Station 27 (off St. John's, Newfoundland) was made available by the Canadian Department of Fisheries and Oceans in St. John's, Newfoundland for the period January 1950-December 1998. Missing monthly data from the Station 27 time series were interpolated from the previous and the following months.

To fit ARIMA univariate and intervention models to the time series of air and bottom water temperatures, the software package Force 4/R Research System (Prat et al., 1998) developed by the Politechnical University of Catalonia was used. This package is based on the Box-Jenkins methodology (Box and Jenkins, 1976), and given a specific time series it is possible to fit univariate and multivariate models with several options of transformations, e.g. logarithmic transformation and differencing. Furthermore, it allows to analyse its seasonal pattern and trend by means of multiple moving averages.

Box-Jenkins (Box and Jenkins, 1986) models are constructed using only the information contained in the series itself. Thus, models are constructed as linear functions of past values of the series and/or previous random shocks (or errors). Forecasts are generated under the assumption that the past history can be translated into predictions for the future.

Box and Jenkins (1976) formalised the ARIMA modelling framework by defining three steps to be carried out in the analysis, as follows: identify the model, estimate the coefficients and verify the model. These linear, stochastic procedures apply to stationary series (time series with no systematic change in mean and variance) whose data are normally distributed. First-or second-order differencing (non-seasonal and/or seasonal) usually takes care of non-stationary mean. Logarithmic transformation of the raw data takes care of non-stationary variance and also of non-normal distribution of original data. Identification of the model (i.e. how many terms to be included in the model) is based on the examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the differenced, log-transformed time series. Estimation of the coefficients of the model is done by means of the maximum likelihood method. Verification of the model is done through diagnostic checks of the residuals (histogram and normal probability plot of residuals, standardised residuals and ACF and PACF of the residuals). The ability to forecast using ARIMA models has been tested by applying the final fitted model to data obtained after the last observation used to estimate the model parameters. Thus, ARIMA models have been constructed using all available data but excluding monthly data of the last year, which was used to compare with forecasts obtained for that year. Detailed description of the non-seasonal and seasonal ARIMA models and the standardised notation used in this paper is developed in Appendix 1.

The intervention analysis is a statistical technique which allows us to detect and quantify non-random changes of a variable in a time series (Box and Jenkins, 1976; Chatfield, 1984; Pankratz, 1991). As documented recently (Stein, 1998) there is large variability in the air temperatures and ocean temperatures in the Northwest Atlantic region. These changes might be either random processes or non-significant in a statistical sense. Owing to the high serial correlation (lack of independence between successive observations), the t-tests for equality of means cannot be used to test for temperature shifts. Therefore intervention analysis is used here to identify the significance, magnitude and form of structural shifts (interventions) at the temperature time series. While the input of an intervention represents a pulse shift in a given month, the output or consequence of that event may be modelled in several ways. Thus, according to the output, two types of interventions are defined, i.e. pulse and step. A pulse intervention represents a temporary event that affects the level of the temperature regime, and can be modelled as abrupt (i.e. a pulse intervention at t=1 shifts the level up or down only during period t=1) or delayed (i.e. a pulse intervention at t=1 causes a decreasing or a increasing reponse during periods t+1, t+2, t+3...). While the first one is also called *Additive Outlier* (denoted AO), the second one is also called *Temporary-Change* intervention (denoted TC). Step interventions may be thought as a permanent change in the level of a time series. They are also called *Level-Shifts* (denoted LS).

By carrying out the intervention analysis we will not only obtain better models (estimated parameters will improve) and better forecasts (in case that the invention occurs in the last values used to model the series) but also it will be possible to know in a better way the time series under study (we will be able to detect possible external events and try to explain them).

Results

Forecasting monthly mean air temperatures and bottom water temperatures

The final ARIMA models fitted to the monthly mean air and bottom water temperatures are presented in Table 1. It was not necessary to log-transform the raw data before fitting the models because of a steady variance. Due to the strong seasonal cycle, seasonal differencing was required in all cases. Regular differencing was only required in the time series of bottom water temperature at Station 27, because of the observed trend. Therefore, seasonal ARIMA models were fitted to all time series: ARIMA model (0,1,1)x12(0,1,1) for the bottom water temperatures at Station 27 and ARIMA models (0,0,1)x12(0,1,1), and (1,0,0)x12(0,1,1) for Egedesminde and Nuuk air temperatures, respectively. Moving-average terms were thus appearing in all fitted models. The forecasts obtained by such a kind of model have a special interpretation: they are exponentially weighted moving averages of the available data.

The amount of variability explained by the models ranged from 80% for bottom water temperature at Station 27 to 90% for both time series of air temperatures. The fitted model was compared with the observed past values for each time series (Figures 1, 3, 5). Nevertheless, the most important check of any model is to compare predicted values with actual data not used in fitting the model. Thus, the ideal situation is to extend the forecast beyond the last observation used in parameter estimation (Pankratz, 1991). Therefore, one year ahead forecasts of the monthly air and bottom water temperatures were made, and were compared with the observed values afterwards (Figures 2, 4, 6, 7). These comparisons demonstrate the predictive ability of the ARIMA models.

Intervention analysis

Globally, 13 significant interventions ($p<0.05$, $t>3.0$ or $t<-3.0$) were incorporated to the ARIMA models, which are given in Table 2. All interventions appeared during wintertime (December-March), e.g. February 1959 *Level-shift* and January 1978 and January 1983 *Additive Outliers* at the bottom water temperatures of Station 27 (Figure 8). The occurrence of these anomalies during wintertime influences the predictions for these months. Thus, for example, the model for the time series of bottom water temperature constructed using data from January 1950 to January 1998 (model 4 in Table 1) produces better forecasts than the model constructed up to December 1997 (model 3 in Table 1; Figures 6, 7).

Discussion

The univariate models for air temperatures and bottom water temperatures revealed that it is reasonable to predict their future dynamics based on its past temperature figures alone. Statistically, the structure of the random processes that generated the temperature time series was specified for most cases as ARIMA models with moving-average terms. These models often arise in practice (Pankratz, 1991) and their forecasts have a special interpretation: they are exponentially weighted moving averages of the available data. This might suggest that factors affecting the time-series of air temperatures and bottom water temperatures in the Northwest Atlantic region, e.g. currents, North Atlantic Oscillation (NAO) index, etc, tend to persist for a long time.

Intervention models developed in this paper for air temperatures and bottom water temperatures indicated the existence of important and re-appearing temperature anomalies during winter in the Northwest Atlantic region. The fact that all these interventions appeared during wintertime indicates that this season shows a high variability and therefore predictions for winter months must be taken with caution. It must be realised that interventions cannot be predicted with the used methodology. Thus, incorporation of winter data into the ARIMA models and forecasting afterwards would increase the reliability of predictions.

Although the Box-Jenkins methodology (Box and Jenkins, 1976) has been used mainly by the industrial business managers, economists and engineers, the use of time-series analysis to model fish population dynamics (e.g. Lloret et al. 1999; Stergiou et al. 1997; Quinn and Marshall, 1989) and physical variables (Irvine and Eberhardt,1992) has increased in recent years. These models are important not only because they may tell us something about the nature of the system generating the time series and their ability to produce optimal forecasts of future values of the series, but also because they do not present the statistical problems usually appearing when using the deterministic and stochastic-static models, i.e. autocorrelation of the observations, colinearity, residual autocorrelation and non-normal distribution of residuals, which finally may bias the fit. Furthermore, time series models do take into account time-lagged relationships between the variables and, when two or more related series are under study, the models can be extended to represent dynamic relationships between the series.

Acknowledgements

The authors would like to thank Dr. Ignasi Solé, Dr. Albert Prat and Dr. J.M. Catot from the Politechnical University of Catalonia for allowing us to use the time series software package Force 4/Research System.. We extend our gratitude to E.B. Colbourne, Department of Fisheries and Oceans, St. John's, Newfoundland, Canada for making available the monthly bottom water temperature data at Station 27. Josep Lloret was financially supported by the D.G. Research of the Government of Catalonia.

References

BOX, G.E.P., and G.M. JENKINS 1976. Time Series Analysis: Forecasting and Control. Prentince Hall, Inc. 575 pp.

CHATFIELD, C. 1984. The analysis of time series: an introduction. Chapman & Hall. 286 pp.

GÓMEZ, V., and A. MARAVALL 1997. Tramo-Seats-SCA. Politechnical University of Catalonia. Dpts. of Statistics. High School of Engineering. Barcelona.

IRVINE, K.N., and J. EBERHARDT 1992. Multiplicative, seasonal ARIMA models for Lake Erie and Lake Ontario water levels. Water Resources Bulletin, 28(2):385-396.

LLORET, J., LLEONART, J., and I. SOLÉ 1999. (in press). Forecasting and modelling fishery resources of Gulf of Lions-Cape Creus (NW Mediterranean) by time series analysis. ICES Journal of Marine Science.

PANKRATZ, A. 1983. Forecasting with Univariate Box-Jenkins Models. John Wiley & Sons, Inc. 562 pp.

PRAT, A., CATOT, J.M., and I. SOLÉ 1998. Force 4/R Research System. Politechnical University of Catalonia. Dpts. of Statistics. High School of Engineering. Barcelona.

QUINN, T.J., and R.P. MARSCHALL 1989. Time series analysis: quantifying variability and correlation in SE Alaska salmon catches and environmental data. Can Spec. Publ. Fish. Aquat. Sci., 108: 67-80.

STERGIOU, K.I., E.D. CHRISTOU, and G. PETRAKIS 1997. Modelling and Forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods. Fisheries Research, 29: 55-95.

STEIN, M. 1998. Integrating Fisheries Observations with Environmental Data – Towards a Better Understanding of the Conditions for Fish in the Sea. J. Northw. Atl. Fish. Sci., Vol. 23: 143-156.

Table 1. Seasonal ARIMA models fitted to the monthly temperatures of Egedesminde, Nuuk and Station 27 (NFLD). Time period used to fit the models, and the resulting models with the r^2 are shown. AR is the non-seasonal autoregressive term, while MA and MAS are the non-seasonal and seasonal moving average terms, respectively.

Station	Time period	Fitted model	Model No	r^2	AR-1	MA-1	MAS-1
Egedesminde Air Temperature	January 1991-December 1997	(0,0,1)x12(0,1,1)	1	0.93		-0.429	0.296
Nuuk Air Temperature	January 1960-December 1997	(1,0,0)x12(0,1,1)	2	0.92	0.334		0.993
Station 27 (NFLD) Bottom Water Temperature	January 1950-December 1997	(0,1,1)x12(0,1,1)	3	0.80		0.006	0.955
Station 27 (NFLD) Bottom Water Temperature	January 1950-January 1998	(0,1,1)x12(0,1,1)	4	0.80		0.006	0.958

Table 2. Interventions incorporated to the ARIMA models: LS (Level-Shifts), AO (Additive Outliers), TC (Temporary-Changes) and their respective values and significance (t-value); only significant interventions at t-values > +3 or < -3 are given, which denotes positive/negative anomalies, respectively.

Station	Date	LS		AO		TC	
		Value	t	Value	t	Value	t
Egedesminde Air Temperature	Feb.92			-9.55	-4.83		
Nuuk Air Temperature	Mar.62			7.45	4.63		
	Dec.71					-6.92	-4.54
	Dec.78					6.45	4.23
	Jan.83					-9.38	-6.14
	Jan.84					-12.45	-8.06
	Mar.84			7.19	4.46		
	Feb.86			7.57	4.74		
	Feb.92			-9.68	-6.04		
	Jan.93					-6.99	-4.55
Station 27 (NFLD) Bottom Water Temperature	Feb.59	-0.90	-4.70				
	Jan.78			-0.64	-4.58		
	Jan.83			0.79	5.67		

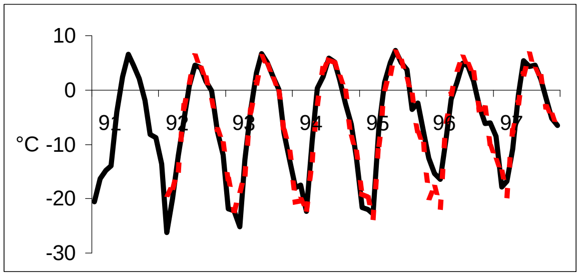

Fig. 1 Fit of the ARIMA model 1 (see Table 1) of Egedesminde monthly mean air temperatures (dashed line) to observed data (solid line) for the period 1992-1997.

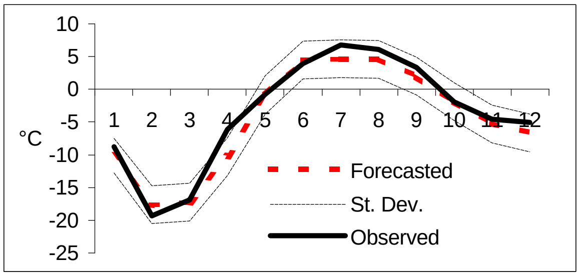

Fig. 2 Comparison between observed Egedesminde monthly mean air temperatures and 1 year ahead forecasts predicted from model 1 (see Table 1) for the period January–December 1998

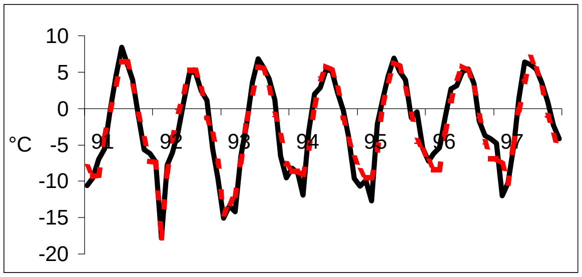

Fig. 3 Fit of the ARIMA model 2 (see Table 1) of Nuuk monthly mean air temperatures (dashed line) to observed data (solid line) for the period 1991-1997.

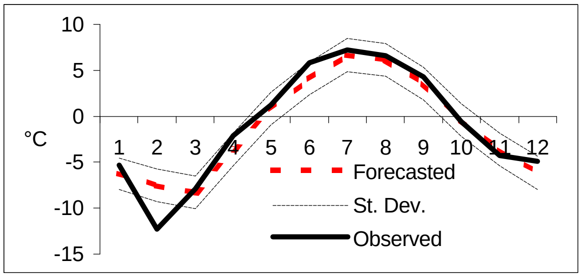

Fig. 4 Comparison between observed Nuuk monthly mean air temperatures and 1 year ahead forecasts predicted from model 1 (see Table 1) for the period January–December 1998

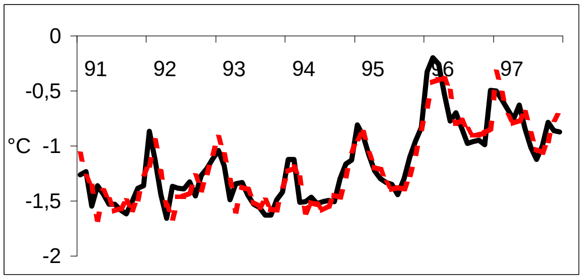

Fig. 5 Fit of the ARIMA model 3 (see Table 1) of Station 27 monthly mean bottom water temperatures (dashed line) to observed data (solid line) for the period 1991-1997.

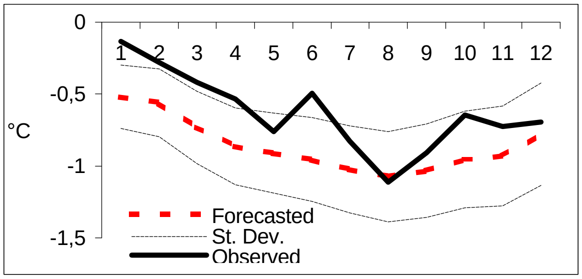

Fig. 6 Comparison between observed monthly mean bottom water temperatures and 1 year ahead forecasts predicted from model 3 (see Table 1) for the period January–December 1998 at Station 27.

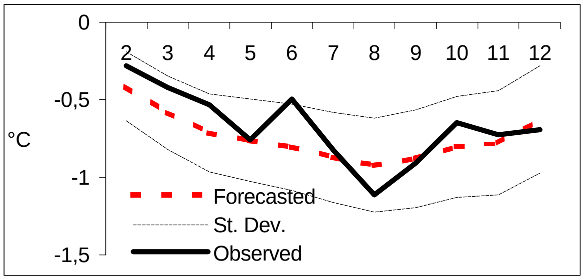

Fig. 7 Comparison between observed monthly mean bottom water temperatures and 1 year ahead forecasts predicted from model 4 (see Table 1) for the period February–December 1998 at Station 27.

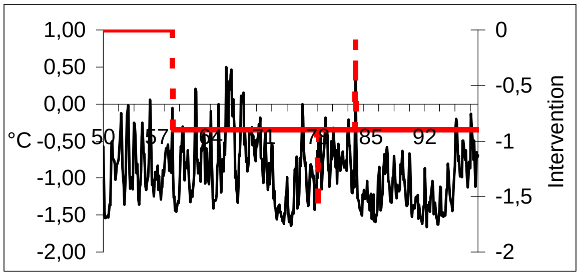

Fig. 8 Level-Shift (LS) and Additive Outliers (AO) interventions in February 1959, January 1978 and January 1983, respectively, for the whole time series (1950-1978) of the bottom water temperature in Station 27. The effect of the intervention is shown by a dashed line while the observed series is plotted with a solid line.

Appendix 1: Standardized ARIMA notation

Simple non-seasonal ARIMA model has a general form of (p,d,q) where p is the order of the non-seasonal autoregressive term (AR), q is order of the the non-seasonal moving average term (MA) and d is the order of non-seasonal differencing. AR is describing how a variable Z_t depends on some and well defined previous temperature values Z_{t-1} (AR-1), Z_{t-2} (AR-2)... while MA describes how this variable Z_t depends on a weighted moving average of the available data Z_{t-1} to Z_{t-n}. For example, for a one-step ahead forecast (say, for period t, an October), with an AR-1, all weight is given to the temperature of the immediately previous month (September), and with an AR-2 the weight is given to the temperature of the 2 immediately previous months (September and August). By contrast, with an MA-1 or MA-2, a certain weight is given to the temperature of the immediately previous month (September), a smaller weight is given to the temperature observed two months ago (August) and so forth, i.e. the weights decline in value exponentially.

The seasonally exhibited in all temperature time series renders simple ARIMA modelling inadequate. In the case of a seasonal time series there is a relationship between month Z_t and Z_{t-s} where s is the seasonal timespan. Thus, the multiplicative, seasonal modelling approach which has a general form of ARIMA(p,d,q) x s(P,D,Q) has been used in this paper. In this general form, P is order of the seasonal autoregressive term (ARS), Q is the order of the seasonal moving average term (MAS), D is the order of seasonal differencing and s is the seasonal span (e.g. s=12 for an annual trend in monthly data). Thus, ARS is describing how a variable Z depends on some and well defined previous values Z_{t-12} (ARS-1), Z_{t-24} (ARS-2)..., while MAS describes how this variable Z depends on a weighted moving average of the available data Z_{t-12} to Z_{t-12n}. For example, for a one-step ahead forecast (say, for period t, an October) and with an ARS-1, all weight is given to the temperature of the same season (October) 1 year while with an ARS-2 the weight is given to October temperature 1 and 2 years ago. By contrast, with an MAS-1 or MAS-2, the model gives a certain weight to October temperature 1 year ago, to the October temperature 2 years ago, and so on. These weights decline exponentially too. The standardised notation used in this paper to represent ARIMA (p,d,q)x12(P,D,Q) models from which the model equation is obtained is $\phi_p(B)\Phi_P(B^s)\ \nabla_s^D\ \nabla^d\ Z_t = C+\theta_q(B)\Theta_Q(B^s)a_t$, where:

- $\phi_p(B)=1-\phi_1B-\phi_2B^2-...\phi_pB^p$ is the nonseasonal autoregressive operator of order p.
- $\Phi_P(B^s)=1-\Phi_1 B^s -\Phi_2 B^{2s}-...\Phi_PB^{Ps}$ is the seasonal autoregressive operator of order P.
- ∇_s^D is the seasonal differencing operator of order D.
- ∇^d is the nonseasonal differencing operator of order d.
- Z_t is the value of the variable of interest at time t.
- $C=\mu\phi_p(B)\ \Phi_P(B^s)$ is a constant term, where μ is the true mean of the stationary time series being modeled. It was estimated from sample data using the approximate likelihood estimator approach.
-$\theta_q(B)= 1-\theta_1B-\theta_2B^2-...\theta_qB^q$ is the nonseasonal moving average operator of order q.
- $\Theta_Q(B^s)= 1-\Theta_1 B^s -\Theta_2 B^{2s}-...\Theta_QB^{Qs}$ is the seasonal moving average operator of order Q.
- $\phi_1, \phi_2,..., \phi_p; \Phi_1, \Phi_2, ..., \Phi_P; \theta_1,\theta_2,...\theta_q; \Theta_1, \Theta_2,..., \Theta_Q$ are unknown coefficients that were estimated from sample data using the approximate likelihood estimator approach.
- a_t is the error term at time t
- s is the seasonal span

When modelling, two different options can be used: tramo-seats and SCA options (Gómez and Maravall, 1997). The first one (tramo-seats) is able to construct good models in a reasonable time for time series whose structure is relatively easy. This option has been used in our study in all cases.