Time series analysis to forecast temperature change

Tanja Van Hecke[1]
University College Ghent

Abstract

ARIMA models are often used to model the evolution in time of economic issues. We demonstrate that an ARIMA model is also valuable in the environmental field, where the evolution of climate change is causing many concerns. Can we confirm the global warming by mathematical prediction theories?

Keywords: time series, ARIMA model, prediction, temperature change
2000 Mathematics Subject Classification:
Primary 60G25; Secondary 62M20

1 Introduction

In Belgium temperature has been recorded since 1850. To model the evolution of this temperature as part of the climate change, we have used data from the environmental report Flanders (MIRA) [7]. As a measure of the change of temperature we used the variable y, where

$$y = \text{mean yearly temperature} - \text{mean temperature during the period 1850-1899}.$$

This is plotted in Figure 1 together with the 10 year moving average. In Figures 2 and 3 the same measure is presented for European and worldwide data respectively, during the same time period [7]. Our aim is to forecast the evolution of temperature based on these time series.

In real-life research and practice, data patterns are hidden and individual observations are subject to errors. But to make reliable forecasts, we need a mathematical model of the process. Time series are characterized by the dependence of their data. As the independence of the data is one of the assumptions of regression analysis where time is the independent variable, this method is inappropriate to model the trend of a time series. To analyze dependence data, Box & Jenkins [3] [4] developed ARIMA models for time series; these are well described in [6] and [2].

[1]Received 16 November 2009; revision received
Postal address: University Collgeg Ghent, Faculty of Applied Engineering Sciences, Ghent, Belgium, Email address: tanja.vanhecke@hogent.be

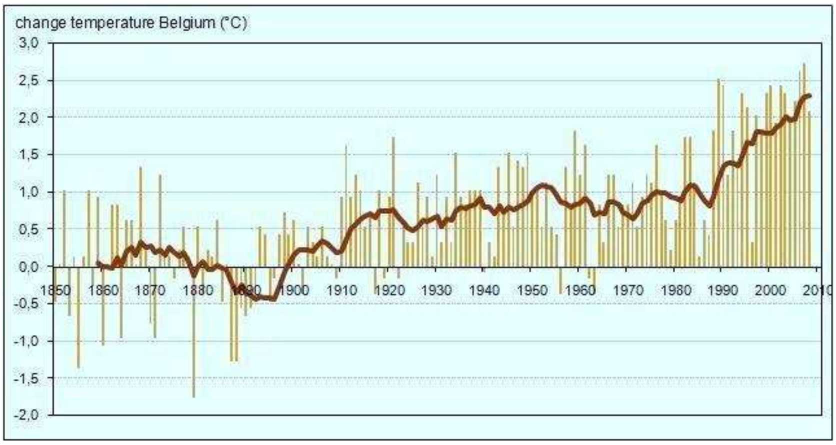

Figure 1: Difference data and 10 year moving average of Belgian temperatures.

2 The ARIMA model

2.1 Definitions

The acronym ARIMA stands for *Auto-Regressive Integrated Moving Average.* Algebraically the ARIMA model can be defined by

$$y_t = \mu + b_1\, y_{t-1} + \ldots + b_p\, y_{t-p} + \epsilon_t - a_1\, \epsilon_{t-1} + \ldots - a_q\, \epsilon_{t-q} \qquad (1)$$

at time $t = 1, \ldots, n$, where ϵ_{t-j} $(j = 0, 1, \ldots, q)$ are the lagged forecast errors. The $p+q+1$ unknown parameters μ, $b_1, \ldots b_p$ and $a_1, \ldots a_q$ are determined by minimizing the squared residuals. In the first part of the righthand side of (1) the dependent variable y_t is predicted, based on its values at earlier time periods. This is the *autoregressive* (AR) part of the equation (1). In the second part, the dependent variable y also depends on the values of the residuals at earlier time periods, which may regarded as prior random shocks. This is the *moving average* (MA) part of the equation. It is the task of the researcher analyzing a given time series, to find the relevant parameters of the ARIMA(p, d, q) model with

p the number of autoregressive terms

d the number of nonseasonal differences

q the number of lagged forecast errors in the prediction equation.

Two goals must be met, namely to find the most effective model and to restrict the number of parameters. The residuals should also fulfil the conditions of independence and normality.

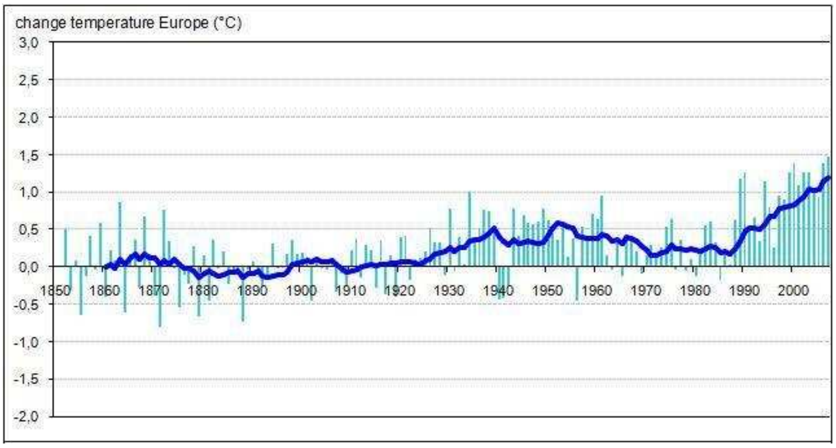

Figure 2: Difference data and 10 year moving average of European temperatures.

2.2 Stationary time series

An important condition for ARIMA models and possibly their weak point is that the time series should be stationary (i.e. the mean and variance should be constant as a function of time) before the analysis can be carried out. Otherwise, past effects would accumulate and the values of successive y_t's would move towards infinity, that is, the series would not be stationary. The observations with ARIMA models should be filtered first by differencing the observations d times, using $\Delta^d y_t$ instead of y_t as the time series to obtain stationary data, with

$$\Delta y_t = y_t - y_{t-1}. \tag{2}$$

A time series which needs to be differenced to be made stationary is said to be an *integrated* version of a stationary series. To analyze the stationary character of the time series we have used the Box-Pierce test with the summed squares of the sample correlations as the test statistic [4]. The p-value of this test should be compared with the significance level α. When the p-value is too small, the null hypothesis that states that the time series is stationary, will be rejected.

2.3 Identification and evaluation of the Model

To identify the appropriate ARIMA model for a time series, one begins by identifying the order of differencing needed to obtain a stationary series and to remove the major seasonality. If one stops after the first differencing and predicts that the differenced series is constant, one has merely fitted a random

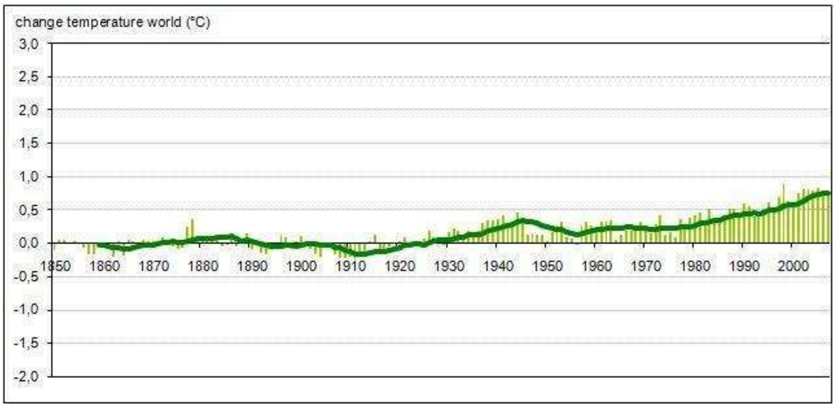

Figure 3: Difference data and 10 year moving average of global temperatures.

walk or random trend model. However, the best random walk or random trend model may still have autocorrelated errors, suggesting that additional factors of some kind are needed in the prediction equation.

In addition to the standard autoregressive and moving average parameters, the ARIMA model (1) may also include a constant μ. The interpretation of such a (statistically significant) constant depends on the model that is fitted:

- if there are no autoregressive parameters in the model, then the expected value of the constant is the mean of the series,
- if there are autoregressive parameters in the series, then the constant represents the intercept,
- if the series is differenced, then the constant represents the mean or intercept of the differenced series.

To evaluate the parameter estimates, a t-test is performed. If not significant, the respective parameter can in most cases be dropped from the model without substantially affecting the overall fit of the model.

A good model should produce statistically independent residuals that contain only noise and no systematic components. To ensure this one can plot the residuals and inspect them for any systematic trends.

To compare models the AIC (Akaike information criterion) [1] is used as a measure of goodness of fit. The test statistic (3)

$$\text{AIC} = 2\,k + n\,[\ln(2\,\pi\,RSS/n) + 1] \tag{3}$$

is used, with k the number of parameters in the model, n the number of observations and RSS the residual sum of squares. A model with a lower AIC value is preferred as it better explains the data with a minimum of free parameters.

3 Analyzing the data

We used the 10 year moving average described in Section 1. The computations were carried out by means of the software *Time Series Modelling v4.30* [8]. We used the Box-Pierce test to investigate whether the series is stationary. For all data one differentiation was needed to make the time series stationary, so $d = 1$.

Tables 1, 2 and 3 contain several measures to evaluate the ARIMA model (restricted to 3 parameters) for different p and q values based on Belgian, European and global data respectively. The AIC value (3), the p-value of the Box-Pierce test and the significance levels of the different parameters of the model (with and without constant) for the null hypothesis H_0 : parameter=0, are given.

p	0	0	1	1	2	2
q	no μ	with μ	no μ	with μ	no μ	with μ
0			152 0.401 0.13(b_1)	153 0.126 0.06(μ) 0.229(b_1)	150 0.438 0.19(b_1) 0.168(b_2)	151 0.169 0.094(μ) 0.29(b_1) 0.271(b_2)
1	153 0.388 0.14(a_1)	153 0.128 0.053(μ) 0.241(a_1)	151 0.415 0.178(b_1) 0.397(a_1)	152 0.140 0.158(μ) 0.110(b_1) 0.306(a_1)	149 0.260 0.274(b_1) 0.072(b_2) 0.162(a_1)	
2	153 0.305 0.181(a_1) 0.103(a_2)	154 0.119 0.095(μ) 0.252(a_1) 0.180(a_2)	151 0.301 0.822(b_1) 0.396(a_1) 0.103(a_2)			

Table 1. Identification and evaluation parameters of the ARIMA model for Belgian temperature changes (restricted to three parameters).

p	0	0	1	1	2	2
q	no μ	with μ	no μ	with μ	no μ	with μ
0			226 0.761 $0.594(b_1)$	227 0.804 $0.066(\mu)$ $0.777(b_1)$	227 0.643 $0.624(b_1)$ $0.001(b_2)$	228 0.623 $0.115(\mu)$ $0.78(b_1)$ $0.003(b_2)$
1	227 0.786 $0.589(a_1)$	229 0.847 $0.060(\mu)$ $0.772(a_1)$	229 0.887 $0.000(b_1)$ $0.000(a_1)$	226 0.767 $0.051(\mu)$ $0.000(b_1)$ $0.000(a_1)$	227 0.513 $0.178(b_1)$ $0.002(b_2)$ $0.081(a_1)$	
2	231 0.507 $0.403(a_1)$ $0.001(a_2)$	231 0.480 $0.117(\mu)$ $0.501(a_1)$ $0.002(a_2)$	229 0.439 $0.160(b_1)$ $0.098(a_1)$ $0.001(a_2)$			

Table 2. Identification and evaluation parameters of the ARIMA model for European temperature changes (restricted to three parameters).

p	0	0	1	1	2	2
q	no μ	with μ	no μ	with μ	no μ	with μ
0			417 0.300 $0.000(b_1)$	418 0.464 $0.078(\mu)$ $0.000(b_1)$	414 0.290 $0.000(b_1)$ $0.732(b_2)$	415 0.488 $0.073(\mu)$ $0.000(b_1)$ $0.969(b_2)$
1	411 0.607 $0.000(a_1)$	414 0.645 $0.004(\mu)$ $0.000(a_1)$	417 0.234 $0(b_1)$ $0.419(a_1)$	418 0.362 $0.362(\mu)$ $0.011(b_1)$ $0.659(a_1)$	413 0.238 $0(b_1)$ $0.162(b_2)$ $0.045(a_1)$	
2	413 0.603 $0.000(a_1)$ $0.008(a_2)$	416 0.648 $0.008(\mu)$ $0.000(a_1)$ $0.024(a_2)$	419 0.327 $0(b_1)$ $0.003(a_1)$ $0.006(a_2)$			

Table 3. Identification and evaluation parameters of the ARIMA model for global temperature changes (restricted to three parameters).

Based on Table 1 and Table 3, one can see that the ARIMA(1, 1, 0) method with constant is appropriate to predict the temperature evolution in Belgium and worldwide. Hence

$$\Delta y_t^b = 0.01387 + 0.11145\, y_{t-1} + \epsilon_t \tag{4}$$

is a suitable model for Belgium, while

$$\Delta y_t^w = 0.00223 + 0.58293\, y_{t-1} + \epsilon_t \tag{5}$$

is suitable worldwide. Based on Table 2, the ARIMA(1, 1, 1) method without constant seems to be more appropriate to predict the temperature evolution in Europe, so that

$$\Delta y_t^e = 0.95734\, y_{t-1} + 0.88786\, \epsilon_{t-1} + \epsilon_t. \tag{6}$$

The predicted values over 40 years for the three geographical regions are plotted in Figures 4, 5 and 6 together with the confidence interval in which the future value will lie with a probability of 95%.

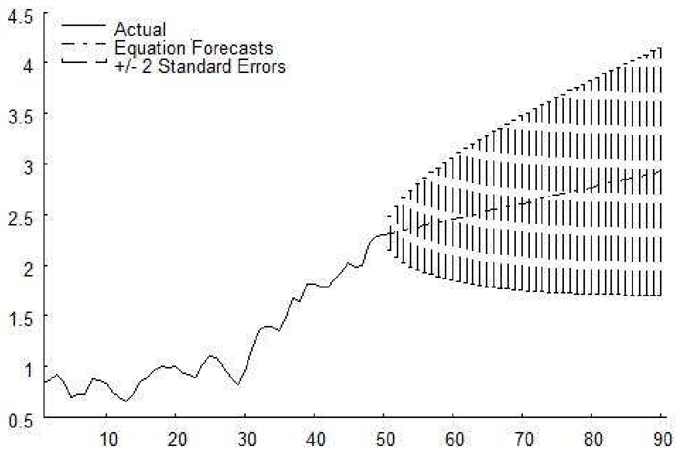

Figure 4: Predicted evolution of Belgian temperatures over 40 years with 95% confidence intervals.

4 Conclusions

By means of ARIMA models it is possible to predict the evolution of temperature based on collected data of the past 150 years. Based on the model that fits best,

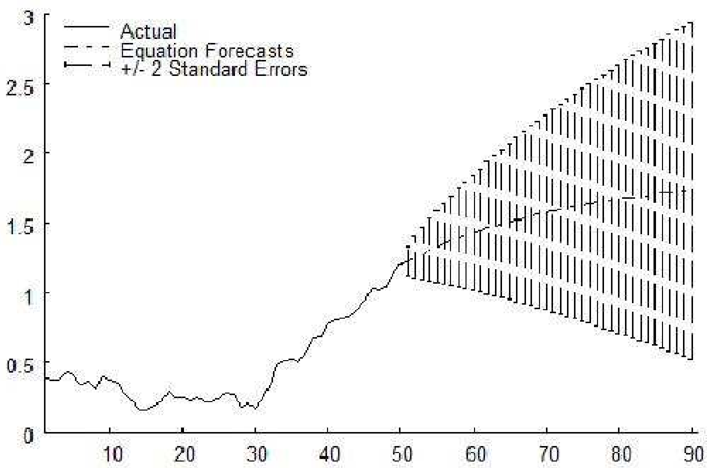

Figure 5: Predicted evolution of European temperatures over 40 years with 95% confidence intervals.

the temperature will still slightly rise from that of the reference period 1850-1899 (more pronounced for Europe and Belgium). Of course, as we advance in time, the uncertainty about the predictions grows, so these results must be treated as tentative.

References

[1] Akaike, H.(1974). A new look at the statistical model identification, IEEE Transactions on Automatic Control **19** (6), pp. 716-723.

[2] Anderson, O.D. (1975). Time Series Analysis and Forecasting, Butterworth & Co.

[3] Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis, Forecasting and Control, Holden-Day.

[4] Box, G.E.P. and Pierce, D.A. (1970). Distribution of the Autocorrelations in Autoregressive Moving Average Time Series Models, Journal of American Statistical Association **65**, pp. 1509-1526.

[5] Gaynor, P.E. and Kirkpatrick, R.C. (1994). Introduction to Time-series Modeling and Forcasting in Business and Economics, McGraw-Hill.

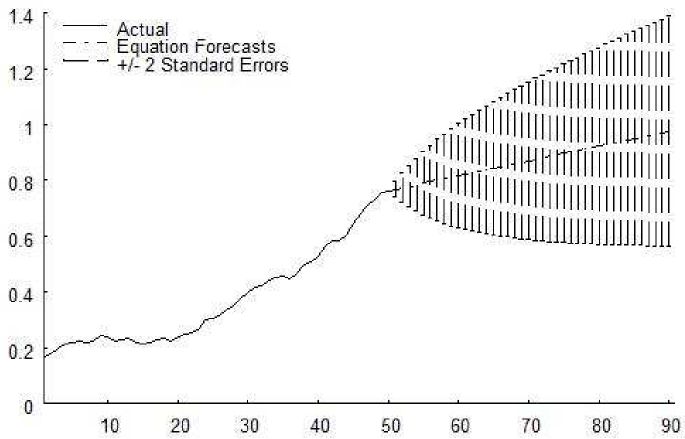

Figure 6: Predicted evolution of worldwide temperatures over 40 years with 95% confidence intervals.

[6] McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A. (1980). Interrupted Time Series Analysis, Sage Publications Ltd.

[7] http://www.milieurapport.be/nl/feiten-cijfers/MIRA-T/milieuthemas/

[8] http://www.timeseriesmodelling.com/