Space Time ARMA Models and their application to Radioactivity in Portugal

Salvador Freire
Instituto Superior Técnico
Universidade de Lisboa
November 2019

Abstract: Space time models have been proven useful to model many time series simultaneously. A review of space time models and univariate time series models are presented, and also a modelling procedure for both cases. The modelling procedure includes an order determination method, a parameter estimation and a diagnosis checking of the selected models. As a case study, radioactivity data collected in eight portuguese sites were analysed. The data were modelled for both cases: with a space time model and with separated univariate models. Then a comparison have been made in terms of the capability of forecasting.

1 Introduction

Energy of radioactivity origin, has been really helpful and efficient for a long time. It's application covers a wide range of areas, such as agriculture, medicine and the production of nuclear energy. However, any leakage, may cause tremendous environmental effects. For example, the accident at the nuclear power station in Chernobyl, in 1986, is still polluting areas within a radius of one hundred kilometers, causing 46 deaths.

Given the possible damage caused by any radioactivity accident, this thesis will focus mainly on describing radioactivity data collected in eight portuguese sites between 2008 and 2017. In this thesis, a detail of the theory of time series models will be complied (for both univariate and space time models), their application to the radioactivity data set, discussing the capability of the selected models to the data set and the accuracy of forecasts.

The main feature of this thesis is the application of space time autoregressive and moving average models (STARMA) to radioactivity data. In a space time series model, each observation at a time instant t and at a location i, is given by a linear combination of the past observations lagged in time and space. This spatial dependency between the N sites is given through weighting matrices $N \times N$. Theses matrices are defined by the model builder, and must reflect the space structure of the data. In order to simplify, in this thesis, it will be considered only a single weighting matrix, where each entry is the inverse of the euclidean distance between all sites.

The rest of the thesis is organised as follows: firstly, a review of autoregressive and moving average (ARIMA) and STARMA models, will be detailed in the section 2. This section includes an overview of some properties of the models, such as the conditions of the existence of a stationary solution, identification methods to select suitable models to fit a data set, a parameters estimation and a diagnosis checking.

In section 3, autoregressive and moving average models have been tested in order to check if a space-time series produces a better fit than univariate time series. Both univariate and space time models produced accurate models to fit the observed data. In terms of forecasting, both modelling procedures have predicted correctly the data, even though the univariate modelling has produced slightly better results.

In conclusion, it must be highlighted that both the univariate and the space time modelling have produced accurate results to model the eight radioactivity time series.

The modelling procedure for the

application was implemented in the software R Studio.

2 Time Series Modelling

The modelling procedure are going to be shown upon the incoming section. Both univariate and space time modelling will be defined, and also a useful procedure to select suitable models to represent data sets.

2.1 ARMA processes

The process (X_t) is an ARMA (p,q) process if X_t is represented as

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} \\ &\quad + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots + \beta_q \epsilon_{t-q} + \epsilon_t, \end{aligned}$$

where $\alpha's$ and $\beta's$ are constants, and ϵ_t is white noise with zero mean and constant variance. As an alternative, the process can be also defined using the backshift operator B, which operates on an element to produce the previous one, so

$$\begin{aligned} &(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p) X_t = \\ &\qquad (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q)\epsilon_t, \end{aligned}$$

or

$$\phi(B) X_t = \theta(B)\epsilon_t.$$

where

$$\phi(B) \; = \; (1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p),$$

and

$$\theta(B) \; = \; (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q),$$

In order to described a data set using an ARMA model, the Box-Jenckins procedure will be applied. This procedure is divided into the following stages:

- Identification of suitable models to fit the data by analysing the sample autocorrelation function (ACF - is the correlation between series values that are fixed intervals apart) and the sample partial autocorrelation function (PACF - is the correlation between series values that are fixed intervals apart, accounting for the values between the time intervals);
- the parameter estimation of the selected models;
- Diagnosis checking of the selected models.

If some selected model failed on the last stage, the procedure should be repeated until the selected model meet the diagnosis checking requirements.

2.2 Space Time ARMA processes

In general, Space Time ARMA processes, are highly effective to model many time series with a spatial dependence.

A spatial dependence is included in the model through a spatial lag operator. Let $\mathbf{L^1}$ be the spatial lag operator (it operates on an element to produce the previous ones taking into account a space dependence) and a weighting matrice $\mathbf{W^1}$ of spatial order l, such that

$$L^0 \mathbf{X_t} = \mathbf{X_t},$$

$$L^{(l)} \mathbf{X_t} = \mathbf{W}^{(l)} \mathbf{X_t} \; \forall \; l > 0,$$

and

$$\sum_{j=1}^{N} w_{ij}^{(l)} = 1,$$

where $w_{ij}^{(l)}$ is each entry of $\mathbf{W}$, $\mathbf{W^0} = \mathbf{I_N}$ and $i = 1, \ldots, N$, where N is the number of sites. The set of weighting matrices $(\mathbf{W^0}, \mathbf{W^1}, \ldots)$ have to be defined by the model builder, which must also decide the weights in order to reflect physical properties of the configuration of the data set. The first weights have to reflect the closest neighbours and so on.

Once defined the spatial lag operator and the weighting matrices, the space-time autoregressive and moving average model is defined as:

$$\mathbf{X_t} = -\sum_{k=1}^{p}\sum_{l=0}^{\lambda_k}\phi_{kl}\mathbf{W^{(l)}}\mathbf{X_{t-k}} + \sum_{k=1}^{q}\sum_{l=0}^{m_k}\theta_{kl}\mathbf{W^{(l)}}\boldsymbol{\epsilon_{t-k}} + \boldsymbol{\epsilon_t},$$

where:

- $\mathbf{X_t}$ is a matrix or data containing the space-time series: row-wise should be the temporal observations, with each column corresponding to a site;
- p is the autoregressive order;
- q is the moving average order;
- λ_k is the spatial order of the k-th AR term;
- m_k is the spatial order of the k-th MA term;
- ϕ_{kl} is the AR parameter at temporal lag k and spatial lag l;
- θ_{kl} is the MA parameter at temporal lag k and spatial lag l;
- $\mathbf{W^{(l)}}$ is the matrix of weights for spatial order l;
- $\boldsymbol{\epsilon_t}$ is the normally distributed random error at time t with:

$$E[\boldsymbol{\epsilon_t}] = 0,$$

and

$$E[\boldsymbol{\epsilon_t}\boldsymbol{\epsilon'_{t+s}}] = \begin{cases} \sigma^2\mathbf{I_N}, & s = 0 \\ 0, & s \neq 0 \end{cases}.$$

The modelling procedure is described in 3 stages:

- Identification of suitable models to fit the data by analysing the space time autocorrelation function (STACF) and space time partial autocorrelation function (STPACF);
- the parameters estimation of the selected models;
- Diagnosis checking of the selected model.

If the selected model does not meet the requirements in the diagnosis checking, the procedure is repeated until an appropriate model is found.

As mentioned before, the identification stage is performed by the sample estimation of STACF given by

$$\widehat{\boldsymbol{\rho}}_{\boldsymbol{l0}} = \frac{\widehat{\boldsymbol{\gamma}}_{\boldsymbol{l0}}(\boldsymbol{s})}{[\widehat{\boldsymbol{\gamma}}_{\boldsymbol{ll}}(\mathbf{0})\widehat{\boldsymbol{\gamma}}_{\mathbf{00}}(\mathbf{0})]^{\frac{1}{2}}}.$$

where $\widehat{\boldsymbol{\gamma}}_{\boldsymbol{l0}}(\boldsymbol{s})$ is the estimate of the space time autocorrelation function between the l^{th} order beighbor and zero order neighbors at lag s. The STPACF is derived from the space time autoregressive model (STAR) that leads to

$$\boldsymbol{\gamma}_{\boldsymbol{h0}}(\boldsymbol{s}) = -\sum_{j=1}^{k}\sum_{l=0}^{\lambda}\phi_{jl}\boldsymbol{\gamma}_{\boldsymbol{hl}}(s-j),$$

and $s = 1, 2, ..., k$ e $h = 0, 1, ..., \lambda$. These equations are the analogues Yule-Walker equations. The theoretical characteristiscs of the STACF and STPACF are summarised in the table 1.

Process	**STACF**	**STPACF**
STAR(p)	tails off with both space and time	cuts off after p lags in time and λ_p lags in space
STMA(q)	cuts off after q lags in time and m_q lags in space	tails off
STARMA(p,q)	tails off	tails off

Table 1: Theoretical behavior of the STACF and STPACF. Adapted from Rao S. and Antunes C.M. (2003)

The estimation of the parameters is performed by calculating the conditional log likelihood function for the parameters (Φ, Θ, σ^2), given by:

$$L(\Phi, \Theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{TN}{2}} exp\left\{ -\frac{S_*(\Phi, \Theta)}{2\sigma^2} \right\},$$

minimizing the conditional sum of squares function given by:

$$S_*(\Phi, \Theta) = \widehat{\boldsymbol{\epsilon}}' \widehat{\boldsymbol{\epsilon}}.$$

where $\widehat{\epsilon}$ is the vector of estimated errors $N \times 1$.

As a last stage, the main goal is to verify if the model suits the data. "If the selected model represents adequately the data, the residuals should be gaussian (...) One way of testing for correlation is to calculate the sample space time autocorrelations of the residuals and check for additional significant structure" (Rao and Antunes, 2003, pag 132).

3 Results

The data consists of daily mean radioactivity, in Rutherford (Rd), collected from 2008 to 2017, at thirteen sites in Portugal: Beja, Castelo Branco, Coimbra, Elvas, Faro, Fratel, Funchal, Lisboa, Penhas Douradas, Ponta Delgada, Portalegre, Porto and Sines. Given the high number of missing values, some sites were neglected. Only sites with no more than 10% of missing data were included in the study. Therefore, these eight sites will be part of the study:

- Bragança;
- Castelo Branco;
- Coimbra;
- Elvas;
- Faro;
- Ponta Delgada;
- Portalegre;
- Porto.

However, most of the selected sites are not complete, so the KNN Algorithm (K-Nearest Neighbours Algorithm) has been applied to fulfil the data set. In table 2 some descriptive measures were calculated of the original data.

Site	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Bragança	74.1	80.6	83.0	82.9	85.0	100.6
Castelo Branco	151.3	164.6	170.0	170.0	175.5	189.2
Coimbra	109.9	117.4	121.5	121.4	124.9	143.9
Elvas	138.8	145.3	149.8	149.8	154.2	163.4
Faro	74.4	77.9	79.2	79.4	80.8	86.1
Ponta Delgada	126.2	130.5	132.1	132.2	133.8	138.8
Portalegre	146.8	153.6	157.1	157.7	161.7	172.1
Porto	150.3	169.5	177.4	177.6	185.4	223.9

Table 2: Descriptive measures of the eight sites

3.1 Univariate Modelling

In this section, modelling each one of the eight sites separately was considered. All of the sites were analysed by decomposing the observed data in the following way:

$$X_t = T_t + S_t + Z_t$$

where

- X_t is the original data;
- T_t is the tendency component;
- S_t is the seasonal component;
- Z_t is the residual component.

Throughout this thesis, the univariate modelling was applied to the data removing the tendency and the seasonal components. So, following the Box-Jenkins procedure to identify which ARMA model fit the data, the ACF and the PACF are computed. Furthermore, the autocorrelations out of the confident bands were a guess to identify the autoregressive, moving average and the seasonal order of each model. The selected models that suit better the time series observed at each portuguese sites are the following:

- Bragança: *ARMA(1,1)*
- Castelo Branco: *ARMA(2,1)*
- Coimbra: *ARMA(3,1)*
- Elvas: *AR(3)*
- Faro: *AR(1)*
- Ponta Delgada: *ARMA(2,1)*
- Portalegre: *ARMA(2,1)*
- Porto: *ARMA(3,1)*

Site	AIC	BIC	Standard Error	Ljung-Box Test p-value
Bragança	16015.9	16034.5	4.7	0.99
Castelo Branco	17365.8	17396.9	6.8	0.68
Coimbra	17361.2	17398.4	6.8	0.62
Elvas	14424.1	14455.1	3.0	0.92
Faro	10022.9	10035.4	0.9	0.37
Ponta Delgada	8457.5	8488.5	0.6	0.07
Portalegre	14123.3	14154.3	2.8	0.87
Porto	20641.1	20678.3	16.7	0.60

Table 3: AIC, BIC and standard errors of the selected models and the p-values of the Ljung-Box Test for 10 lags.

To ensure the accuracy of the selected models, the residuals were analysed not only by plotting the respective ACF's, but also by applying a Ljung-Box Test, to verify the correlation between the residuals of the model. The summary of each model is presented in the table 3, and it can be drawn that all of the selected models fit correctly the data and the residuals are non correlated according to the Ljung Box Test.

3.2 Space Time Modelling

After modelling the radioactivity time series separately, a space time modelling was performed. As detailed before, first a weighting matrix has to be defined in order to describe spatial distribution of the eight sites. In this study, only spatial orders 0 and 1 were considered: where $W^{(0)}$ is the identity matrix $N \times N$ and $W^{(1)}$ is obtained by the inverse of the euclidean distance between sites.

$$\mathbf{W^{(1)}} = \begin{bmatrix} 0 & 0.06 & 0.07 & 0.09 & 0.16 & 0.48 & 0.09 & 0.05 \\ 0.09 & 0 & 0.03 & 0.04 & 0.12 & 0.62 & 0.02 & 0.07 \\ 0.09 & 0.03 & 0 & 0.07 & 0.14 & 0.59 & 0.05 & 0.04 \\ 0.12 & 0.04 & 0.07 & 0 & 0.08 & 0.59 & 0.02 & 0.10 \\ 0.15 & 0.09 & 0.10 & 0.06 & 0 & 0.42 & 0.07 & 0.12 \\ 0.15 & 0.14 & 0.14 & 0.15 & 0.14 & 0 & 0.14 & 0.14 \\ 0.11 & 0.02 & 0.05 & 0.02 & 0.10 & 0.61 & 0 & 0.09 \\ 0.06 & 0.06 & 0.04 & 0.10 & 0.16 & 0.51 & 0.08 & 0 \end{bmatrix}$$

After defining the matrices $W^{(0)}$ and $W^{(1)}$, the following step is the observation of the sample space-time autocorrelation and sample space-time parcial autocorrelation function (STACF and STPACF) to identify some possible models to describe correctly the data.

The data set was differentiated for a $lag = 365$. The sample STACF and STPACF of the differentiated time series are display in the figures 1 and 2.

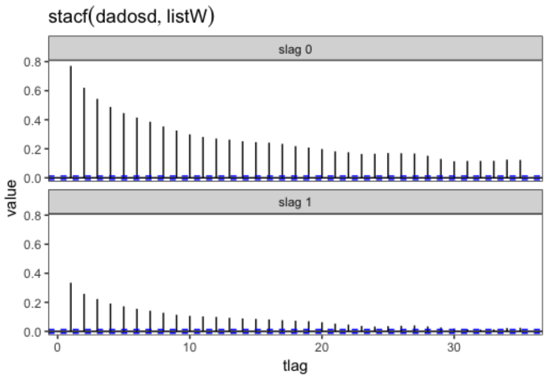

Figure 1: STACF of the differentiated data

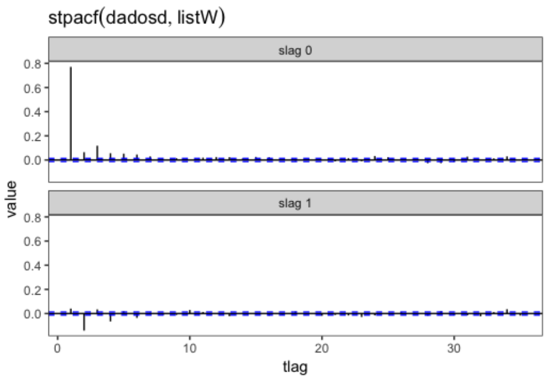

Figure 2: STPACF of the differentiated data

By observing figures 1 and 2, it was possible to identify many candidate models to describe the time series. The models are presented in table 4.

As previously mentioned, the residuals have also been examined through the observation of STACF of the residuals. Several correlations were above the confident bands. In addition, the hypothesis of residuals null correlation was tested. By the Box Pierce Test, is was concluded that the model 6 is the only in which null hypothesis can not be rejected. Nevertheless, additional STARMA models were tested.

More models were selected, and displayed in table 5. Most were more suitable than the first six. The residuals of the models eight, nine and eleven demonstrate null correlation between the residuals, as proven by the Box-Pierce Test.

Once more, plot the STACF of the residuals is part of the model diagnosis. To summarise, only the residuals STACF of model nine and eleven will be shown.

#	Model	BIC	Standard Error	Box Pierce Test p-value
1	$(2,0,0) \times (0,1,1_1)$	130596.5	8.5	0
2	$(3,0,0) \times (0,1,1_1)$	133560.5	11.7	0
3	$(3_1,0,1) \times (0,1,1_1)$	142063.7	25.5	0
4	$(3_1,0,1) \times (1,1,1_1)$	268289.7	4294.4	0
5	$(3_1,0,2) \times (0,1,1_1)$	231168.8	1040.5	0
6	$(3_2,0,2) \times (0,1,1_1)$	130536.5	9.7	0.3

Table 4: AIC, BIC and standard errors of the selected models and the p-values of the Ljung-Box Test for 10 lags. The notation 3_1 in the model 3, for example, represents a max time lag of 3 and include spatial dependence of order 1.

In both models, the correlations can be detected by the residuals STACF. But, it can be observed that the STACF of the residuals of model eleven can catch more correlations than the model nine. So, by this reason the model eleven is chosen as the best model, and it will be used to forecasting.

Although it is extremely hard to catch all the correlations, there are many reasons to explain this issue:

- No spatial correlation between sites;

- covariance matrix different from $\sigma^2\mathbf{I_N}$;
- the model is not suitable.

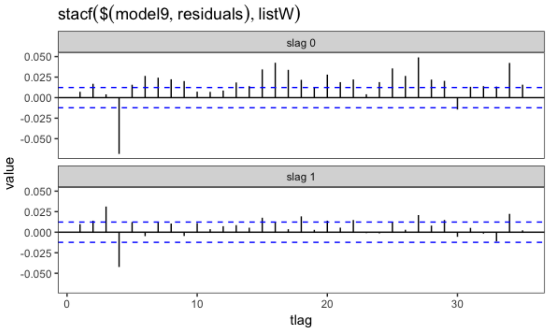

Figure 3: STACF of residuals of model 9

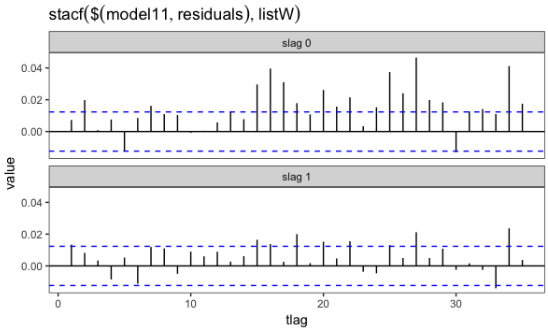

Figure 4: STACF of residuals of model 11

#	Modelo	BIC	Standard Error	Box Pierce Test p-value
7	$(4_1, 0, 2) \times (0, 1, 1_1)$	130109.8	8.9	0.0
8	$(3_1, 0, 3) \times (0, 1, 1_1)$	130254.8	9.2	0.7
9	$(4_1, 0, 3) \times (0, 1, 1_1)$	130069.2	8.7	0.1
10	$(3_1, 0, 4) \times (1, 1, 1_1)$	183024.5	158.9	0.0
11	$(3_1, 0, 4_1) \times (0, 1, 1_1)$	130513.5	9.7	0.1

Table 5: AIC, BIC and standard errors of the selected models and the p-values of the Ljung-Box Test for 10 lags

Site	ARMA	STARMA
Brangança	3.4	3.1
Castelo Branco	5.3	5.2
Coimbra	5.0	8.9
Elvas	6.6	7.5
Faro	1.9	2.5
Ponta Delgada	2.8	3.5
Portalegre	4.5	6.6
Porto	15.0	16.0

Table 6: Standard error of the forescasts by the ARMA and STARMA model.

3.3 Comparing Forecasts

In this study, the two ways of modelling the eight time series, were compared also by the ability to forecast. These forecasts, has been made by removing the last five observations of each time series, then predicting these values and comparing them to the observed data. The comparison between the univariate and the space-time predictions has been made in terms of the Mean Square Error and the standard deviation of the forecast errors.

The forecasts were accurate enough and quite similar using ARIMA and STARMA models. By table 6, it can be proved that ARIMA models are slightly better than the STARMA model. Even so, the difference is not significant.

4 Conclusion

In this thesis ARMA and STARMA models were described as well as their application to model data collection. Eight time series were modelled independently by ARMA models, and also all together by a STARMA model. Suitable models were found in both univariate and space time case that fit accurately the data set.

After selecting the models (both in univariate and STARMA) that best describe the radioactivity in the eight sites, a comparison was made regarding the forecasting of the last five observations. The results were slightly better in the univariate case, considering mean squared errors from the forecasts of the observed data.

References

[1] Andre M., Dabo-Nian S., Soubdhan T. and Ould-Baba H. (2016). Predictive spatio-temporal model for spatially sparse global solar radiation data. *Energy* **111**, 599-608.

[2] Barbosa S. and Huisman J.A. and Brito Azevedo A. (2018). Meteorological and soil surface effects in gamma radiation time series - Implications for assessment of earthquake precursors. *Jounal of Environmental Radioactivity* **195**, 72-78.

[3] Giacomini R. and Granger C. (2004). Aggregation of space-time process. *Journal of Econometrics* **118**, 7-26.

[4] Glasbery C. and Allcroft D. (2007). A spatio temporal auto-regressive moving average model for solar radiation. *Applied Statistics* **57**, 343-355.

[5] Martin R.L. and Oeppen J.E. (1975). The identification of regional forecasting models using space-time correlation functions. *Transactions of the Institute of British Geographers* **66**, 95-118.

[6] Mercieca J. and Kadirkamanathan V. (2016). Estimation and identification of spatio-temporal models with applications in engineering, healthcare and social science. *Annual Reviews in Control* **42**, 285-298.

[7] Pace R., Barry R., Sirmans C. and Gilley O. (2000). A method for spatial-temporal forecasting with application to real estate prices. *International Journal od Forecasting* **16**, 229-246.

[8] Pfeifer P. and Deutsch S. (1980a). A three-stage iterative procedure for space-time modeling. *Technometrics* **22**, 35-47.

[9] Pfeifer P. and Deutsch S. (1980b). Identification and interpretation of first order-space-time ARMA models. *Technometrics* **22**, 397-408.

[10] Pfeifer P. and Deutsch S. (1981a). Variance of the sample space-time autocorrelation function. *Journal of the Royal Statistical Society Series B* **43**, 28-33.

[11] Pfeifer P. and Deutsch S. (1981b). Space-time arma modeling with contemporaneously correlated innovations. *Technometrics* **23**, 401-409.

[12] Pfeifer P. and Deutsch S. (1981c). Seasonal space-time aroma modeling. *Geographical Analysis* **13**, 117-133.

[13] Rao S. and Antunes C.M. (2003). Spatio-temporal modeling of temperature time series - Time Series Analysis and Applications to Geophysical Systems. *Technometrics* **139**, 123-150.

[14] Zhour J. and Mitchell K. (2005). Spatial autoregressive models for resource demand prediction in mobile wireless networks. *Beijing University of Posts and Telecommunications Press*, 1071-1080.