

**A METODOLOGIA DE GEOESTATÍSTICA  
APLICADA NA ANÁLISE DE SÉRIES TEMPORAIS**

**Sueli Aparecida Mingoti**- Departamento de Estatística - UFMG  
**Otaviano Francisco Neves** - CDTN/CNEN  
e-mail: [sueli@est.ufmg.br](mailto:sueli@est.ufmg.br) ; [otaviano@urano.cdtm.br](mailto:otaviano@urano.cdtm.br)

**Resumo**

Neste artigo apresentamos um exemplo de utilização da metodologia de Geoestatística aplicada a análise de séries temporais. Para facilitar o entendimento a série temporal tratada é a do número de manchas solares de Wolfer, série bastante conhecida devido a sua alta variabilidade e dificuldade de predição. O objetivo principal é mostrar que a metodologia de Geoestatística pode resultar em previsões de boa qualidade mesmo no contexto temporal.

---

Palavras Chaves: Geoestatística, Séries Temporais, Sazonalidade.

**Abstrat**

In this article we present an example of the Geostatistics methodology applied to the time series analysis. In order to make it easier the understanding, the time series chosed to be analysed was the Wolfer sunspots number, which is very well known because it has a high variability and it is difficult to predict. The main objective is to show that the Geostatistics methodology can result in predictions of good quality even in the time series context.

---

Key Words: Geostatistics, Time Series, Seasonal series.

**1.0 Introdução**

Uma das áreas mais importantes da Estatística é a que trata da Análise de Séries Temporais. Dentre as várias metodologias existentes encontram-se o método de Box e Jenkins (Box e Jenkins, 1976), que se resume na identificação de um modelo ARIMA para descrever o comportamento da série; a análise espectral (Fuller, 1976) que trata a série no domínio das frequências e a metodologia Bayesiana (Harrison e West, 1987) na qual distribuições à priori são incorporadas para representar o comportamento dos parâmetros do modelo que descreve a série. Entretanto, uma outra forma de se tratar uma série temporal, e que tem sido pouco explorada na literatura, é aquela em que a série é vista dentro do contexto de Geoestatística (Cressie, 1992) metodologia esta que utiliza a informação do Semi-Variograma ou outras medidas alternativas

como Madograma ou Rodograma, como forma principal de identificação do modelo e previsão de valores. O objetivo deste artigo é mostrar como a metodologia de Geoestatística pode ser aplicada no caso de Análise de Séries Temporais. Escolhemos como exemplo de análise a série do número de manchas solares de Wolfer no período de 1770 a 1869, por ser esta uma série bastante conhecida nesta área, devido a sua alta variabilidade e dificuldade de predição (ver gráfico 1). Denotamos por  $Z_t$  o valor do número de manchas solares de Wolfer no ano  $t$ ;  $t=1,2,3,\dots,100$ , onde  $t=1$  corresponde ao ano de 1770,  $t=2$  ao ano de 1771, até  $t=100$  que corresponde ao ano de 1869.

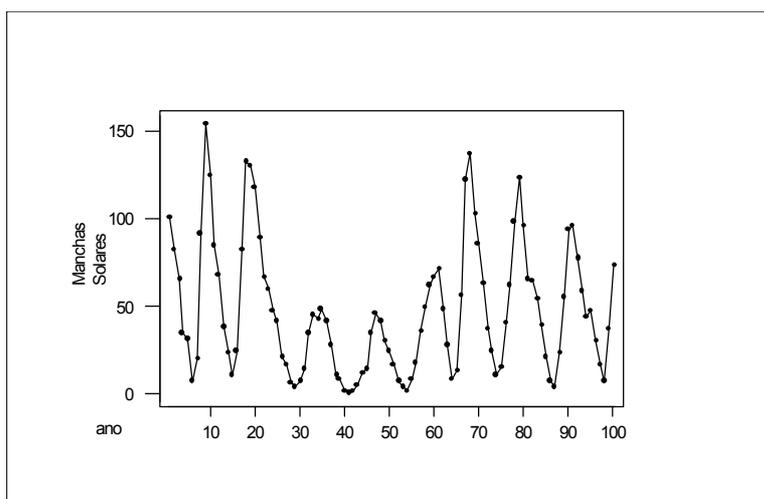


Gráfico1:Série de manchas solares de Wolfer período:1770 a 1869.

## 2.0 Metodologia de Geoestatística em Séries Temporais

### 2.1 Medidas que Descrevem a Variabilidade Espacial dos Dados

Para se proceder a análise dos dados através de métodos de Geoestatística é necessário a imposição de que o processo estocástico gerador dos dados amostrais seja intrinsecamente estacionário e isotrópico ou seja, deve satisfazer as seguintes condições:

- i)  $E[Z(t)] = \mu, \forall t \in T$ ;
- ii)  $Var[Z(t)-Z(s)] = 2 \gamma(|t-s|), \forall t \neq s \in T$ ,

isto é, o processo deve ter média constante ao longo do tempo e para cada  $t \neq s$  a variância das diferenças  $[Z(t)-Z(s)]$  deve ser uma função que depende apenas da distância no tempo das observações  $Z(t)$  e  $Z(s)$ , isto é  $|t-s|$ . A quantidade  $2\gamma(\bullet)$  é chamada de Variograma do processo estocástico gerador da série e  $\gamma(\bullet)$  é chamado de Semi-Variograma. O conhecimento da função de Variograma tem importância fundamental na

construção do modelo de projeção, uma vez que as equações de previsão são todas funções dos valores do Variograma do processo estocástico gerador dos dados (Mingoti e Silva, 1995). Embora o Variograma seja a função mais conhecida para descrever a variabilidade espacial dos dados existem outras medidas que, em alguns casos, podem resultar em projeções de melhor qualidade (Mingoti, 1996). As mais conhecidas são o Madograma e o Rodograma, que para um processo intrinsecamente estacionário e isotrópico são definidas respectivamente por:

$$2\Gamma_w(t-s) = E[ | Z(t) - Z(s) |^w ], t \neq s, \in T$$

Para  $w=1$  tem-se o Madograma, para  $w=1/2$  o Rodograma e  $w=2$  representa o Variograma. Estas medidas alternativas são mais robustas no que se refere a influência dos valores discrepantes e do tamanho da amostra na identificação de modelos de projeção espacial (Journel, 1988). Alguns exemplos mais comuns de modelos de Semi-Variograma, Madograma e Rodograma são o modelo linear e o esférico. Na prática, estas funções de variabilidade espacial são estimadas a partir dos dados amostrais sendo os estimadores clássicos não viciados usuais, para o caso de séries temporais, definidos por:

$$2\hat{\Gamma}_w(s) = \frac{\sum_{t=1}^{n-s} [ | Z(t) - Z(t-s) | ]^w}{n-s}$$

onde para  $w=1$  tem-se o estimador do Madograma, para  $w=1/2$  o Rodograma e para  $w=2$  o Variograma. Em geral as estimativas assim obtidas são chamadas de Madograma, Rodograma e Variograma Experimentais. Através destes estimadores identifica-se o modelo matemático de Variograma, Madograma e Rodograma do processo estocástico gerador dos dados amostrais. Prossegue-se então, com a estimação dos parâmetros do mesmo. O modelo estimado é utilizado nas equações de projeção. No caso específico de séries temporais, os estágios de identificação e estimação de parâmetros dos modelos de Variograma, Madograma e Rodograma, podem ser suprimidos na grande maioria dos casos de interesse prático. Isto decorre do fato de que os Variogramas, Madogramas e Rodogramas experimentais contêm toda informação necessária para a construção das equações de predição, como será mostrado ao longo deste artigo.

## **2.2 Identificação e Ajuste dos Modelos para a Série de Números de Manchas Solares de Wolfer**

Os Semi-Variograma, Madograma, e Rodograma experimentais da série de manchas solares de Wolfer estão apresentados nos gráficos 2 a 4. Estes gráficos indicam a existência de uma

sazonalidade de ordem de aproximadamente 10,5 na série  $Z_t$ , já que existe uma sequência de picos inferiores iniciando-se na defasagem próxima  $h=10$  e repetindo-se de 10 em 10 unidades. Além disso, percebe-se nitidamente que o modelo senoide completo descrito na equação (1) poderia ser ajustado para os 3 casos. Duas outras alternativas a considerar seriam os dois parcionamentos do modelo senoide completo, na defasagem  $h=5$ , conforme é apresentado nas equações (2) e (3).

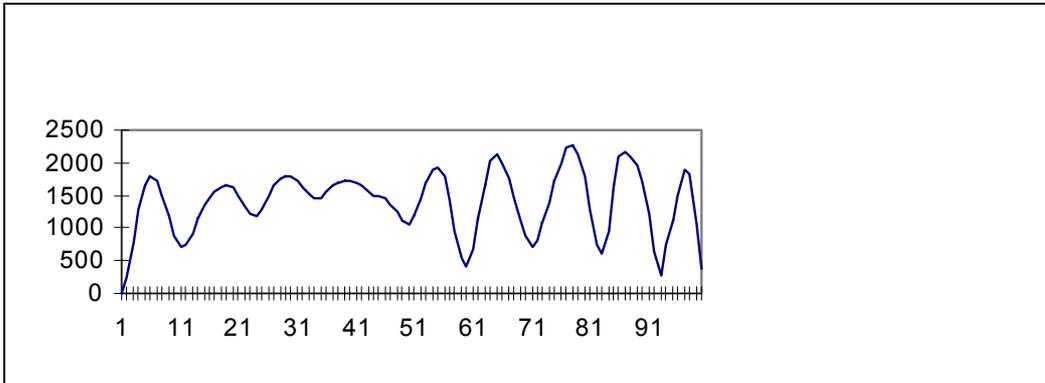


Gráfico 2: Semi-Variograma experimental da série manchas solares de Wolfer

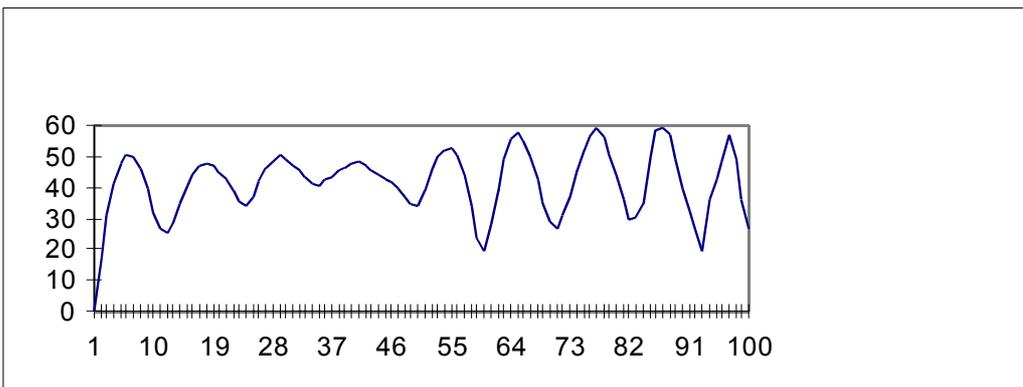


Gráfico 3. Madograma experimental da série manchas solares de Wolfer

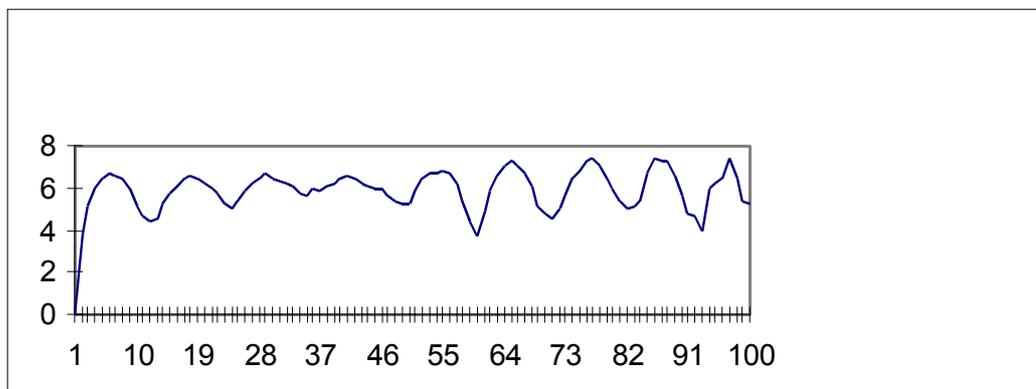


Gráfico 4: Rodograma experimental da série manchas solares de Wolfer

$$\gamma(h, \tilde{\theta}) = \begin{cases} 0 & , \quad h = 0 \\ c_0 + c_w \{1 - a_w \sin(h / a_w) / h\} & , \quad h \neq 0 \end{cases} \quad (1)$$

onde  $\tilde{\theta} = (c_0, c_w, a_w)$  ,  $c_0 \geq 0$  ,  $c_w \geq 0$  ,  $a_w \geq 0$  são os parâmetros do modelo.

$$\gamma(h, \tilde{\theta}) = \begin{cases} 0 & , \quad h = 0 \\ c_0 + c_w \{1 - a_w \sin(h / a_w) / h\} & , \quad 0 \leq h \leq 5 \\ c_0 + c_{1w} \{1 - b_w \sin(h / b_w) / h\} & , \quad h > 5 \end{cases} \quad (2)$$

$$\tilde{\theta} = (c_0, c_w, a_w, c_{1w}, b_w) \quad c_0 \geq 0, \quad c_w \geq 0, \quad a_w \geq 0, \quad c_{1w} \geq 0, \quad b_w \geq 0.$$

$$\gamma(h, \tilde{\theta}) = \begin{cases} 0 & , \quad h = 0 \\ c_0 + c_w \{1 - a_w \sin(h / a_w) / h\} & , \quad 0 \leq h \leq 5 \\ c_0 + c_{1w} \{1 + b_w \sin(h / b_w) / h\} & , \quad h > 5 \end{cases} \quad (3)$$

$$\tilde{\theta} = (c_0, c_w, a_w, c_{1w}, b_w) \quad c_0 \geq 0, \quad c_w \geq 0, \quad a_w \geq 0, \quad c_{1w} \geq 0, \quad b_w \geq 0.$$

Os parâmetros dos modelos foram estimados pelo método de mínimos quadrados ponderados (Cressie, 1992). Para efeito de ajuste dos parâmetros utilizou-se apenas as informações do Semi-Variograma, Madograma e Rodograma experimentais, relativas aos lags  $h=1, 2, 3, \dots, 50$ , uma vez que a alta variabilidade dos dados observados a partir de  $h = 50$  comprometem a qualidade de estimação dos parâmetros do modelo. De um modo geral, os modelos senoide parcionados apresentaram melhores ajustes que os modelos senoide completo, sendo que o modelo senoide parcionado positivo foi melhor para o ajuste de Semi-Variograma, enquanto que os parcionados negativos foram melhores para Rodograma e Madograma. O Rodograma foi o que apresentou o melhor ajuste para os parâmetros. A qualidade de ajuste foi medida em termos da avaliação da Soma dos quadrados dos erros para cada modelo (ver tabela 1).

**Tabela 1: Soma de quadrados dos erros para os modelos ajustados**

Modelos ajustados	Semi-Variograma SQE	Madograma SQE	Rodograma SQE
senoide completo	130,78	79,14	34,38
senoide negativo	119,47	64,81	29,16
senoide positivo	97,88	77,76	33,80

### 2.3 Predição de Valores: Kriging Ordinário e Intervalos de Confiança

As equações de previsão para um ano "s" qualquer não pertencente à série observada, são obtidas utilizando-se a técnica de predição espacial denominada Kriging Ordinário (Cressie, 1992). Neste caso, o preditor linear não viciado de  $Z(s)$  é dado por:

$$\hat{Z}(s) = \sum_{i=1}^n \lambda_i Z(t_i) \quad s > t_n \quad (\text{onde } \sum_{i=1}^n \lambda_i = 1)$$

e  $Z(t_1), Z(t_2), \dots, Z(t_n)$  representam respectivamente os valores observados da série. Para o caso de séries temporais, quando a medida de variabilidade espacial é o Semi-Variograma tem-se que os coeficientes  $\lambda_i$  são obtidos através da seguinte equação:

$$\lambda_0 = \Gamma_0^{-1} \gamma_0 \quad (4)$$

onde  $\gamma_0 = (\gamma(s-t_1), \gamma(s-t_2), \dots, \gamma(s-t_n), 1)'$  (5)

e  $\Gamma_0$  é a matriz  $(n+1) \times (n+1)$  dada por:

$$\Gamma_0 = \begin{cases} \gamma(t_i - t_j), & i=1, \dots, n, j=1, \dots, n, \\ 1 & , i=n+1, j=1, \dots, n, \\ 0 & , i=n+1, j=n+1, \\ 1 & , i=1, \dots, n, j=n+1 \end{cases} \quad (6)$$

onde  $\gamma(\bullet)$  representa o valor do Semi-Variograma com os parâmetros já estimados. Neste caso, o estimador linear  $\hat{Z}(s)$  é o que minimiza a variância do erro de predição. Quando o Madograma e o Rodograma são as medidas de variabilidade espacial utilizadas, os coeficientes  $\lambda_i$  são obtidos das equações (4) a (6), substituindo-se no lugar de  $\gamma(\bullet)$  o valor do Madograma, ou do Rodograma correspondente. Neste caso o preditor linear obtido não mais minimiza a variância do erro de predição, mas sim a norma ao quadrado do erro de predição. Em todos os casos, o erro de predição e a respectiva variância são definidos por:

$$erro = (Z(t) - \hat{Z}(t))$$

$$Var(erro) = \hat{\sigma}^2(s) = [2 \sum_{i=1}^n \lambda_i \gamma(s-t_i)] - [\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(t_i - t_j)]$$

sendo  $\lambda_i$  os valores obtidos da equação(4) usando-se respectivamente o Semi-Variograma, Madograma e o Rodograma estimados. Os intervalos de confiança de 95% para  $Z(s)$  são dados aproximadamente por:  $(\hat{Z}(s) \pm 1,96 \hat{\sigma}(s))$ , supondo-se normalidade dos dados. Embora teoricamente a predição de  $Z(s)$  seja efetuada com base na informação da vizinhança  $t_1, t_2, \dots, t_n$ , ou seja com base em todas as observações coletadas da série, na prática isto não ocorre. As predições de valores é efetuada de acordo com a informação de uma vizinhança menor, representada por valores da série que estão mais próximos do tempo "s". Um software como o GEO-EAS, por exemplo, permite que o pesquisador escolha a vizinhança de interesse para efeito de projeção (Englund e Sparts, 1991). No caso específico de séries temporais os valores destas medidas de variabilidade espacial (temporal) podem ser obtidas diretamente através do cálculo do Semi-Variograma, Madograma e Rodograma Experimentais, uma vez que na maioria dos casos estes podem ser calculados para todos os valores de  $h$  necessários para a projeção de valores à curto e médio prazo. Deste modo, uma forma alternativa de se obter projeções para valores futuros da série, é utilizando-se as medidas de variabilidade experimentais nas equações(3) a (6) de Kriging Ordinário. Portanto, predições de valores futuros da série sempre são possíveis de serem obtidas com o uso da metodologia de Geoestatística, independentemente do bom reconhecimento e do ajuste dos parâmetros do modelo teórico de Semi-Variograma, Madograma ou Rodograma do processo estocástico gerador dos dados amostrais, o que caracteriza um ponto positivo em favor desta metodologia uma vez que no caso de séries temporais, os modelos de Variograma podem ser mais complicados de serem identificados e ajustados. A título de exemplo, o gráfico 5 apresenta as previsões para a série  $Z_t$ , no período de 1860 a 1869, tendo-se como origem fixa de previsão o ano de 1859. As previsões foram geradas de acordo com o modelo Madograma senoide paracionado positivo e com vizinhança 11, ou seja, a previsão do ano de 1860, por exemplo, foi efetuada utilizando a informação da série de 1849 a 1859. Os respectivos intervalos de 95% de confiança para  $Z_t$  e os correspondentes valores reais observados da série estão apresentados no gráfico. Apesar da grande variabilidade dos dados e da amplitude dos intervalos de confiança, apenas três observações ficaram fora dos limites (anos de 1866, 1867 e 1869). As demais observações se mostraram bem próximas do valor real. Os limites dos intervalos de confiança são consistentes com a definição da variável  $Z_t$ .

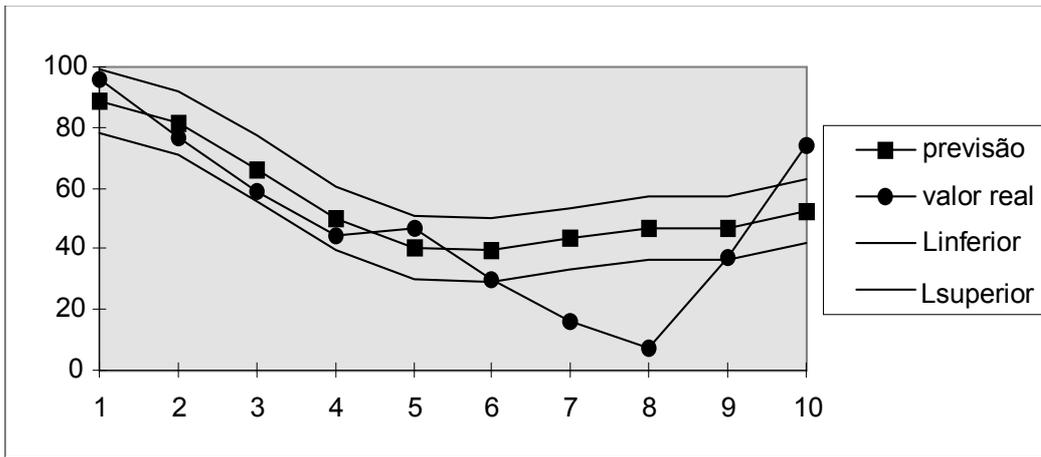


Gráfico 5: Previsões a dez passos para  $Z_t$  com origem de previsão no ano de 1859

## 2.4 Análise das Previsões

Uma comparação mais detalhada dos modelos ajustados foi efetuada considerando-se as previsões obtidas a partir de origens fixas e a partir de origens móveis. Numa primeira análise calculou-se as predições a um passo para o período de 1781 a 1869. A escolha do ano inicial de 1781 foi feita para que os modelos com vizinhança 11 (vizinhança mais próxima da sazonalidade 10,5 observada na série) fossem avaliados. Numa segunda análise a origem de previsão foi fixada em 1864 e previsões foram geradas para o período de 1865 a 1869. Numa análise complementar a qualidade das predições foi avaliada de acordo com o comportamento das mesmas para origens de previsão diferentes. A origem de previsão "t" variou de 1849 à 1865 para as previsões a 5 passos à frente, e de 1849 à 1867 para previsões a 2 passos. Para comparação dos modelos utilizamos o Erro Médio (EM) e o Erro Quadrático Médio (EQM) definidos por:

$$EM = \frac{1}{m} \left[ \sum_s (Z(s) - \hat{Z}(s)) \right] \quad EQM = \frac{1}{m} \left[ \sum_s [Z(s) - \hat{Z}(s)]^2 \right]$$

onde, m é número de predições que foram feitas,  $Z(s)$  representa o valor real da série no ano "s", e  $\hat{Z}(s)$  representa a predição feita pelo modelo para o valor de  $Z(s)$  quando se tem como origem de previsão um tempo t, fixado à priori. Os resultados obtidos estão na tabela 1. Para as previsões com origem móvel os valores EM e EQM apresentados nas tabela 2 representam a média aritmética dos valores de EM e EQM obtidos para cada origem de previsão "t" considerada. Embora existam vários modelos bem similares em termos de EM e EQM, pode ser observado que para previsões a um passo os melhores desempenhos foram do Semi-Variograma (vizinhança 2 e 4) e o Semi-Variograma teórico senoide completo (vizinhança 4). Os piores desempenhos ocorreram para o Rodograma teórico

completo e o positivo. Para previsões a 5 passos com origem de previsão fixa, o Madograma (11) e Rodograma (11) experimentais foram os melhores. Para previsões a 2 e 5 passos com origem de previsão móvel, os melhores modelos foram o Madograma (11) e Rodograma (11) experimentais. É interessante observar os altos valores de EM e EQM quando a previsão é feita para 5 passos a frente (com origem fixa e móvel) nos modelos ajustados com vizinhança 2. Aparentemente modelos com vizinhança 2 não seriam as melhores escolhas para previsões a longo prazo.

Tabela 2: Análise de ajuste das previsões dos modelos de geoestatística ajustados

Modelos	Previsões a um passo 1781-1869		Previsões a 5 passos origem 1864		Previsões a 5 passos origem móvel t=1849 a 1864		Previsões a 2 passos origem móvel t=1849 a 1867	
	EQM	EM	EQM	EM	EQM	EM	EQM	EM
S.V.Exp.(4)	14,43	-0,12	37,79	-28,35	38,68	-8,80	20,64	-0,32
S.V.Exp.(2)	16,11	0,42	28,49	-17,01	39,64	-4,52	22,80	1,49
M.Exp.(11)	19,19	-1,11	13,44	-5,82	19,00	-7,30	15,20	-4,22
M.Exp.(4)	22,03	-0,57	29,81	-18,35	44,43	-11,35	28,03	-5,29
M.Exp.(2)	21,49	-0,22	26,98	-13,93	43,88	-9,78	27,19	-3,64
R.Exp.(11)	23,29	-1,60	14,88	-7,44	21,37	-8,15	17,06	-5,69
R.Exp.(4)	27,81	-1,08	29,56	-18,25	46,65	-12,54	32,31	-7,46
R.Exp.(2)	24,77	-0,44	26,73	-13,43	45,13	-10,64	29,13	-4,80
S.V.T.C.(4)	16,00	0,12	36,71	-27,85	35,35	-8,12	22,37	-0,77
M.T.C.(11)	19,99	-0,75	29,57	-21,65	27,84	-8,49	22,45	-5,13
M.T.C.(4)	24,13	-0,54	27,85	-15,46	42,14	-10,37	47,80	44,11
M.T.C.(2)	23,97	-0,39	26,78	-13,53	42,06	-9,81	28,69	-4,56
M.T.P.(11)	20,39	-0,68	28,23	-16,24	23,49	-8,52	20,35	-4,49
M.T.N.(11)	19,60	-0,72	22,33	-10,35	21,26	-8,06	33,05	30,75
R.T.C.(11)	24,02	-1,48	24,10	-15,96	27,15	-9,03	21,76	-6,85
R.T.C.(4)	30,19	-1,20	28,80	-17,20	44,25	-11,88	33,44	-7,76
R.T.C.(2)	26,45	-0,55	26,64	-13,24	42,79	-10,27	30,00	-5,23
R.T.P.(11)	32,04	12,38	20,78	2,28	27,23	10,60	25,40	9,70
R.T.N.(11)	24,06	-1,44	21,72	-11,74	24,86	-9,03	20,77	-6,52

Nota: S. V. = Semi-Variograma ; R=Rodograma; M=Madograma; E=Experimental;  
T. C. = Teórico Completo; T.N.= teórico negativo; T.P. = teórico positivo.

#### 4.0 Considerações Finais

A análise das previsões efetuadas na seção 2.4 indicou que os modelos de projeção construídos a partir da metodologia de Geoestatística tiveram bom desempenho no caso da previsão da série do número de manchas solares de Wolfer. O uso desta metodologia permitiu a identificação de uma sazonalidade de ordem de aproximadamente 10,5 na série. A facilidade no entendimento das funções de Semi-Variograma, Madograma e Rodograma, faz com que esta metodologia de Geoestatística possa ser uma grande competidora com a

metodologia de análise espectral, quando esta é utilizada para detecção de possíveis sazonalidades na série. Os conceitos envolvidos com a construção do Variograma, Madograma e Rodograma, são mais simples de serem entendidos pelo usuário comum, do que aqueles relacionados com o Peridiograma. Em relação a outras metodologias para análise de séries temporais, a Geoestatística teria uma vantagem, devido ao fato de poder tratar de uma forma simples séries não igualmente espaçadas no tempo, uma vez que sua implementação não depende da igualdade de espaçamento entre as observações da série. Pode-se observar também que as previsões obtidas a partir da inclusão da informação pura e simples de Semi-Variograma, Madograma e Rodograma Experimentais nas equações de previsão, resultaram numa qualidade muitas vezes superior àquela obtida através da identificação e estimação dos parâmetros teóricos destas medidas de variabilidade espacial dos dados.

Deve ser notado que apesar dos modelos senoide positivo e negativo terem apresentado um melhor ajuste de parâmetros que o modelo senoide completo, nem sempre estes modelos resultaram em previsões de melhor qualidade. Finalmente, é importante destacar que para uma avaliação mais precisa da qualidade dos modelos de Geoestatística aplicados a Análise de Séries Temporais, outros trabalhos semelhantes à este deverão ser conduzidos utilizando-se séries temporais reais ou simuladas.

### **Referências Bibliográficas**

- BOX, G.P. e JENKINS, G.M. *Time series analysis: forecasting and control*. New York: Holden Day, 1976, 575 pp.
- CRESSIE, N. *Statistics for spatial data*. New York: John Wiley & Sons, 1992, 900 pp.
- ENGLUND, E. e SPARTS, A. *GEO-EAS User's Guide*. Las Vegas: Environmental Monitoring Systems Laboratory, 1991, 120 pp.
- FULLER, W. A. *Introduction to statistical time series*. New York: John Wiley & Sons, 1976, 470 pp.
- HARRISON, P.J. & WEST, M. Practical bayesian forecasting. *The Statistician*, 36, p. 115-125, 1987.
- JOURNEL, A.G. New distance measures: the route forward truly non-gaussian geostatistics. *Mathematical Geology*, vol. 20, 4, p. 459-475, 1988.
- MINGOTI, S.A. e SILVA, A.F. Um exemplo de aplicação das técnicas de geoestatística: uma análise do nível piezométrico do aquífero de Wolfcamp (Estados Unidos). *Revista Escola de Minas (REM)*, Ouro Preto, Vol 49, 2, p. 96-100, 1995.
- MINGOTI, S.A. As funções de madograma e rodograma como alternativas para descrever a variabilidade espacial dos

dados. *Revista Escola de Minas (REM)*, Ouro Preto. vol 50, 2,  
p.71-74, 1996.

Endereço dos Autores:

Universidade Federal de Minas Gerais

Instituto de Ciências Exatas - Departamento de Estatística

Av. Antonio Carlos, 6627 - Pampulha

30161-970 - Belo Horizonte - Minas Gerais