

ESTUDIO DE EMISIVIDAD: PREDICCIÓN ESTADÍSTICA

E. Frigerio¹, O. Avila Blas²

Av. Bolivia 5150 - 4400 Salta

Tel. : (0387) 4255423

Email : frigerio@unsa.edu.ar, oblas@unsa.edu.ar

RESUMEN: Se busca establecer un modelo de la emisividad de cielo para la zona, realizando un estudio estadístico de las características de la radiación nocturna. Para ello se comenzó a usar un tratamiento de análisis estadístico de series de tiempo con modelos estructurales, que resulta ser al presente una técnica novedosa. Se buscó mejorar el modelado de una serie de emisividad nocturna formada por promedios mensuales mediante el tratamiento de series irregulares, lo que lleva a trabajar con modelos no lineales y no gaussianos. Se obtuvo un modelo estructural que permite ajustar los datos observados y realizar pronósticos con una confiabilidad del 95%.

Palabras clave: Radiación, Atmósfera, Emisividad, Estadística, Análisis Estructural.

INTRODUCCIÓN.

El establecer un modelo estadístico apropiado de la emisividad para la región de Salta es de importancia para poder predecir los valores esperados en el corto plazo.

Para poder realizar este estudio se han usado los datos de la estación Salta proporcionados por el Servicio Meteorológico Nacional que abarcan el período junio/1965-diciembre/1984. Al promedio mensual de éstos se les aplicó una correlación que da la emisividad en función de la temperatura de rocío obtenida de ajustar medidas de emisividad en Salta (Frigerio y Saravia, 1991). La serie de datos obtenida sirvió de punto de partida para el análisis.

En un trabajo previo (Frigerio y Avila Blas, 1999), se hizo el modelado de una serie más restringida de datos, abarcando desde julio/1971 hasta diciembre/1984. El modelo óptimo obtenido en esa oportunidad fue

$$y_t = -4,1e-4 + 6,93e-3*y_{t-1} - 3,47e-3*\gamma_{11} - 9,63e-4*\gamma_{12} + 7,96e-5*x_t + 0; \quad \epsilon_t = 7,96e-5*x_t \sim N(0; 2,02e-4). \quad (1)$$

Este modelo presenta la desventaja de mostrar un corrimiento de la serie pronosticada respecto de la serie original. En este sentido, el modelo encontrado en el presente trabajo no presenta tal dificultad, aumentando de esta manera la confiabilidad en las predicciones.

ANÁLISIS ESTADÍSTICO USADO.

La serie bajo estudio es de tipo irregular debido a que hay datos faltantes, como se muestra en la figura 1.

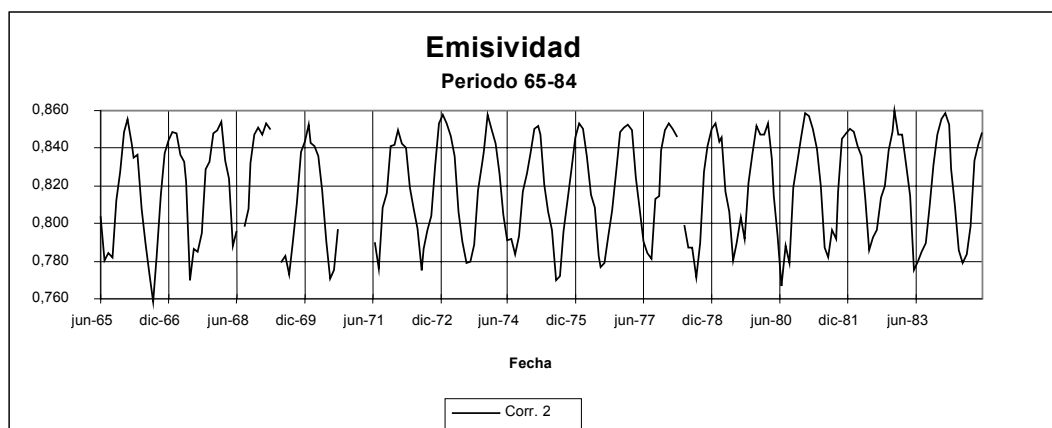


Figura 1: serie de emisividad calculada sobre los datos de la estación Salta del SMN.

¹ INENCO, Instituto UNSa - CONICET

² Cátedra de Probabilidades y Estadística, Fac. Cs. Exactas - UNSa

Para poder tratar esta serie de datos, se han rellenado los datos faltantes con los promedios obtenidos con la media entre los valores de los años anterior y posterior para cada mes en cuestión. La justificación de este procedimiento radica en la regularidad notada en la serie.

Además la distribución empírica de ellos tiene “colas pesadas” tal como se puede observar en la figura 2. La herramienta básica empleada para el tratamiento de este tipo de series consiste en modelar las densidades de los errores de las observaciones y del vector de estado mediante una mezcla de densidades, uno de cuyos componentes es una densidad normal con varianza grande (Durbin y Cordero, 1994). La otra componente puede poseer una distribución distinta a una normal (p. ej. una t de Student), pero la idea subyacente consiste en estimar al vector de estado mediante los procesos de linealización, iteraciones y el suavizador y filtro de Kalman.

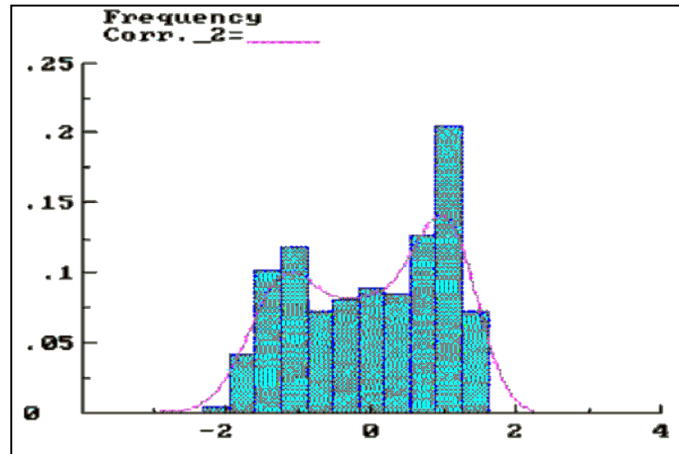


Figura 2: distribución de la serie de datos.

Considerando el modelo de nivel local:

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t \\ \alpha_t &= \alpha_{t-1} + \eta_t, \quad t=1, \dots, n. \end{aligned} \quad (2)$$

donde se relaciona al vector de disturbios $\{\varepsilon_t\}$ con el vector de observaciones $\{y_t\}$ a través de un proceso de Markov $\{\alpha_t\}$ y donde α_t es el vector de estado. Los disturbios son ruido blanco mutuamente no correlacionados con media cero y varianza constante.

El primer objetivo buscado fue construir un modelo para el vector de disturbios de modo que se pudiera suavizar las colas pesadas. Para ello se supuso que la distribución de los disturbios es una mezcla de dos distribuciones normales de la siguiente manera:

$$h(\varepsilon_t) = (1-\beta) \cdot h_1(\varepsilon_t) + \beta \cdot h_2(\varepsilon_t), \quad (3)$$

con $h_1 = N(0, \sigma^2)$ y $h_2 = N(0, \lambda^2 \sigma^2)$ y donde β es pequeño y λ es grande (p. ej., valores aconsejados son $\beta = 0,01$ y $\lambda = 10$). En este modelo β puede considerarse como una medida de la ocurrencia de datos extremos, y λ puede pensarse como un indicador de la magnitud de la desviación de un dato extremo de la media relativa correspondiente a una observación ordinaria.

Para realizar la linealización del modelo, se tomó como valor experimental del vector de estado, $\alpha_t^* = 5,515E-3$, el que se obtuvo estimando el valor de y_t correspondiente a enero de 1966, con el modelo óptimo anterior. Partiendo de este valor experimental inicial, se formó una nueva serie de disturbios

$$\varepsilon_t^* = y_t - \alpha_t^*. \quad (4)$$

Con esta serie, se estimó la nueva varianza de los disturbios por medio de la siguiente ecuación (Abril, 1999)

$$\frac{1}{\sigma_{\varepsilon^*}^2} = \frac{1}{h(\varepsilon_t^*)} \left[\frac{(1-\beta)}{\sigma^2} h_1(\varepsilon_t^*) + \frac{\beta}{\lambda^2 \sigma^2} h_2(\varepsilon_t^*) \right] \quad (5)$$

El valor estimado de la misma fue $\sigma_{\varepsilon^*}^2 = 4,0804.E-4$. Éste permite la resolución de las ecuaciones que intervienen en la optimización del proceso de suavizado y filtrado de Kalman, empleando el software estadístico STAMP 5.0. Tomando como nuevos valores del vector de estado a las soluciones obtenidas, se repite el proceso hasta alcanzar una convergencia adecuada.

Se han intentado varios modelos estructurales a fin de modelar la serie mencionada. En todos ellos se tomaron los valores $\beta = 0,01$ y $\lambda = 10$, para la construcción del valor h y, además, según lo observado en la gráfica de la serie, se ha tomado la componente de nivel fija, es decir no aleatoria. Esos modelos contemplan la descomposición de las series generadas por el proceso de optimización, como una combinación lineal de componentes de diversa índole. Se indica a continuación un resumen de los mismos, indicando el por qué no fueron aceptados como apropiados desde el punto de vista de la bondad de ajuste :

- 1) se trató primeramente la serie de los disturbios sin estacionalidad y con una diferenciación de segundo orden, escrita como: nivel fijo + AR(1)+ 1 ciclo + irregular. Se obtuvo un coeficiente de correlación moderado (0,49) y la prueba de bondad de ajuste de Chocran arroja un índice suficientemente bueno (140,1) como para aceptar el modelo propuesto;
- 2) se trabajó con la serie original de disturbios descomponiéndola como la suma: nivel fijo+ AR(1)+ 1 ciclo + irregular. Se logró una convergencia muy fuerte en el proceso de estimación de hiperparámetros en 13 iteraciones, el coeficiente de determinación aumentó a 0,66, pero en el proceso de estimación se observa un desplazamiento con respecto a la serie original, lo que también se observó en el modelo 1;
- 3) se realizó un estudio similar a los ítems anteriores trabajando con la segunda serie de disturbios obtenida con y_t fijo y haciendo variar los valores del vector de estado. En el primer caso, el coeficiente de determinación bajó a 0,28 y en el segundo caso, el índice de Chocran es muy bajo (4,26) lo que no permite aceptar a los modelos como apropiados;
- 4) cuando se trató la segunda serie de disturbios con el modelo: nivel fijo+AR(1)+1 ciclo + estacionalidad fija + irregular, se consiguió una convergencia muy fuerte en 19 iteraciones, comenzando el proceso de estimación de parámetros en julio de 1965. Cuando se comenzó en junio de 1965, el proceso de estimación de parámetros daba una convergencia muy débil. Este comportamiento se debió a que el valor de junio de 1965 es un valor atípico al comienzo de la serie. Con estas consideraciones, este último modelo es considerado el óptimo.

Los resultados obtenidos para este último modelo en el resumen de estadísticos, estimación de las desviaciones estándar de los disturbios, estimación de los coeficientes del vector de estado final y para la bondad de ajuste, se muestran en la tabla, junto con las predicciones de los errores de estimación y pronóstico de valores a partir de la serie modelada para enero a diciembre de 1984.

Corr. 2 = Level + AR(1) + 1 Cycle(s) + Fixed seasonal + Irregular = $\mu + \phi_1 * y_{t-1} + (k_1 * \psi_{11} + k_2 * \psi_{12}) + \gamma_t + \epsilon_t$ (Ec. 6)			
Estimation report		Ecuación 6: Diagnostic summary report.	
Estimation report Model with 6 hyperparameters (2 restrictions). Parameter estimation sample is 1965. 7 - 1983.12. (T = 222). Log-likelihood kernel is 4.590124. Very strong convergence in 19 iterations. (likelihood cvg 1.528632e-014 gradient cvg 6.883383e-010 parameter cvg 8.139981e-008)		Estimation sample is 1965. 7 - 1983.12. (T = 222, n = 221). Log-Likelihood is 1019.01 (-2 LogL = -2038.02). Prediction error variance is 4.77225e-005 Summary statistics Corr_2 Std.Error 0.0069081 Normality 1.215 H(73) 0.86326 r(1) -0.015989 r(13) 0.027519 DW 2.010 Q(13, 8) 2.789 Rs ² 0.41579	
Ecuación 6 : Estimated standard deviations of disturbances.		Ecuación 6 : Estimated parameters of Cyl.	
Component Corr_2 (q-ratio) Irr 0.00596873 (1.0000) Cyl 0.00000 (0.0000) Ar1 0.00335871 (0.5627)		The cycle variance is 0. The rho coefficient is 0.8524262. The cycle period is 2.11636 (0.1763632 'years'). The frequency is 2.96887.	
Ecuación 6 : Estimated coefficients of final state vector.		Ecuación 6 : Estimated autoregressive coefficient.	
Variable	Coefficient	R.m.s.e.	t-value
Lvl	0.0208759	4.6192e-006	4519.4 [0.0000] **
Cyl_1	0.00000	0.00000	
Cyl_2	0.00000	0.00000	
Ar1	5.1721e-006	2.2763e-005	
		The AR(1) rho coefficient is 0.5786485.	
		Ecuación 6 : Cycle analysis for Cyl.	
		Failure Chi ² (12) test is 8.88533 [0.7127]	
		Cusum t(221) test is -1.42986 [1.8458].	

Goodness-of-fit results for eps2,t		Coefficient of determination	R ²	0.934658
Prediction error variance (p.e.v)	0.000000	... based on differences	RD ²	0.824995
Prediction error mean deviation (m.d)	0.000000	... based on diff around seas mean	RS ²	0.415792
Ratio p.e.v. / m.d in squares (ñ 1.0)	1.102804	Information criterion of Akaike	AIC	-19.740745
		... of Schwartz (Bayes)	BIC	-19.480180
Ecuación 6 : Pronóstico para Corr._2.		Corr._2: Serie original		
Period	Forecast	R.m.s.e.	- Rmse	+ Rmse
1984. 1	0.852051	0.00711546	0.844936	0.859167
1984. 2	0.848396	0.00722244	0.841173	0.855618
1984. 3	0.843411	0.00725972	0.836152	0.850671
1984. 4	0.828110	0.00727320	0.820837	0.835384
1984. 5	0.808841	0.00727831	0.801562	0.816119
1984. 6	0.787484	0.00728036	0.780204	0.794765
1984. 7	0.781134	0.00728126	0.773852	0.788415
1984. 8	0.782765	0.00728167	0.775483	0.790046
1984. 9	0.794607	0.00728188	0.787325	0.801889
1984.10	0.816632	0.00728198	0.809350	0.823914
1984.11	0.835337	0.00728204	0.828055	0.842619
1984.12	0.846532	0.00728208	0.839250	0.853814
Period				
1984. 1	0.855			
1984. 2	0.858			
1984. 3	0.852			
1984. 4	0.829			
1984. 5	0.809			
1984. 6	0.785			
1984. 7	0.779			
1984. 8	0.783			
1984. 9	0.799			
1984.10	0.834			
1984.11	0.842			
1984.12	0.848			

Tabla 1. Resultados estadísticos para la ecuación tratada.

Los resultados de la bondad de ajuste tanto para la predicción de valores futuros como para el tratamiento de los errores de estimación pueden observarse en los siguientes gráficos.

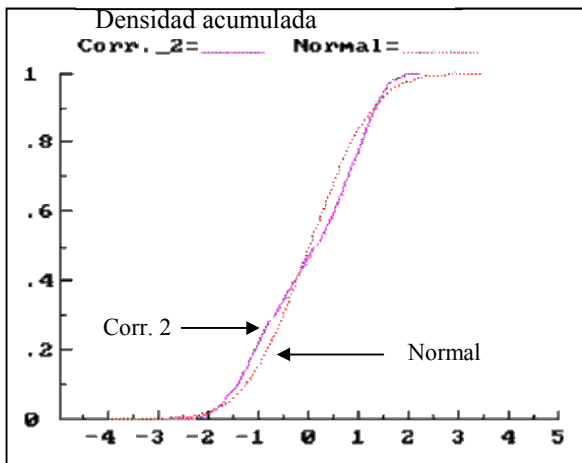


Figura 3: densidad acumulada estimada vs. observada.

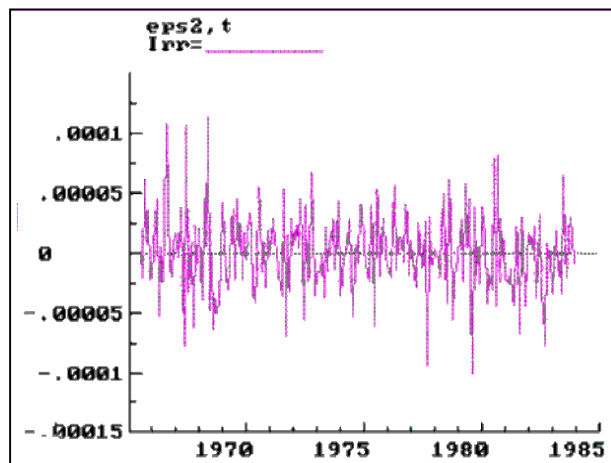


Figura 4: predicciones para la variable de ajuste.

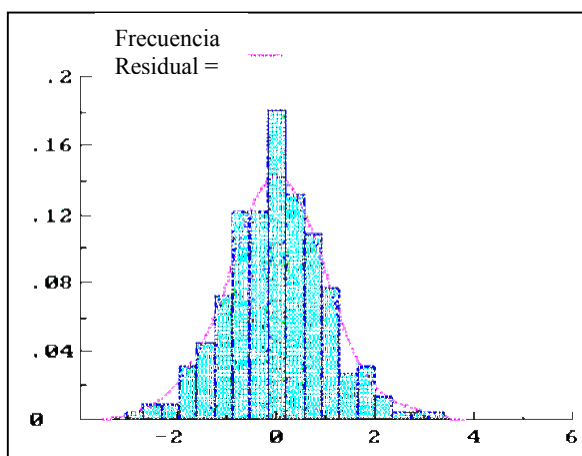


Figura 5: distribución estimada de los residuos.

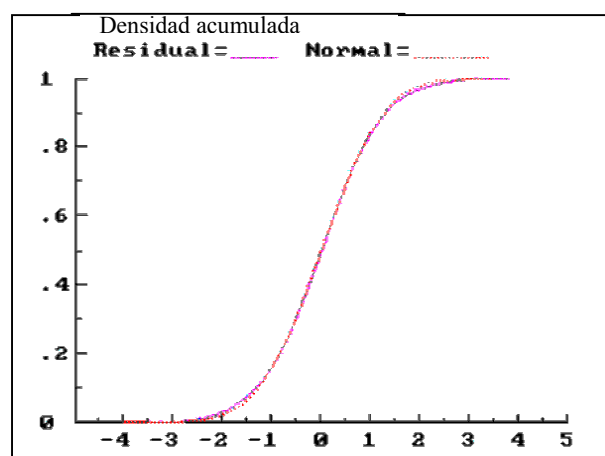


Figura 6: densidad acumulada estimada residual.

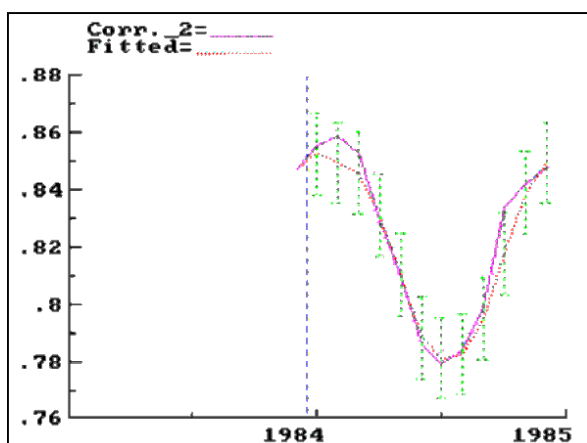


Figura 7: predicciones 12 meses hacia delante para y_t .
(modelo actual)

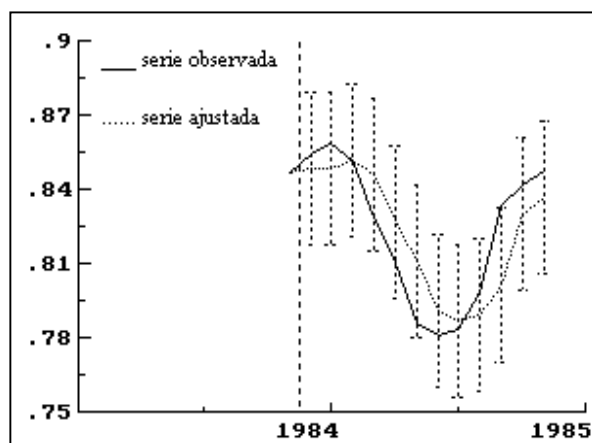


Figura 8: predicciones 12 meses hacia delante para y_t .
(modelo trabajo anterior)

Las predicciones mostradas en las figuras 7 y 8 reflejan comparativamente las mejoras introducidas con el nuevo modelo.

CONCLUSIONES.

Mediante el empleo del análisis estadístico estructural para series de tiempo irregulares, se consiguió modelar una serie de valores promedios mensuales de emisividad para la localidad de Salta.

El modelo obtenido significa una mejora sustancial respecto a la que se había obtenido el año pasado, permitiendo realizar estimaciones con un 95 % de confiabilidad. La mejora proviene del hecho de haber trabajado con una serie más larga, lo que permite disminuir el error en las predicciones (barras en las figuras 7 y 8).

En ambos casos, todos los valores caen dentro de los correspondientes intervalos de confianza, salvo el punto asociado a octubre. Este evento está contemplado dentro del nivel de significancia.

NOTA. El presente trabajo fue financiado por el Consejo de Investigación de la UNSa.

REFERENCIAS.

- Abril (1999). *Análisis de Series de Tiempo Basado en Modelos de Espacio de Estado*. Apuntes de curso. Tucumán.
- Durbin, J. y Cordero, M. (1994). *Handling Structural Shifts, Outliers, And Heavy-Tailed Distributions In State Space Time Series Models*. London School of Economics Statistics Research Report.
- Frigerio, E. y Avila Blas, O.,(1999). Análisis estadístico estructural de emisividad. *Avances en Energías Renovables y Medio Ambiente, Vol. 3, N° 1, 03.29-03.32*.
- Frigerio, E. y Saravia, L.,(1991). Emisividad infrarroja de la atmósfera. Medidas en el Noroeste Argentino. *Anales AFA, vol. 3, pp 432 – 438*.
- Siem Jan Koopman, Andrew C. Harvey, Jorgen A. Doornik and Neil Shepard.(1995). STAMP 5.0, Structural Time Series Analyser, Modeller and Predictor. *Chapman and Hall*.

ABSTRACT

A model of the sky emisivity for the zone was intended making a statistical study of the nocturnal radiation characteristics. A statistical analysis treatment with structural models was used which, at present, is a innovating technic. The improvement of the model of a nocturnal emisivity mensual data serie was searched using irregular series treatment, that implies to work with no lineal and no gaussian models. A structural model was obtained that allows to fit the observed data and make forecasts with a 95% of confidential level.