

# Modelos Autorregresivos espaciales para la simulación y pronósticos de enfermedades desde condiciones climáticas.

Autores. PAULO LÁZARO ORTIZ BUTO<sup>1</sup>, ALINA RIVERO VALENCIA<sup>1</sup> Y ANTONIO E. PÉREZ RODRÍGUEZ<sup>2</sup>

1. Centro del Clima. Instituto de Meteorología. E-mail paulo@met.inf.cu

2. Instituto de Medicina Tropical «Pedro Kourí».

---

## Resumen.

*El objetivo de este trabajo es presentar una breve introducción de los modelos espaciales y analizar el potencial uso de estos modelos para la simulación y el pronóstico de las condiciones de riesgos epidemiológicos desde situaciones climáticas anómalas, descritas por el índice climático  $IB_{t,p,c}$ . En este artículo son considerados índices de autocorrelación espacial, I de Moran y C de Geary que son usados para el estudio de la asociación y estructura espacial. Se dan los modelos Autorregresivos espacial (SAR) y promedio de medias móviles espacial (SAM) para la simulación y pronósticos de diferentes enfermedades a escala mensual. Los resultados muestran que los procesos son un buen enfoque para la simulación del comportamiento de las series epidemiológicas, teniendo en cuenta al índice climático como variable exógena en los modelos.*

*Palabras Claves: Sistema de predicción bioclimático (SPBC), predicción de enfermedades, modelos Autorregresivos espaciales, Índice de Moran, C de Geary.*

## Introducción.

De todos es conocido que el clima de una región está determinado por un conjunto integrado de factores (astronómicos, meteorológicos, geográficos) que se manifiesta a través del comportamiento de sus elementos principales (temperatura, humedad, presión, régimen de lluvia, etc). Esto presupone que el presente estudio se enfoque de forma compleja e integrada, mediante índices climáticos construidos bajo estas condiciones ya que se hace necesario que se trate de tener en cuenta que todos estos factores interactúan entre sí y que describen el clima y sus variaciones en un lugar dado.

La Organización Mundial de la Salud (OMS) en su trabajo WHO(2004) «Using climate to predict disease outbreaks: a review» plantea que los sistemas de vigilancia de las enfermedades proporcionan un medio de monitorear la incidencia de las entidades desde el tiempo y el clima. Dependiendo de la naturaleza del mismo puede ser un instrumento apropiado para detectar patrones inapropiados en los datos de incidencia entre otras funciones. Sin embargo, los sistemas de vigilancia epidemiológica de la enfermedad por sí solo no constituyen advertencia temprana, ni siquiera en los lugares donde la vigilancia se lleva a cabo dentro de una red especialmente diseñada con esos fines.

Luego los sistemas de vigilancia proporcionan un medio de descubrir las situaciones en lugar de predecir el comienzo de una epidemia (por consiguiente no hay tiempo de primacía como tal).

Sin embargo, un sistema de vigilancia adecuadamente diseñado debe propiciar de forma anticipada el momento de actuar, aumentando las oportunidades para la intervención con el fin de recomendar medidas de prevención que ayuden al control de las enfermedades teniendo en cuenta la información que brinda el propio sistema, como la que le suministra el sistema de pronóstico, de manera que les permita reaccionar y responder anticipadamente a los impactos y no continuar reaccionando después que ocurren.

Las ideas anteriores junto con la experiencia acumulada en las investigaciones de Clima y Salud, tanto en Cuba como en América Latina nos motivaron al desarrollo de modelos de predicción y de un sistema de pronóstico de enfermedades que incluyeran las predicciones de anomalías climáticas a partir de índices propuestos con tales propósitos (Ortiz, 2003) y (Ortiz, 2004), lo que nos permitió contar en estos momentos con un sistema de pronósticos para la predicción de enfermedades en el país.

Otro elemento a señalar es que los trabajos hasta ahora desarrollados en la temática de pronóstico de impacto de la variabilidad climática sobre la salud humana, presentan tres limitaciones fundamentales (Dowlatabadi, 1997):

- Se dificulta encontrar un factor climático que pueda describir el comportamiento de la enfermedad.

- Se utilizan modelos de regresión, en los cuales no se establece una generalización de causa-efecto, además estos no tienen en cuenta la heteroscedasticidad de las series de datos ni las variaciones temporales del fenómeno.

- Existe una alta multicolinealidad de las variables en estudio, como los de contaminación del aire.

Todos estos elementos, así como las valoraciones de párrafos anteriores hicieron que nos decidiéramos a realizar este estudio, solucionando las dificultades de los modelos empírico-estadísticos (biofísicos) propuestos hasta ahora para la evaluación de los riesgos climáticos; de manera que en este artículo se proponen modelos espaciales por enfermedades para la modelación y pronóstico de las condiciones de riesgo a la salud humana usando las salidas de las predicciones climáticas anómalas como variables exógenas descritas por índices climáticos.

## Materiales y métodos.

### Materiales.

Para la elaboración del estudio fueron utilizadas tres bases de datos con el siguiente contenido:

- información epidemiológica (datos)
- datos climáticos
- algunos índices de circulación.

### Información epidemiológica.

Los datos epidemiológicos fueron aportados por dos instituciones fundamentales, el Instituto de Medicina Tropical «Pedro Kourí» (IPK) y la Unidad de Vigilancia y Lucha Antivectorial (UNLAV) ambas pertenecientes al Ministerio de Salud de Cuba.

### Criterio de Inclusión.

- Un primer conjunto de datos contiene los reportes mensuales de casos de hepatitis virales (HV), enfermedades diarreicas agudas (EDA), las meningitis a escala provincial, para el período correspondiente a 1990-2004, disponibles en la Dirección Nacional de Estadísticas del Ministerio de Salud Pública de Cuba y de la base de datos aportada por el (IPK) en formato electrónico.

### Datos climatológicos

Para describir las condiciones climáticas en el país se tomaron las series de valores de los índices climáticos  $IB_{1,1,C}$ ,  $IB_{1,2,C}$  y  $IB_{1,3,C}$  para el período 1981 al 2004 cuya formulación e interpretación expresan Ortiz y Rivero, 2004.

### Datos sobre Índices de Circulación

Se usaron tres índices de circulación, cuya información está disponible en la base de datos de la NOAA-CIRES Climate Diagnostics Center (CDC), la Oscilación Cuasi-Bienal (QBO), la Oscilación del Atlántico Norte (NAO) y el Índice Multivariado de Eventos ENOS (MEI) para el período comprendido entre 1981 al 2004.

## Métodos.

### Medidas de asociación espacial (autocorrelación).

Probablemente las medidas de asociación espacial usadas más comúnmente para medir la autocorrelación espacial son el estadístico de Moran (Moran, 1948) y el de C de Geary (Geary, 1954). Estas pruebas indican el grado de asociación espacial como reflejo de un juego de datos en conjunto

En ambos casos se necesita calcular previamente la matriz de ponderaciones o pesos espaciales, que no es más que una matriz de unos y ceros, donde se le asigna el valor uno a las observaciones «j» que son vecinas de los puntos «i», y cero en los demás casos.

Esta forma de calcular la matriz está limitada a considerar el área de influencia a aquellos puntos que comparte un borde o límite, sin embargo, como en la práctica otras observaciones podrían estar influyendo, entonces se necesita usar no solo los puntos adyacentes, sino que se tomarán todos los puntos que disten de i a una distancia d, de forma tal que ahora la matriz de ponderaciones contenga uno en aquellos puntos j que se encuentren en un radio de distancia d respecto al punto i y no solo a los adyacentes.

El problema ahora se centra en que distancia tomar para definir la vecindad. Para ello Kelley y Dongya (2000) proponen efectuar el cálculo del vecino más cercano, algoritmo que permite obtener la medida entre las que se encuentra la mínima distancia euclidiana de puntos georeferenciados. Siendo esta la distancia a la cual solo un par de puntos de la muestra son vecinos, también se puede usar la máxima distancia con los mismos propósitos, pero ahora esta representa el segmento sobre el cual todos los puntos de la muestra tienen al menos un vecino.

### Índice de Moran

El contraste de dependencia espacial de Moran, es el índice usado con más frecuencia en el análisis de correlación espacial. El cual viene expresado mediante la siguiente ecuación (Cliff and Ord, 1981).

$$I = \frac{N}{S_o} = \frac{\sum_{i \neq j} \omega_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad i \neq j \quad (1)$$

donde:

N es el tamaño de la muestra,  $S_o = \sum_i \sum_j w_{ij}$  x es la variable sometida al análisis de dependencia espacial y  $w_{ij}$  son las ponderaciones de la matriz binaria W que se construye asignándole 1 a las observaciones vecinas al punto i, y cero a las demás observaciones.

Su interpretación es muy sencilla, pues cuando existe correlación espacial positiva el valor del coeficiente I es positivo, y viceversa.

Para evaluar su significación estadística, se estandariza este índice y se compara con la distribución normal. Para lo cual se procede a calcular la media  $\mu(I)$ , y la desviación estándar,  $\delta(I)$ , mediante la siguiente expresión:

$$\mu(I) = -1/(N-1)$$

$$\delta(I) = \sqrt{\frac{N(N^2 - 3N + 3)S_1 - N^2 S_2 + 3S_3^2 - b_2(N^2 - N)S_1 - 2N S_2 + 6S_3^2}{(N-1)(N-2)(N-3)S_o^2}} \quad (2)$$

siendo,

$$b_2 = \frac{m_2}{m_1^2}; \quad m_1 = \frac{\sum_i z_i^2}{N}; \quad m_2 = \frac{\sum_i z_i^4}{N^2}; \quad z_i = \sum_j w_{ij} w_{ij} \quad (3)$$

$$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2; \quad S_2 = \sum_i \sum_j (w_{ij} + w_{ji})^2; \quad w_{ij} = \sum_{j=1}^N w_{ij} \quad (4)$$

y  $Z_i$  es la variable x estandarizada (expresada como la desviación respecto a la media dividida por la desviación estándar).

El valor del Moran's I estandarizado,  $Z(I)$ , sigue una distribución normal estándar, de tal manera que si  $Z(I) > 1.96$  habrá autocorrelación espacial positiva, significativa al 5%. De igual manera si  $Z(I) < -1.96$  habrá autocorrelación espacial negativa. Se parte siempre de la hipótesis nula de que no existe autocorrelación espacial.

Índice C de Geary.

Otro índice empleado para el análisis de dependencias espaciales es la C de Geary. Con este índice se evalúa la hipótesis nula de la no existencia de dependencia espacial. La formulación de C tiene en cuenta algunos elementos ya definidos en el índice Moran's I y viene dada por la siguiente expresión (Geary, 2001):

$$C = \frac{N-1}{2S_o} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (5)$$

siendo  $i \neq j$

El índice C estandarizado,  $Z(C)$ , se distribuye asintóticamente normal, de tal manera que se debe evaluar la significación de este estadístico calculando su media y la desviación estándar que, bajo la hipótesis de muestreo aleatorio, pueden hallarse mediante:

$$\mu(C) = 1,$$

$$\delta^2(C) = \frac{(N-1)S_1[N^2 - 3N + 3 - (N-1)b_2] + S_2^2[N^2 - 3 - (N-1)^2 b_2]}{N(N-2)(N-3)S_o^2} \quad (6)$$

$$\frac{(N-1)S_1[N^2 + 3N - 6 - (N^2 - N + 2)b_2]}{4N(N-2)(N-3)S_o^2}$$

Los términos empleados para el cálculo son definidos similarmente a como se procedió para el cálculo del índice de Moran. La estandarización del índice, es realizada mediante:

$$Z(C) = \frac{C - \mu(C)}{\delta(C)} \quad (7)$$

Esto nos permite evaluar la significación del estadístico C comparándola con los valores críticos de la distribución normal estándar. La autocorrelación espacial positiva se presenta cuando el índice C toma valores estadísticamente significativos en el rango entre cero y uno. Cuando los valores de C se sitúan entre uno y dos, se presenta autocorrelación negativa, lo que significa que hay autocorrelación espacial significativa, pues en caso contrario se deduce que no existe autocorrelación espacial.

La distancia de Moran mide las asociaciones basada en el cálculo del producto cruzado de las observaciones, mientras que la C de Geary se basa en el cuadrado de las diferencias de las observaciones.

Sin embargo, todos estos indicadores de análisis de estructuras o patrones espaciales, son solo el comienzo de este tipo de análisis, ya que ellos nos indican en que dirección están las mas agrupadas y a que tipo de patrón siguen las enfermedades que se estudian, pero no explican por si mismos por qué tal agrupación ocurre, ni qué factores determinan su forma y fuerza. En otras palabras, la hipótesis alternativa de «autocorrelación espacial» es demasiado vaga para ser muy útil en la construcción de la teoría del problema. Un concepto que formaliza la manera en la que la asociación espacial se genera es el proceso espacial, o proceso espacial

estocástico. Estrictamente hablando, semejante procesos expresan como las observaciones a cada localización  $i$  dependen de los valores a las localizaciones vecinas  $j$ , es decir, en los rezagos (retardos) espaciales.

### Modelos de procesos espaciales.

Análogo a la terminología el análisis del uso de las series de tiempo (Box and Jenkins, 1994), los procesos estocásticos espaciales pueden ser clasificados como procesos Autoregresivo espacial (SAR) o procesos de promedio móviles espacial (SMA). La combinación de los dos procesos, con proceso Autoregresivo para la asociación espacial en la variable dependiente, y un proceso de promedios móviles para los errores, se llama un proceso espacial autoregresivo de promedios móviles (SARMA). Estos pueden ser lineales o Bilineales.

Los errores de medición pueden ser generados por diferentes situaciones por ejemplo: la propia unidad de estudio, ya que no podemos determinar si la delimitación es o no por un criterio pues por lo general se toma aleatoria: las divisiones políticas de los municipios por dar cuenta, la presencia tecnológica y automatización de la agregación de los datos aumentando el riesgo de no detectar los errores y por último la propia combinación de procesos espaciales.

En cambio la heterogeneidad es causada por la falta de estabilidad espacial del comportamiento de las series o de las relaciones que están siendo analizadas, las cuales pueden ser debido al tamaño de la muestra o área observada, otro aspecto no menos importante a considerar es la heterocedasticidad que esta provoca

### Modelos Autoregresivos espaciales.

Los modelos para procesos espaciales presentan la siguiente forma general (Anselin 1992):

$$Y = \rho W_i y + X\beta + \varepsilon \quad (8)$$

Donde  $y$  es la variable dependiente, « $\rho$ » o rho es el coeficiente autoregresivo de la variable dependiente espacialmente ligada o retrasada (retardo espacial;  $W_1, W_2$ ),  $W_i$  es precisamente la matriz de unidades vecinas o matriz de pesos espaciales, « $X$ » es la matriz de variables independientes, « $\beta$ » son los respectivos coeficientes, y « $\Sigma$ » es el error.

El término  $\varepsilon$  lo podemos expresar por:

$$\varepsilon = \lambda W_2 \varepsilon + \mu \quad (9)$$

donde  $\lambda$  es el coeficiente escalar del error espacial, el cual representa los choques,  $\mu \sim N(0, \Omega)$ .

Específicamente, el coeficiente  $\rho$  es el promedio de la variable dependiente en las áreas vecinas. Como se mencionaba, este coeficiente es un término autoregresivo del indicador de la enfermedad a estudiar, dirigido a medir la autocorrelación espacial. Se le denomina «efecto espacial» ya que al ser independiente por alguna causa, permite probar la hipótesis y medir el efecto que tienen los niveles de la variable dependiente en las áreas vecinas y sostener la inferencia de un efecto contextual.

Existen dos tipos fundamentales de modelos espaciales Autorregresivos que son empleados en los análisis estadísticos y que parte de la ecuación general anterior, ello son los Simultáneos Autorregresivos (SAR) y los Condicional Autorregresivos (CAR). Donde la diferencia entre ellos estriba en la especificación de la función de covarianza inversa.

1.2.2 Modelos espaciales (SAR) y (CAR) Anselin, 1993.

La diferencia entre ellos esta dada por la forma que adopta la matriz de covarianza, pues en el caso de los modelos SAR la función de covarianza inversa esta dada por la siguiente especificación:

$$(1-\rho W)^{-1} (1-\rho W) / \sigma^2 \quad (10)$$

Donde  $W$  es un matriz de pesos espaciales estandarizada por regiones y  $1 \leq t \leq 12$ .

Mientras que para el caso del CAR, la matriz de covarianza esta dada por:

$(1-\rho C) / s^2$ , donde  $C$  es la matriz binaria de contigüedades simple.

### Procesamiento de la información.

Para el cálculo del coeficiente de autocorrelación espacial  $I$  de Morán, se predeterminó una matriz de vecinos con aquellas estaciones que se encontraran en un radio de 5 kilómetros, la cual es la mínima distancia en la cual todas las estaciones presentaban por lo menos una vecina. Para efectuar el ajuste y corrida de los modelos y la técnica SARMA se utilizó el SIG «ArcView 3.2» soportado sobre Windows junto con el módulo «Spatial-Statistics» del sistema estadístico «S-Plus 2000» para Windows XP.

### Resultados.

#### Asociación espacial

El coeficiente  $I$  de Morán y  $C$  de Geary son las técnicas estadísticas utilizadas con más frecuencia para probar esta ley geográfica; o en una fase del

análisis probar las hipótesis de la presencia de una autocorrelación espacial de una variable y sus correspondientes niveles de concentración o de dispersión. Su formulación es muy similar al del coeficiente de correlación  $r$  de Pearson (Arselin, 1993). También sus valores varían entre +1 y -1, en donde el primer valor significa una perfecta concentración o autocorrelación espacial y el segundo una perfecta dispersión espacial. El cero significa un patrón espacialmente aleatorio o sin orden.

La autocorrelación espacial en los datos de salud la interpretamos como la similitud de los valores de los patrones entre distintas ubicaciones, es decir los valores altos o bajos de una variable tienden a estar rodeados de vecinos con similitud de comportamiento en sus valores. Como la aparición de casos de una enfermedad en un área condiciona la aparición de los casos en otras áreas vecinas (Kwan et al, 2004), resulta lógico que las enfermedades presenten correlación espacial. Los resultados que se muestran en la Tabla 1, corroboran la idea anterior.

Tabla 1. Asociación espacial de cada enfermedad según los índices  $I$  de Moran y  $C$  de Geary.

Enfermedades	I de Moran	Z. Valor	probabilidad
IRA	0.31	2.92	0.000
MB	0.20	2.01	0.000
MV	0.18	2.00	0.000
Varicela	0.21	2.21	0.000
EDA	0.30	2.63	0.000
HV	0.32	2.88	0.000
Neumococo	0.19	2.89	0.000
	C de Geary	Z. Valor	probabilidad
IRA	0.55	-3.682	0.006
MB	0.69	-4.101	0.000
MV	0.56	-3.591	0.003
Varicela	0.37	2.314	0.009
EDA	0.82	-8.005	0.000
HV	0.61	-6.372	0.000
Neumococo	0.31	2.231	0.000

Según los resultados que se exponen en la tabla 1, se observan valores positivos y significativos para ambas distancias, que indican dependencia espacial positiva, con una tendencia a la concentración de las tasas de las enfermedades en aquellas provincias que presentan condiciones similares tanto climáticas como higiénico-sanitarias.

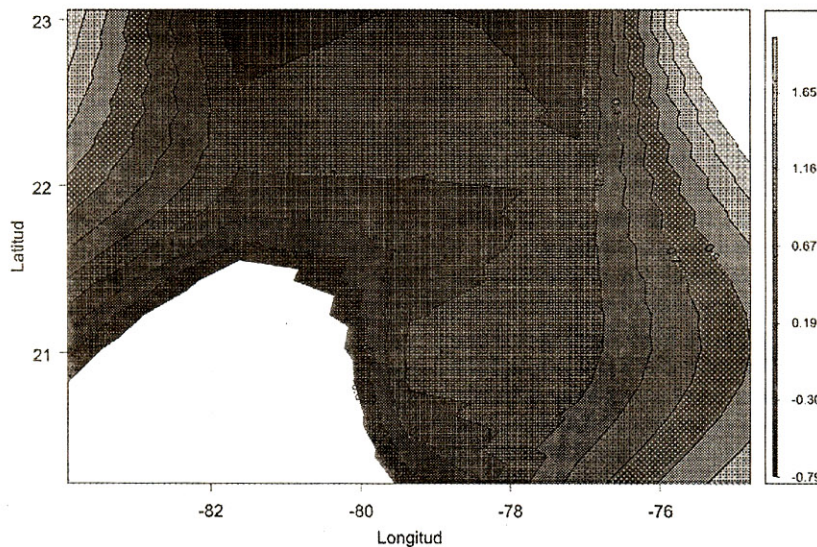


Fig. 1. Distribución espacial de la varicela acorde con la  $I$  de Moran

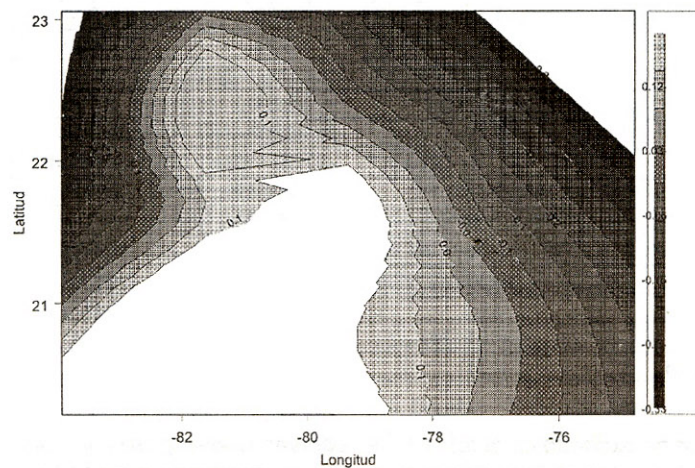
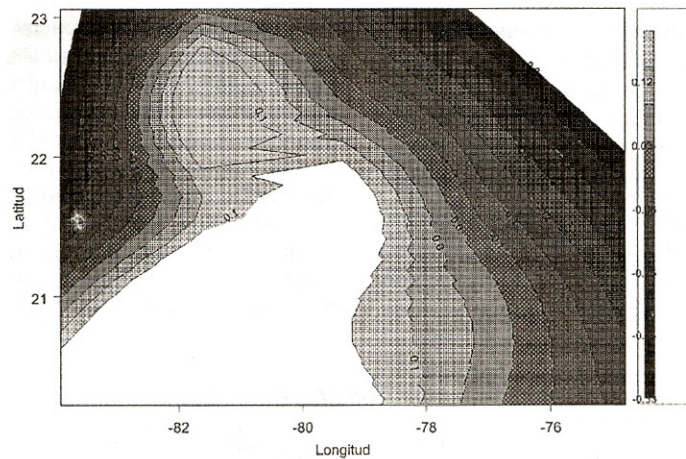


Fig. 3. Distribución espacial de la hepatitis viral acorde con la I de Moran

El análisis de la distribución espacial de las enfermedades según el I-Moran mostró una fuerte correlación espacial positiva con una tendencia a la concentración por regiones en todos los casos. Mostrándose en el caso de las de transmisión respiratoria, que la distribución y conglomerados se acentúan hacia la región central. En el caso de las varicelas, se observan los mayores riesgos en la zona central del país, los cuales se identifican con áreas más frías y secas con mayor radiación en el período poco lluvioso, condiciones que favorecen la irritación de las vías respiratorias y provocan un aumento del riesgo epidémico en la zona occidental y oriental las cuales presentan las tasas más bajas. Estos resultados concuerdan con los del C de Geary los cuales toman valores entre cero y uno, evidenciando una clara asociación espacial. Fig. 1.

En el caso de las enfermedades de transmisión digestiva las áreas más vulnerables se presentan en la región oriental seguida por la región occidental (en su extremo más oriental). En el primer caso están muy asociadas con los procesos intensos de sequía que se han venido produciendo en los últimos años en esa región lo que provoca afectaciones en la calidad de las aguas y la higiene de los alimentos que al conjugarse con las condiciones físico-geográficas y sociales provocan tasas de incidencia elevadas de esta enfermedad haciéndolas particularmente diferente al resto del país. Las condiciones que provocan las alzas en la región occidental son muy diferentes, ya que estas no aparecen vinculadas por procesos de sequías intensos, sino que están asociadas con el comienzo de la temporada lluviosa,

aspecto este que está recogido por el cambio de signo en las asociaciones. Luego tenemos un mismo impacto con dos procesos climáticos diferentes que ambos generan afectaciones que se ven acentuadas debido a las diferentes condiciones higiénico sanitarias en cada área, Fig. 2 y 3.

### Estimación y ajuste de los modelos para cada enfermedad.

A continuación se muestran los resultados del ajuste de los modelos SAR para cada una de las enfermedades Tabla 2

En la misma tabla se muestran los parámetros y coeficientes espaciales para cada uno de los modelos propuestos, destacándose la alta significación y

calidad de los ajustes logrados, con lo que se confirma que la distribución espacial de estas enfermedades no son aleatorias, sino que están sujetas a las características físico geográficas del lugar, las variaciones del clima y las propias características de la enfermedades que encuentran condiciones mas favorables para su desarrollo en unas regiones mas que en otra.

Los resultados mostrados en la tabla evidencian que los modelos espaciales de resolución simultanea (estimación de los coeficientes junto con el resto de los parámetros), son adecuados para nuestros objetivos e hipótesis, debido a que nuestro propósito era modelar la dependencia espacial y no filtrarla y los modelos autorregresivos nos han proporcionado simulaciones adecuadas de esta dependencia espacial encontradas en el análisis del epígrafe anterior.

Tabla 2. Resultado del ajuste de los modelos autorregresivos espaciales (SAR).

Enfermedades	VARIABLES	Coefficiente	Errores estándares	Estadístico t	Probabilidad
Número de focos (NF)	Constante	35.8	9.42	4.770	0.0063
	NF <sub>t-1</sub>	.917	0.022	21.4531	0.0000
	IBP <sub>t,c</sub>	-56.3715	0.121	13.210	0.0000
	Población	5.20	0.001	3.831	0.0501
	Efecto espacial	Rho = -0.84			
	Log-likelihood	-69.99			
EDA	Constante	- 3702.64	1321.44	2.8020	0.0231
	EDA <sub>t-1</sub>	0.8701	0.1283	6.7826	0.0001
	IBP <sub>t,c</sub>	32.2355	12.6	11.1	0.0000
	IBP <sub>t,c</sub>	2221.2	19.32	13.8	0.0000
	longitud	0.0030	0.0008	3.7343	0.0058
	latitud	0.0089	0.0019	4.5911	0.0018
	Efecto espacial	Rho=-0.98			
Log-likelihood	- 141.77				
HIV	Constante	106.6639	100.4509	2.0619	0.03193
	HV <sub>t-1</sub>	0.3361	0.0962	3.4928	0.0082
	IBP <sub>t,c</sub>	- 13.7014	26.5782	-3.5155	0.0062
	IBP <sub>t,c</sub>	93.4096	23.8881	3.9103	0.0045
	longitud	6.5516	2.4104	2.7181	0.0263
	latitud	15.4839	5.8633	2.6408	0.0297
	Efecto espacial	-0.344			
Log-likelihood	-55.68				
IRA	Constante	1212.2677	573.6472	2.1133	0.0582
	IRA <sub>t-1</sub>	202.1370	40.0995	7.3777	0.0000
	IBP <sub>t,c</sub>	1.0999	0.1619	6.7919	0.0000
	IBP <sub>t,c</sub>	1.2	0.121	6.671	0.0000
	Efecto espacial	Rho=-0.139			
	Log-likelihood	-125.7			
MV	Constante	-2.1984	1.0799	-2.2421	0.0481
	MV <sub>t-1</sub>	0.6115	0.1121	5.4567	0.0002
	IBP <sub>t,c</sub>	5.6147	6.1527	3.6887	0.0050
	Efecto espacial	Rho=-.1622			
	Log-likelihood	-40.46			
Varicela	Constante	-31.3346	23.0674	-2.3584	0.0120
	VAR <sub>t-1</sub>	0.4777	0.0117	40.7465	0.0000
	IBP <sub>t,c</sub>	11.0326	13.0569	2.8450	0.0042
	IBP <sub>t,c</sub>	11.9499	10.0739	1.862	0.0263
	Efecto espacial	Rho = .8812			
	Log-likelihood	-53.37			

Por otra parte, en términos de dependencia espacial y según los resultados obtenidos en el test I de Moran, puede decirse que el término  $rWY$  en el modelo está comprobando la dependencia espacial de las variables estudiadas y que en este contexto los modelos propuestos resultan los indicados para estos estudios.

También de los resultados se infiere que las relaciones espaciales están moduladas por efectos endógenos (por el patrón de comportamiento de la propia serie) y exógenos (condiciones climáticas, sociales etc), así como por el comportamiento de los vecinos (áreas cercanas).

### Análisis de la capacidad predictiva de los modelos.

Debido a cuestión de tiempo y espacio nos limitaremos a mostrar tan solo un ejemplo que confirma la plausibilidad del uso los modelos autorregresivos y su capacidad de predicción. En la Fig. 4 se muestran los valores pronosticados y reales del indicador entomológico, donde se constata la gran concordancia entre los valores.

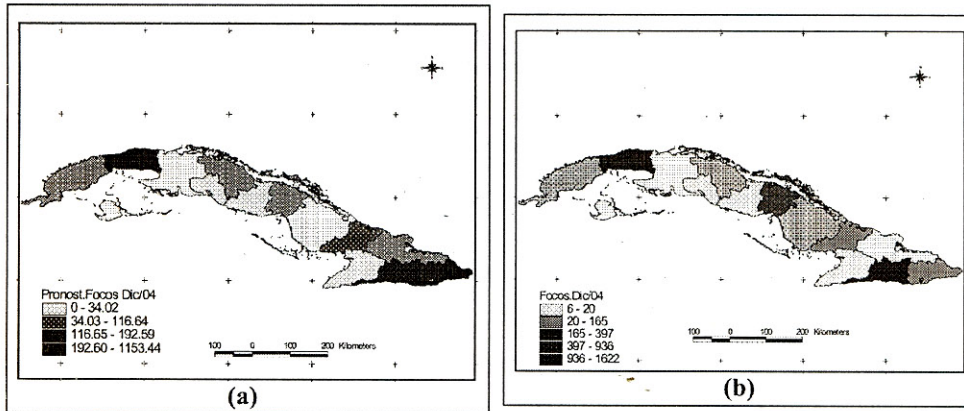


Fig 4. Pronóstico del número de focos (a) y valor real (b) para el mes de diciembre del 2004

Resultados muy similares se obtienen para el resto de las enfermedades, obteniendo una efectividad de pronóstico entre un 85% a 90%, lo que corrobora que los modelos usados para tales fines resultan adecuados para su inclusión en el sistema de pronósticos.

Predicciones por enfermedades.

Teniendo en cuenta las salidas de los modelos de regresión espacial anteriores se muestran a continuación los pronósticos por enfermedades en las Figs. 5 - 10, para el mes de octubre del 2004 de las EDA, IRA, HV, MV, MB, VAR, y Meningitis a neumococo, observándose variaciones del riesgo para las diferentes provincias según las entidades.

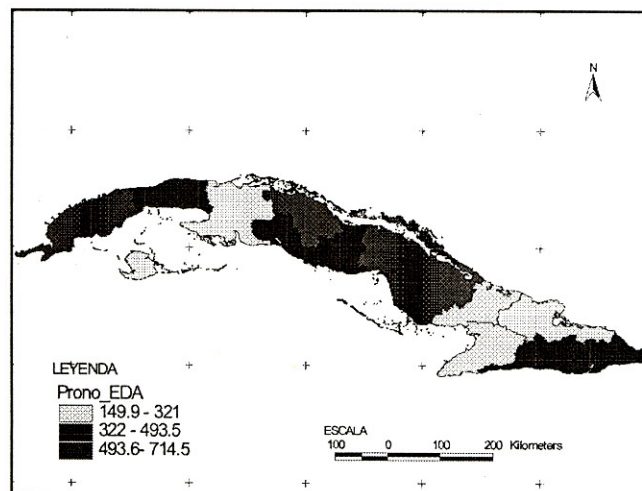


Fig. 5. Tasas (100 000HAB) de atenciones esperadas por EDA para Octubre/2004



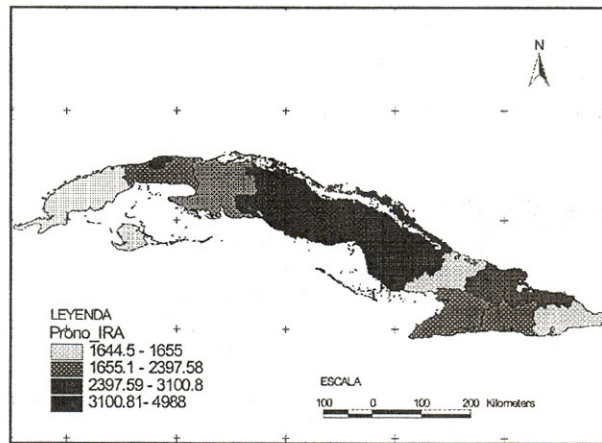


Fig 6. Tasas (100 000HAB) de atenciones esperadas por IRA para Octubre/2004

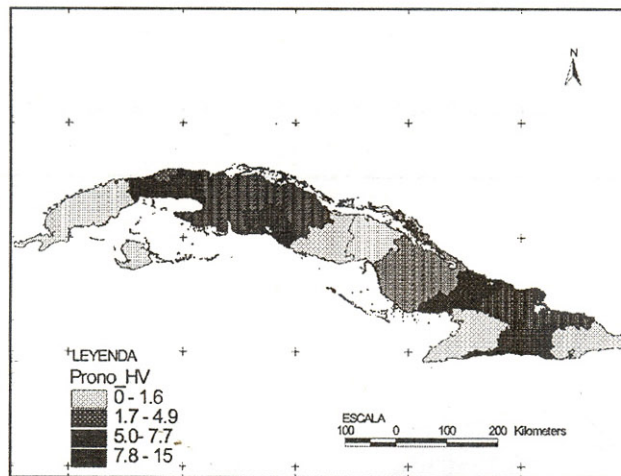


Fig 7. Tasas (100 000HAB) de atenciones esperadas por HV para Octubre/2004.

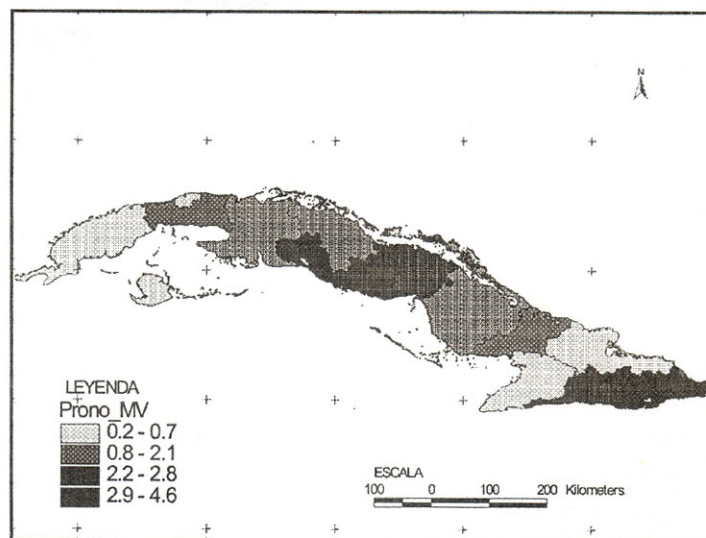


Fig. 8. Tasas (100 000HAB) de atenciones esperadas por MV para Octubre/2004

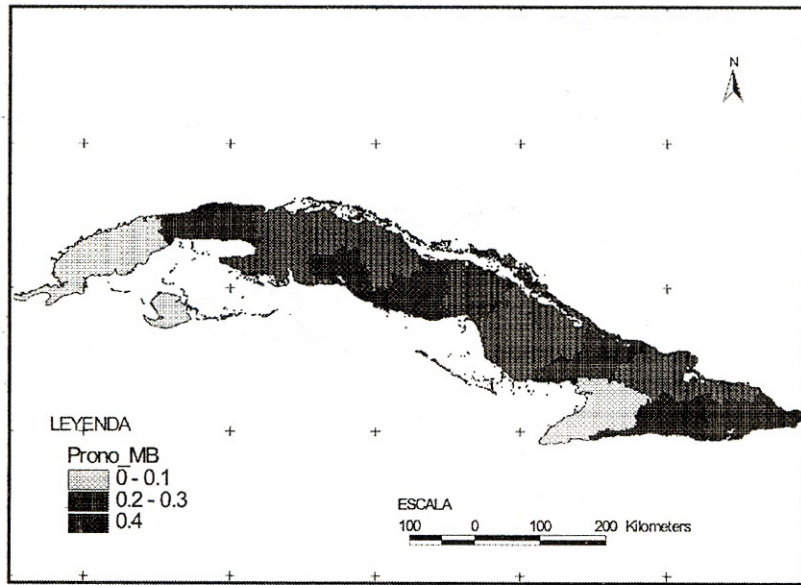


Fig. 9. Tasas (100 000HAB) de atenciones esperadas por MB para Octubre/2004

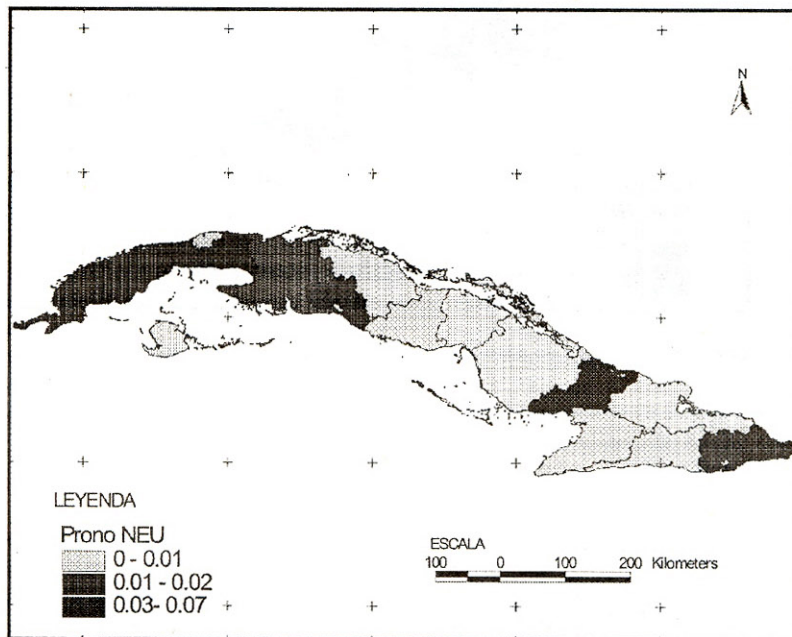
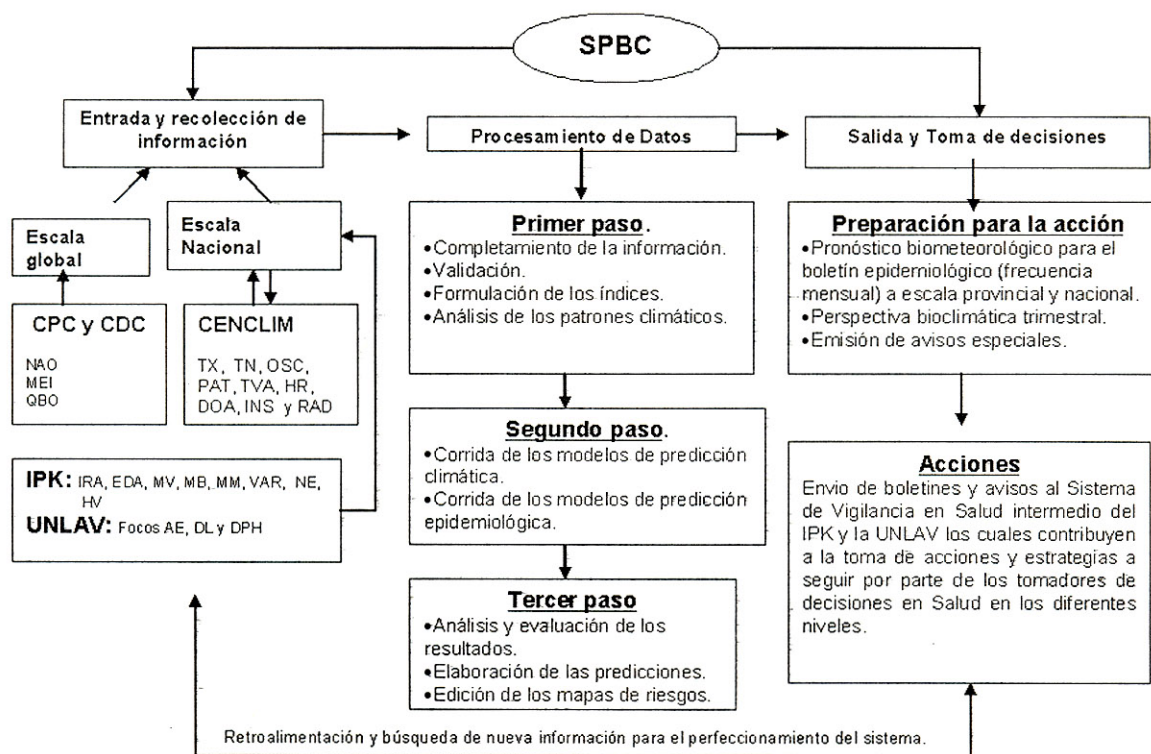


Fig. 11. Tasa (100 000HAB) de atenciones esperadas por meningitis a neumococo para Octubre/2004

Los resultados mostrados forman parte de las salidas del sistema de vigilancia y pueden ser consultados en el boletín del IPK, en la siguiente pagina Web: <http://www.ipk.sld.cu/bolepid/bolepid.htm>, los mismos constituyen la componente de acción para la toma de decisiones dentro del sistema de predicción bioclimático de Cuba. (SPBC), cuyo esquema se muestra a continuación.



Esq.1 Funcionamiento del Sistema de Pronóstico Bioclimático de Cuba (SPBC).

## Conclusiones.

Se evidencia la idoneidad del uso práctico de los modelos Autorregresivos espaciales para la simulación y predicción del comportamiento de las enfermedades, pues estos simulan de manera adecuada los movimientos y variaciones de cada una de ellas, sugiriendo una mayor capacidad predictiva del riesgo en cada área.

Se muestra la utilidad del análisis de datos espaciales, particularmente los índices de correlación espacial de I de Moran y el Scatter-Plot de Moran para los cuales resultaron adecuados para detectar y ubicar patrones espaciales y conglomerados de cada una de las enfermedades.

Con estos modelos quedan superadas las muchas de las limitaciones de los modelos hasta ahora usados, ejemplo los de regresión o análisis canónico alcanzándose con estos modelos, predicciones con alto nivel de efectividad, lo cual contribuyen al perfeccionamiento del sistema de vigilancia en Salud para poder actuar anticipadamente ante una contingencia.

Se corrobora que los modelos Autorregresivos espaciales resultan adecuados para la modelación, y predicción de las contingencias en salud debido a los

brotos epidémicos, así como explicar de alguna manera la distribución espacial de las enfermedades, lo que facilita la propuesta de modelos de predicción de los riesgos a escala nacional y provincial.

## Referencias.

- Anselin, L, 1988.** *Spatial Econometrics: Methods and Models.* Kluwer. Dordrecht.
- Anselin, L, 1992.** *Space Stat, a program for the analysis of spatial data* (National Center for Geographic Information and Analysis. University of California, Santa Barbara, CA, USA.
- Anselin, L, 1993.** "The Moran Scatter plot as an ESDA tool to assess local instability in spatial association", GISDATA Specialist Meeting on GIS and Spatial Analysis, Amsterdam.
- Anselin, L, 2001.** Notas de Clases del Curso Econometría Espacial, <http://geog55.gis.uiuc.edu/ace492se/syllabus2001.html>
- Box, G. E. P., Jenkins, G. M. and Reinsel, G. C, 1994.** *Time Series Analysis: Forecasting and Control.* 3rd edition. Prentice-Hall Inc., New Jersey.
- Cliff, A. and Ord, J. K, 1981.** *Spatial Processes, Models and Applications,* London: Pion.

**Dowlatabadi, H, 1997** Assessing the Health Impacts of Climate Change. *Climatic Change* 35 137-144.

**Geary, R.C., 1954.** The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5, 15-145.

**Geary, R, 2001.** «*The contiguity ratio and statistical mapping*», *The Incorporated Statistician*, No. 5, 1954. Citado por Mike Sawada, «*Global Spatial Autocorrelation indices Moran's I, Geary's C and the General Cross- Product Statistic*», Mimeo, Laboratory of Paleo climatology and Climatology, Dept. Geography, University of Ottawa.

**Moran, P. A. P. 1948.** The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B*, 10:243-251.

**Ortiz, B.P.L, Rivero A, 2004.** Índices climáticos para la determinación y simulación de las señales de la variabilidad climática en diferentes escalas espacio temporales. *Revista Cubana de Meteorología*. Vol 11 No 1: 41-52.

**Ortiz, B.P.L, Rivero AV, Pérez A, Vega L N, Pérez C A, 2003.** The climatic variability and their effects in the variations of the patterns space – time of the diseases and their economic impact. Technical Report 82. Research Climate Center. INSMET. National Program the Global Changes and the Evolution of the environment Cuban. In Spanish.

**S-Plus, 2000. Professional Release 2. Copyright(c) 1998-1999.** Mathsoft .inc.

**Kwan,W; Leszczynsk A; Schaub P; Tan P and Wendw L, 2004.** Tuberculosis spatial Epidemiology in greater vancouver, <http://www2.sfu.ca/geog452spring02/group4/index.html>.

**Kelley P and Dongya Z, 2000.** Closed- Form Maximum Likelihood Estimates of Nearest Neighbour Spatial Dependence. *Geographical Analysis*, Vol, 32, No 2 April.

**WHO, 2004.** Using climate to predict Diseases Outbreaks: A review. Geneva: World Health Organization (WHO/SDE/OEH/04.01.).

## Agradecimientos.

Deseo agradecer en primer lugar al Instituto Interamericano para la Investigación del Cambio Global (IAI) por el apoyo financiero, bajo el amparo del proyecto SG-037 que nos permitió el desarrollo e implementación de esta metodología, además deseo agradecer también al equipo de trabajo y técnicos del grupo de Clima y Salud que sin ellos no fuera posible un trabajo de esta magnitud. Por ultimo y no menos importante deseo dar mi mayores agradecimientos al programa Nacional «Los Cambio Globales y la Evolución del Medio Ambiente Cubano» que garantizo la obtención de los resultados y el resto del financiamiento al grupo de Clima-Salud mediante el proyecto Nacional N° 01304166.

### Abstract.

*The objective of this work is to present a brief introduction the spatial models and analyzes the potential use of its models for simulation and forecasts of epidemic risks from anomalous climatic conditions, described by climatic indexes  $I_{B,C}$ . In this paper are consider two Spatial Autocorrelation indices – Moran's I, Geary's C that are used in the studies structure and spatial association. The spatial Autoregressive process ((SAR) and spatial Moving Average process (SMA) is given to simulate and forecast the different diseases to monthly scale. The results show that processes are a good fit for simulation the behavior of epidemic series, taking into account to the climate index as exogenous variable in the models.*

**Key words:** Bioclimatic system prediction (BSPC), Prediction diseases, spatial autoregressive, spatial moving average Moran's I, Geary's C.