Teledetección y Medio Ambiente. IV Reunión Científica Asociación Española de Teledetección.
Junta de Andalucía - Consejería de Cultura y Medio Ambiente.
Sevilla, 13, 14 y 15 de noviembre de 1991. pp. 78-85.
Editor: J. M. Moreira

ANÁLISIS ESTADÍSTICO PARA LA CORRECCIÓN GEOMÉTRICA DE IMÁGENES DE SATÉLITE

J.A. Ardizone, A. Arozarena, J. Delgado, M. Herrero, G. Villa y P. Vivas.
Area de Teledetección I.G.N. Madrid

RESUMEN

En este trabajo, se realiza un análisis estadístico inferencial para determinar y simplificar los modelos polinómicos de regresión utilizados en la corrección geométrica mediante puntos de control del terreno, de dos imágenes digitales de satélite.

ABSTRACT

In this paper, an inferencial statistical analisys is achieved in order to assess an simplify polinomical models of regression, used in a geometrical correction through ground control points, of two digital satellite images.

1. INTRODUCCIÓN

Las imágenes de satélite no son mapas. Sin embargo, la información extraída de ellas se suele integrar con datos cartográficos en un sistema de información geográfica o se presenta a los usuarios en forma de mapa, como ocurre con los mapas meteorológicos.

A la transformación que es necesario aplicar a una imagen de satélite de manera que adquiera propiedades de mapa, tales como sistema de proyección y escala, se le denomina corrección geométrica (Mather, 1987).

Existen varios procedimientos para realizar la corrección geométrica de una imagen de satélite. Uno de ellos, generalmente conocido como "método de los puntos de control", consiste en establecer una transformación matemática

$$x = x(u,v) \qquad y = y(u,v) \tag{1}$$

entre las coordenadas (u,v) de puntos de un mapa de la zona cubierta por la imagen y las coordenadas (x,y) de puntos bien definidos sobre la misma, denominados puntos de control (Richards, 1986), los cuales son identificados en ambos documentos. Si las ecuaciones (1) fueran conocidas se podría localizar cada punto en la imagen conociendo las coordenadas mapa; y, mediante un proceso de interpolación, asignarse a cada punto de la imagen corregida, la cual desearíamos que coincidiera con el mapa, un nivel de brillo.

Desgraciadamente, la forma explícita de las funciones (1) se desconoce; por lo que generalmente suelen elegirse para su representación polinomios de segundo o tercer grado, los cuales podrían considerarse como los primeros términos de un desarrollo en serie de las funciones (1). Así pues, en el caso de un segundo grado obtendríamos:

$$\begin{aligned} x &= a_0+a_1u+a_2v+a_3uv+a_4u^2+a_5v^2+\epsilon_x \\ y &= b_0+b_1u+b_2v+b_3uv+b_4u^2+b_5v^2+\epsilon_y \end{aligned} \tag{2}$$

donde ϵ_x y ϵ_y serían las perturbaciones o errores introducidos al tomar la forma polinómica en lugar de la forma explícita real de las funciones x e y.

A partir de las coordenadas del conjunto de puntos de control, el cual puede considerarse una muestra de puntos de la imagen y del mapa, y mediante el método de ajuste por mínimos cuadrados, puede estimarse los b_i y ϵ_y, con lo que las expresiones (2) serían ahora:

$$\begin{aligned} x &= \hat{a}_0+\hat{a}_1u+\hat{a}_2v+\hat{a}_3uv+\hat{a}_4u^2+\hat{a}_5v^2+e_x = \hat{x}+e_x \\ y &= \hat{b}_0+\hat{b}_1u+\hat{b}_2v+\hat{b}_3uv+\hat{b}_4u^2+\hat{b}_5v^2+e_y = \hat{y}+e_y \end{aligned} \tag{3}$$

donde $\hat{a}_i$ y $\hat{b}_i$ son los estimadores de los coeficientes a_i y b_i y los términos e_x y e_y son los residuos, los cuales nos permiten estimar las perturbaciones.

El conjunto de puntos de control, también nos define otra transformación, en este caso de imagen a mapa, cuya fórmula análoga a (3), sería

$$\begin{aligned} u &= \hat{c}_0+\hat{c}_1x+\hat{c}_2y+\hat{c}_3xy+\hat{c}_4x^2+\hat{c}_5y^2+e_u = \hat{u}+e_u \\ v &= \hat{d}_0+\hat{d}_1x+\hat{d}_2y+\hat{d}_3xy+\hat{d}_4x^2+\hat{d}_5y^2+e_v = \hat{v}+e_v \end{aligned} \tag{4}$$

y aunque no suele considerarse interesante, en algunos casos, se suele determinar junto con la (3), para ajustar las coordenadas mapa de los límites de la imagen corregida (Mather, 1987).

El presente trabajo está dedicado al análisis estadístico inferencial de los modelos polinómicos (3) y (4), aplicando las técnicas proporcionadas por la teoría estadística de la regresión multivariable.

2. ANÁLISIS ESTADISTICO

Cada una de las expresiones de (2) puede considerarse compuesta de dos partes, una primera parte polinómica, en general de grado m, denominada modelo, y una segunda parte denominada perturbación.

Los modelos polinómicos reciben también el nombre de modelos lineales porque lo son respecto a los parámetros a_0, a_1,..., a_k y b_0, b_1,..., b_k y no porque las relaciones entre las variables x con u y v, e y con u, v sean lineales. En el caso concreto de (2) bastaría con hacer un sencillo cambio de variables para que los modelos de (2) se convirtieran en dos formas lineales de cinco variables, tales como

$$x = a_0+a_1u_1+a_2u_2+a_3u_3+a_4u_4+a_5u_5+\epsilon_x$$
$$y = b_0+b_1u_1+b_2u_2+b_3u_3+b_4u_4+b_5u_5+\epsilon_y$$

Para determinar los modelos sólo contamos con una muestra de puntos de control que se extienden de forma aproximadamente uniforme sobre la imagen y la zona correspondiente de mapa por lo que, como en el apartado anterior, sólo podemos determinar estimaciones de los modelos, tales como $\hat{x}$ e $\hat{y}$. Sin embargo, para que dichas estimaciones posean un valor inferencial, esto es, para que nos permitan sacar conclusiones sobre la totalidad de la imagen a partir de la muestra de puntos de control, son necesarias las siguientes condiciones fundamentales:

1) El modelo ha de ser lineal.
2) Las perturbaciones han de distribuirse normalmente (Normalidad).
3) La varianza de la perturbación ha de ser constante, σ^2 Homocedasticidad).
4) Las perturbaciones deben estar incorreladas entre si (Independencia).

2.1. Estimación del modelo y de su precisión.

Para estimar los coeficientes de cada uno de los cuatro modelos aplicaremos el método de máxima verosimilitud el cual, al ser normales las perturbaciones, coincide con el método de mínimos cuadrados. Así pues, si la variable dependiente o de respuesta fuera y, y las variables independientes, explicativas o regresores fueran $u_1, u_2,..., u_k$ la solución, en forma matricial, sería (Peña, 1989).

$$\hat{b} = (U'U)^{-1} U' Y \qquad (5)$$

donde

$$U = \begin{bmatrix} 1 \; u_{11} \; u_{21} \; .... \; u_{k1} \\ 1 \; u_{12} \; u_{22} \; .... \; u_{k2} \\ . \; .... \; ..., \\ . \; .... \; ... \\ 1 \; u_{1n} \; u_{2n} \; .... \; u_{kn} \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix} \qquad \hat{b} = \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ b_k \end{bmatrix}$$

Las componentes del vector $\hat{b}$ proporcionan la estimación puntual de los coeficientes del modelo. Además se demuestra (Peña, 1989) que, bajo las condiciones impuestas al modelo, cada b_j posee una distribución normal de media b_j y varianza σ^2 q_{jj},siendo q_{jj} los elementos de la diagonal de la matriz $(U' U)^{-1}$.

La estimación por intervalos de cada uno de los coeficientes ha de hacerse determinando sus respectivos intervalos de confianza, los cuales se pueden hallar sin más que tener en cuenta que la variable aleatoria

$$t = \frac{\hat{b}_j - b_j}{S_R\sqrt{q_{jj}}}$$

con

$$S_R = \sqrt{\frac{\sum_1^n e^2_{yi}}{n-k-1}}$$

posee una distribución **t** de Student de (n-k-1) grados de libertad (Peña, 1989).

De esta manera el intervalo de confianza, de grado $(1-\alpha)$, del coeficiente b_j será

$$(\hat{b}_j \pm t_{(n-k-1),\alpha/2}\, S_R\sqrt{q_{jj}}) \qquad (6)$$

La precisión del ajuste del modelo debería medirse mediante σ^2 (varianza de las perturbaciones). Sin embargo, al ser este valor desconocido usaremos como estimador de la misma la magnitud

$$S^2_R = \frac{\sum_1^n e^2_{yi}}{n-k-1}$$

por ser éste un estimador insesgado de (4).

La estimación por intervalos σ^2, puede hacerse teniendo en cuenta que la variable aleatoria

$$\chi^2 = \frac{(n-k-1)S^2_R}{\sigma^2}$$

posee una distribución χ^2 con (n-k-1) grados de libertad (Peña, 1989), por lo tanto el intervalo de confianza correspondiente a $(1-\alpha)$ será

$$\frac{(n-k-1)S^2_R}{\chi^2_a} \leq \sigma^2 \leq \frac{(n-k-1)S^2_R}{\chi^2_b} \qquad (7)$$

siendo χ^2_a y χ^2_b los extremos superior e inferior de un intervalo que, para la distribución $\chi^2_{(n-k-1)}$, dejen cada uno una probabilidad de $\alpha/2$ fuera de dicho intervalo.

2.2. Comprobación de las condiciones de validez del modelo.

Cada uno de los modelos lineales ha sido determinado bajo la suposición de que se cumplen las condiciones de aplicación del modelo de regresión. Las consecuencias que tendría el incumplimiento de las mismas son diversas (Peña, 1989). En nuestro caso podría decirse de forma resumida que, además de la pérdida de eficiencia (mínima varianza) de los estimadores y posibles sesgos en las estimaciones, los contrastes de hipótesis y las expresiones de los intervalos de confianza dejan de tener validez, tal como ocurre con la falta de normalidad. Así pues, el valor del modelo sería puramente descriptivo y las consecuencias que de él pudieran extraerse quedarían restringidas al conjunto de los puntos de control.

Habitualmente, no se conoce con anterioridad si se cumplen tales condiciones, por esta razón la fase de verificación de los modelos se realiza una vez establecida la ecuación de regresión. El análisis de la distribución de los residuos permite detectar si se han transgredido algunos de los supuestos. Cada uno de los modelos lineales asume que los errores ϵ_i son variables aleatorias independientes y normales de medio cero y varianza constante, por lo tanto si el modelo es adecuado para describir los datos, los residuos observados e_i deberán presentar una estructura similar, es decir, deberán cumplir tales condiciones.

La detección del cumplimiento de las condiciones y el tratamiento adecuado para su corrección, constituyen la diagnosis del modelo. Un estudio más detallado sobre este tema, puede verse en (Peña, 1989).

2.3. Simplificación y selección del modelo óptimo.

Una vez comprobadas las hipótesis del modelo es necesario averiguar si pueden eliminarse variables (en nuestro caso términos polinómicos) irrelevantes, las cuales podrían producir errores de especificación, en cuyo caso, pese a que los estimadores de los coeficientes y de la precisión del modelo siguen siendo centrados, la varianza de los primeros puede aumentar perdiéndose la eficiencia (Peña, 1989).

En nuestro caso, tal como más adelante veremos, los términos irrelevantes pueden ser, por el orden de magnitud de sus coeficientes, los términos de mayor grado. Para averiguar si un coeficiente o un grupo de coeficientes posee un valor significativamente bajo, se recurre a los contrastes de hipótesis individuales o conjuntos de los coeficientes.

2.3.1. Contrastes individuales y conjuntos de los coeficientes.

Para comprobar estadísticamente si un determinado coeficiente de un modelo, por ejemplo: el coeficiente b_j de (2), es o no cero, se recurre al siguiente contraste de hipótesis

H_0:b_j = 0 (hipótesis nula)
H_1:b_j = 0 (hipótesis alternativa)

en el que el estadístico de contraste, cuando la hipótesis nula es cierta, posee una distribución **t** de Student con (n-k-1) grados de libertad (Peña, 1989).

La hipótesis nula se rechazará cuando, fijado un nivel de significación, el valor observado del estadístico sea significativamente alto.

Cuando este estudio queremos extenderlo a un grupo de r coeficientes, el contraste es

H_0:b_r (las r componentes de b_r son 0)
H_1:b_r (alguna de las componentes de b_r no es 0)

siendo b_r un vector cuyas r componentes son los coeficientes del grupo que quiere estudiarse.

El estadístico del contraste posee, si la hipótesis nula es cierta, una distribución F de Snedecor con r y (n-k-1) grados de libertad (Peña, 1989), realzándose el contraste, por lo demás, como en el caso anterior.

2.3.2. Selección del modelo.

Cuando se dispone de un modelo con varias variables, es preciso deficir un procedimiento operativo de reducción para construir un modelo final de regresión.

Existen tres métodos distintos para ello (Peña, 1989):

- Eliminación progresiva de variables.
- Introducción progresiva de variables.
- Regresión paso a paso.

En nuestro caso (Ardizone, 1990) se ha aplicado el primero, el cual consiste en emplear, en principio, todas las variables potencialmente influyentes; a continuación se calculan los estadísticos para cada coeficiente. Si alguno de estos valores es menor que un valor prefijado, se elimina dicha variable, repitiéndose de nuevo la regresión con las k-1 variables restantes. Este procedimiento posee el inconveniente de su excesivo cálculo. En contrapartida es excelente para evitar la exclusión de alguna variable significativa, por lo que se suele aplicar con frecuencia.

Los procedimientos anteriores definen un conjunto de modelos posibles, el problema entonces consiste en seleccionar el más adecuado.

El criterio elegido para la selección del modelo es el del estadístico C_p de Mallows, el cual consiste en elegir aquel cuyo número de parámetros p minimiza el error cuadrático medio de predicción en los puntos observados:

$$\min \left[\frac{1}{\sigma^2} \sum_1^n E\,(\hat{y}_p(i)-m_i)^2 \right]$$

siendo m_i la esperanza matemática condicionada de y en el punto de observación.

Se demuestra (Peña, 1989) que este criterio equivale a minimizar

$$C_P = p + (n-p) \left[\frac{S^2_R(P) - S^2_R(k+1)}{S^2_R\,(k+1)} \right] \qquad (8)$$

donde $S^2_R\,(k+1)$ es la varianza residual del modelo con todas las k variables, S^2_R (P) es la del modelo con p-1 variables, p es el número de parámetros y n es el número total de datos.

2.4. Estudio de variables omitidas.

La influencia de una variable no incluida puede detectarse dibujando respecto a ella los residuos, por ejemplo: e_y = f(w), dado que ellos contienen información ajena a las variables independientes por estar incorrelacionados con ellas (Peña, 1989).

El estudio puede completarse determinando la recta de ajuste definida por la variable residuo y la nueva variable y, mediante un análisis de la varianza, comprobar si existe alguna tendencia lineal significativa entre ambas variables.

3. APLICACIÓN AL ESTUDIO DE DOS CORRECCIONES GEOMÉTRICAS

Los fundamentos teóricos expuestos en el apartado anterior se han aplicado a la corrección geométrica de dos cuartos de escena LANDSAT 5 (sensor TM): 205-32-IV y 200-34-IV. El primero cubre una extensión aproximada de 93 Km por 93 Km, incluyendo una zona de Castilla-La Mancha próxima a Aranjuez, mientras que el segundo, de la misma superficie, incluye una zona de Andalucía próxima a Guadix.

Los datos de partida son las coordenadas de las parejas de los puntos de control, determinadas en el Area de Teledetección del Instituto Geográfico Nacional, para la corrección geométrica de ambos cuartos de escena.

El estudio realizado se compone de las siguientes partes:

- Preparación de los datos.
- Estimación de los modelos de regresión y de sus precisiones.
- Comprobación de las hipótesis de los modelos.
- Simplificación y selección del modelo.
- Estudio de variables omitidas.

3.1. Preparación de los datos.

Las coordenadas mapa (u,v) (coordenadas UTM) han sido transformadas para referirlas al mismo sistema de referencia que las coordenadas imagen (x,y) y que pasen a las mismas unidades. Dicha transformación se compone simplemente de una traslación, una simetría axial y un cambio de unidades. Las altitudes z también han sido transformadas según el siguiente cambio de variable

$$w = (z_i - z_0)/D \qquad (9)$$

donde z_0 es la altitud del punto de control más bajo y D es el tamaño de remuestreo de la imagen (en nuestro caso 25 m). Las Tablas 1 y 2 muestran las coordenadas mapa, una vez transformadas, y las coordenadas imagen de los puntos de control asi como la altitud modificada W de ambas imágenes. Además, para linealizar los modelos polinómicos, inicialmente de segundo grado, se han realizado los siguientes cambios de variables

$$x^2 = X^2 \qquad u^2 = U^2$$
$$y^2 = Y^2 \qquad v^2 = V^2$$
$$xy = XY \qquad uv = UV$$
$$x = X \qquad u = U$$
$$y = Y \qquad v = V$$

3.2. Estimación del modelo y de su precisión.

Partiendo de los datos anteriores se han estimado los coeficientes de los modelos para la transformación inversa (de mapa a imagen) y directa (de imagen a mapa) de ambas imágenes, asi como los intervalos del 95% de confianza de los mismos (apartado 2.1). En las Tablas 3, 4, 5 y 6 se presentan únicamente los resultados referentes a la transformación inversa (de mapa a imagen) de ambas imágenes, un estudio más detallado puede verse en Ardizone, 1990.

La estimación puntual y por intervalos del 95% de confianza de la varianza del modelo (apartado 2.1), que es considerada como medida de precisión del ajuste, de cada imagen se muestra a continuación:

Imagen TM 205-32-IV (ARANJUEZ)

Modelo	S^2_R	Extr. Inf.	Extr. Sup.	(S^2_R)
X(U,V)	0,71	0,39	1,63	0,84
Y(U,V)	1,01	0,56	2,34	1,00
U(X,Y)	1,02	0,57	2,37	1,01
V(X,Y)	1,45	0,81	3,36	1,20

Imagen TM 200-34-IV (GUADIX)

Modelo	S^2_R	Extr. Inf.	Extr. Sup.	(S^2_R)
X(U,V)	1,20	0,65	2,90	1,09
Y(U,V)	0,30	0,17	0,73	0,55
U(X,Y)	1,66	0,90	3,98	1,29
V(X,Y)	0,50	0,27	1,20	0,71

Para la primera imagen de forma aproximada podríamos decir que la desviación típica se encuentra, en el caso de X, entre 0,6 y 1,3 y, en el caso de Y, entre 0,7 y 1,5; mientras que en la segunda imagen la desviación típica de las perturbaciones estaría, en el caso de X, entre 0,8 y 1,7 y, en el caso de Y, entre 0,4 y 0,8.

En las Tablas 1 y 2 se presentan las estimaciones de los valores de cada modelo en los puntos de control, así como los valores de los residuos.

3.3. Comprobación de las hipótesis del modelo.

Antes de comprobar las hipótesis del modelo, es necesario detectar si existen valores de los residuos que sean atípicos, ya que su presencia podría alterar las estimaciones e incluso afectar a dichas condiciones. Así, por ejemplo, la falta de normalidad puede deberse a la presencia de observaciones atípicas y en consecuencia de los estimadores dejarían de ser eficientes. Además, los contrastes **t** y **F** sólo podrían usarse de forma aproximada.

Para detectar la presencia de residuos atípicos, posiblemente debidos a errores de medida o a la codificación de los datos, se examinan los residuos para ver si alguno de ellos sobrepasa un margen de 3 S_R alrededor del 0, en cuyo caso se consideraría atípico.

De la comparación de los residuos con los valores de S_R correspondientes, no se puede afirmar que tales observaciones existan y por ello pueden considerarse homogéneos los conjuntos de residuos.

Para completar el estudio, se realiza un contraste de observaciones atípicas usando como estadístico de contraste el coeficiente de apuntamiento. La hipótesis de homogeneidad se rechazaría cuando el valor del coeficiente fuera significativamente alto respecto al de una distribución normal (Peña, 1989).

La comprobación de las condiciones de validez se ha realizado (Ardizone, 1990) utilizando los siguientes procedimientos gráficos y contrastes de hipótesis:

LINEALIDAD: Gráfico de residuos frente a valores estimados de la variable dependiente (Peña, 1989).

NORMALIDAD: Gráfico de probabilidad normal (Peña, 1989) y los contrastes de Kolmogorov-Smirnov (Peña, 1989; Ruíz, 1986; Sachs, 1978), de D'Agostino (Sachs, 1978) y de Saphiro y Wilk (Ruíz, 1986).

HOMOCEDASTICIDAD: Gráfico de residuos frente a valores estimados de la variable dependiente y frente a valores de cada variable independiente. Cuando los resultados no son concluyentes (Ardizone, 1990), se aplica el contraste de homogeneidad de Bartlet (Ruíz, 1986).

INDEPENDENCIA: Contraste de rachas (Sachs, 1978) y de Durbin-Watson (Peña, 1989).

Todos los contrastes se han realizado fijando un nivel de significación del 5%, de sus resultados y de los gráficos se deduce que pueden aceptarse las condiciones de validez con el 95% de confianza (Ardizone, 1990).

3.4. Simplificación y selección del modelo.

Las Tablas 3, 4, 5 y 6 proporcionan los contrastes individuales (t) de los coeficientes. En las dos primeras se aprecia que, para un nivel $\alpha = 5\%$, la influencia de las variables U2 y V2 no es significativa en los modelos de corrección correspondientes a la imagen de Aranjuez. Además, sólo es significativa la influencia de la variable UV en el modelo X, no siéndolo en el Y.

Por otro lado, tal como puede verse en las tablas 3 y 4, los intervalos de los coeficientes de las variaables que no poseen influencia significativa contienen en su interior el valor cero, lo cual confirma la posibilidad de que realmente puedan ser nulos.

A iguales conclusiones se llega en el caso de la transformación de mapa a imagen (Ardizone, 1990).

En el caso de la imagen de Guadix los resultados son bastante diferentes, teniéndose que rechazar la hipótesis nula, al nivel $\alpha = 5\%$, en X e Y, por lo que en estos casos consideramos todos los coeficientes cuadráticos distintos de cero.

Los contrastes por grupos de coeficientes (apartado 2.3.1) en la imagen de Aranjuez nos permiten comprobar que los términos que no poseen influencia individual, tampoco la poseen de forma conjunta (Ardizone, 1990).

Además de averiguar las posibles simplificaciones de cada modelo de regresión es necesario comprobar sus condiciones de validez, estas han sido verificadas en la imagen de Aranjuez, comprobándose que se cumplen con un nivel de confianza del 95 %.

Para seleccionar el modelo más apropiado, en el caso de la imagen de Aranjuez, se ha aplicado el criterio del estadístico C_p de Mallows (8), calculándose su valor para los modelos que a continuación se relacionan:

MODELOS	C_p
X(U,V)(C.C)	6
X(U,V)(C.I)	2,47
X(U,V)(L)	9,099
Y(U,V)(C.C)	6
Y(U,V)(C.I)	4,096
Y(U,V)(L)	3,029
U(X,Y)(C.C)	6
U(X,Y)(C.I)	2,43
U(X,Y)(L)	6,95
V(X,Y)(C.C)	6
V(X,Y)(C.I)	5,35
V(X,Y)(L)	5,017

C.C = Modelo cuadrático completo
C.I = Modelo cuadrático incompleto (el único término cuadrático es el producto cruzado).
L = Modelo lineal.

Tal como puede observarse, el valor mínimo del estadístico de Mallows lo poseen los modelos X e U en forma cuadrática completa e Y y V en forma lineal, lo cual confirma los resultados obtenidos, mediante contrastes individuales y por grupos de coeficientes. Sin embargo, dado que, al reducir de forma cuadrática completa a lineal los modelos Y y V, la disminución del estadístico C_p es pequeña, adoptamos como modelo definitivo en los cuatro casos la forma cuadrática incompleta (Ardizone, 1990), figurando la estimación de la precisión en la siguiente tabla:

Modelo	S^2_R	Extr. Inf.	Extr. Sup.	(S_R)
X(U,V)	0,65	0,37	1,41	0,80
Y(U,V)	1,02	0,58	2,23	1,01
U(X,Y)	0,93	0,53	2,04	0,96
V(X,Y)	1,56	0,89	3,41	1,25

(intervalos de confianza del 95% para S^2_R)

3.5. Estudio de variables omitidas.

Es bien conocida la influencia del relieve del terreno en la precisión de la corrección geométrica de una imagen de satélite. Al no haberse incluido, por el momento, este factor, es necesario hacerlo para averiguar su influencia. Por ello, se ha realizado un nuevo estudio, usando como nueva variable la altitud.

Los datos que se han empleado son las altitudes de los puntos de control de las imágenes estudiadas, que en la imagen de Aranjuez oscilan entre 480 m y 775 m, y en la imagen de Guadix lo hacen entre 10 m y 1810 m.

Dado que los residuos están incorrelados con las variables explicativas, contendrán la información de las variables que no han sido incluidas por ahora. Por ello, se ha realizado una regresión lineal entre los residuos de cada uno de los modelos seleccionados en el apartado anterior y la altitud transformada (9).

Los resultados en el caso de la imagen de Guadix pueden verse en las Tablas 7 y 8.

Como se desprende de los valores del coeficiente de determinación, la explicación de las variaciones de los residuos frente a las de altitud es muy baja (4,4% en el caso de e_x y 0,24% en el caso de e_y). Por otro lado los contrastes de Student para la pendiente de ambas rectas indican que para un nivel $\alpha = 5\%$, no podemos decir que sean distintos de cero. Sin embargo, viendo los valores tan bajos del estadístico t podemos considerarlos nulos.

Al mismo resultado se llega en el caso de la imagen de Aranjuez .

No obstante, con la imagen de Guadix se ha realizado una prueba más, la cual consiste en eliminar del modelo multivariable todos los términos cuadráticos, quedándonos únicamente con los lineales, con lo que lógicamente la varianza residual aumentará. Sin embargo, al estudiar los residuos correspondientes frente a W (Tablas 9 y 10) los coeficientes de determinación aumentan considerablemente (hasta 18,25% en el caso de e_x y hasta el 17,14% en el caso de e_y). Además, aunque los contrastes de hipótesis de las pendientes de ambas rectas no dan un rechazo estricto para un nivel α del 5%, el valor del estadístico t se encuentra muy próximo a los límites de la región crítica o de rechazo (aproximadamente $t>2$, $t<-2$) e incluso puede rechazarse la hipótesis nula si aumentamos el valor de α al 10%.

A pesar de todo, la porción de varianza de los residuos que aun queda por explicar es todavía muy grande y se debe, en su mayor parte, a las variables cuadráticas suprimidas inicialmente.

Un estudio análogo se ha seguido con la imagen de Aranjuez, pero en este caso puede concluirse que, para niveles del 5% o del 10%, la variable W no presenta una influencia significativa en el modelo, por lo que su presencia en el mismo no lo mejora en precisión. Probablemente este resultado sea debido a que la variación de altitudes en este caso es mucho menor que en el de la zona de Guadix.

4. CONCLUSIONES

Pese a que en el apartado 3 ya se han expuesto los resultados y conclusiones más importantes de este trabajo, presentaremos aquí un resumen de las mismas.

1) En este estudio se establece una metodología rigurosa para analizar, desde el punto de vista estadístico, la corrección geométrica de imágenes de satélite, ya que en él además de estimarse las transformaciones necesarias, se comprueban sus condiciones de validez y se estiman las precisiones de los modelos. Este último aspecto es importante dada la aplicación cartográfica que poseen las imágenes de satélite.

2) El modelo más sencillo que explica de forma más rigurosa la transformación inversa resulta ser la forma cuadrática incompleta, en el caso de la imagen de Aranjuez, y el cuadrático completo en la imagen de Guadix.

3) En la imagen de Aranjuez la variación de los residuos de la transformación de imagen a mapa no queda explicada, desde el punto de vista estadístico, por la variación de la altitud. Mientras que en la imagen de Guadix, las variaciones de la altitud explican cerca del 20% de la variación de los residuos de la transformación en forma lineal, el resto se debe, en su mayor parte, a los términos cuadráticos que faltan para completar este modelo.

Tabla 1

N	M	X	Y	e_X	e_Y	H	$\hat{X}$	$\hat{Y}$
1239	511	472	346	-0.68789	1.23888	5.6	472.688	344.7610
1856	745	1016	435	-0.31838	0.38719	1.6	1016.320	434.6130
2129	341	1174	59	0.30881	-1.07032	2.3	1173.690	60.0703
3653	929	2519	291	0.84446	1.00607	3.2	2518.160	289.9930
1662	1548	988	1123	-1.30948	0.68706	0.4	989.309	1122.3100
2514	1512	1681	951	0.22336	-1.93681	1.6	1680.780	952.9370
3279	1206	2258	578	0.93498	0.54807	2.4	2257.070	577.4520
4057	1200	2892	444	-1.27765	-0.85027	8.6	2893.280	444.8500
853	1354	297	1096	1.66147	-0.69427	3.2	295.339	1096.6900
989	2259	229	1880	0.48247	0.05995	0.0	228.518	1879.9400
1294	1930	750	1494	-0.99130	-0.98039	1.6	750.991	1494.9800
3847	2137	2874	1245	-0.22534	0.64779	11.8	2874.230	1244.3500
614	2916	357	2413	-0.40091	-0.29610	10.7	357.401	2413.3000
1735	2704	1239	2057	-0.08570	1.74673	5.2	1239.090	2055.2500
2106	2976	1588	2217	0.76373	0.39044	6.8	1587.240	2216.6100
3811	2826	2958	1813	0.62745	-0.20437	10.2	2957.370	1813.2000
795	3315	971	2710	-0.24411	-0.00783	11.2	971.244	2710.0100
1390	3558	1098	2811	0.32686	0.20240	9.8	1097.670	2810.8000
2631	3438	2092	2507	-0.47539	-0.99305	7.6	2092.480	2507.9900
3351	3510	2693	2449	-0.06917	0.80377	8.0	2693.070	2448.2000
2196	3870	1808	2933	0.12496	-0.21167	8.0	1807.880	2933.2100
3636	3547	2932	2431	-0.21324	-0.47332	8.3	2932.210	2431.4700

N	M	X	Y	H	e_N	e_M	$\hat{N}$	$\hat{M}$
1239	511	472	346	5.6	1.07885	-1.29721	1238.030	512.297
1856	745	1016	435	1.6	0.46830	-0.39249	1855.600	745.382
2129	341	1174	59	2.3	-0.61403	1.18430	2129.890	339.816
3653	929	2519	291	3.2	-0.73381	-1.38799	3653.880	930.388
1662	1548	988	1123	0.4	1.70031	-0.49928	1660.130	1548.500
2514	1512	1681	951	1.6	-0.71750	2.22598	2514.570	1509.770
3279	1205	2259	578	2.4	-0.96764	-0.86757	3279.950	1206.870
4097	1200	2892	444	8.6	1.50104	1.30274	4095.770	1198.700
853	1354	297	1096	3.2	-2.11744	0.42580	854.998	1353.570
989	2259	229	1880	0.0	-0.55467	-0.18306	889.426	2259.180
1294	1930	780	1494	1.6	0.93222	1.38717	1292.870	1929.610
3847	2137	2874	1245	11.8	0.41726	-0.70748	3846.420	2137.710
614	2916	357	2413	10.7	0.39949	0.44384	613.614	2915.560
1735	2704	1239	2057	5.2	0.51399	-2.03209	1734.340	2706.030
2106	2976	1588	2217	6.8	-0.80465	-0.64008	2106.700	2976.640
3811	2826	2958	1813	10.2	-0.78506	0.09419	3811.640	2825.910
795	3315	971	2710	11.2	0.20419	0.06804	794.858	3314.930
1390	3558	1098	2811	9.8	-0.35631	-0.31489	1390.310	3558.310
2631	3438	2092	2507	7.8	0.32486	1.28023	2630.690	3436.720
3351	3510	2693	2449	8.0	0.27086	-0.92696	3350.750	3510.930
2196	3870	1808	2933	8.0	-0.19640	0.21899	2196.450	3869.780
3636	3547	2932	2431	8.3	0.13914	0.60892	3635.900	3546.390

Tabla 2

X	Y	N	M	H	$\hat{X}$	e_X	$\hat{Y}$	e_Y
246	256	977	377	32.4	246.0154545	-0.015455	255.894779	0.105221
973	159	1855	422	45.2	973.3322275	-0.332227	159.095191	-0.095191
1955	307	2979	810	34.4	1956.2149983	-1.214998	306.697912	0.302088
2935	118	4171	799	17.0	2935.0612079	-0.061208	118.600162	-0.600162
1456	953	2247	1464	56.0	1454.5975561	1.402434	953.290781	-0.290781
2142	691	3112	1303	64.4	2140.5797159	1.420286	690.483719	0.516281
1232	1676	1828	2268	49.6	1232.2630700	-0.263070	1676.652479	-0.652479
1679	1310	2435	1935	47.6	1679.7940180	-0.794018	1309.446119	0.553881
1017	1164	2629	1791	32.8	1017.6410843	-0.641084	1163.560015	0.439985
2730	1378	3659	2242	21.2	2731.3591760	-1.359176	1378.051060	-0.051060
300	2066	644	2523	66.8	300.7197664	-0.719766	2067.154578	-1.154570
433	2300	750	2826	36.4	432.8774512	0.122549	2299.921070	0.078930
586	2343	921	2910	39.6	585.3722121	0.627789	2342.744836	0.255164
3086	2002	3943	3084	2.8	3085.0637326	0.936267	2002.001319	-0.001319
93	2727	257	3252	30.8	93.1578125	-0.157812	2726.094452	0.905548
358	2793	555	3309	15.2	357.1819876	0.818012	2792.945255	0.054745
2253	2325	2892	3254	1.0	2251.5484919	1.451508	2325.082343	-0.082343
2663	2167	3412	3157	0.0	2664.0623440	-1.062344	2166.684881	0.315119
1736	805	2609	1350	72.0	1734.4918197	1.508180	804.990710	0.009282
1103	2031	1600	2657	35.2	1103.6551049	-0.655105	2030.924251	0.075749
1346	2703	1744	3502	2.4	1347.0098808	-1.009881	2703.684080	-0.684080

X	Y	N	M	H	$\hat{N}$	e_N	$\hat{M}$	e_M
246	226	977	377	32.4	976.9551057	0.044014	377.120339	-0.1205392
973	159	1855	422	45.2	1854.6350698	0.364930	421.815146	0.1848557
1955	307	2979	810	34.4	2977.5075645	1.492435	810.093027	-0.0930272
2935	118	4171	799	17.0	4171.0459920	-0.045993	798.270316	0.7216541
1456	953	2247	1464	56.0	2248.7198264	-1.719826	1463.962628	0.0373723
2142	691	3112	1303	64.4	3113.3610176	-1.361018	1303.915779	-0.9157791
1232	1676	1828	2260	49.6	1827.8307008	0.169299	2257.175307	0.0246912
1679	1310	2435	1935	47.6	2433.9456878	1.054512	1933.479774	-0.4797735
1017	1164	2629	1791	32.8	2628.1918447	0.840185	1791.378833	-0.3788831
2730	1378	3659	2242	21.2	3657.4090119	1.590986	2241.644953	0.3550474
300	2066	644	2523	66.0	643.3908995	0.601101	2521.489402	1.5105184
433	2390	750	2826	36.4	750.1275020	-0.127502	2826.123213	-0.1232129
586	2343	921	2910	39.6	921.6860348	-0.686035	2910.439904	-0.4399040
3086	2002	3945	3054	2.8	3744.1045654	-1.104565	3054.202487	-0.2024872
93	2727	257	3252	30.8	256.6235883	0.376412	3203.029362	-1.0273628
358	2793	555	3309	15.2	555.9574340	-0.957434	3309.240349	-0.2403489
2253	2320	2092	3254	1.0	2093.7260418	-1.726042	3254.217593	-0.2175932
2663	2167	3412	3157	0.0	3410.6798019	1.320118	3157.141087	-0.1418974
1736	805	2604	1350	72.0	2610.7779895	-1.777969	1350.330062	-0.3300123
1103	2031	1600	2657	35.2	1599.2088177	0.791182	2656.946419	0.0535012
1346	2703	1744	3502	2.4	1742.9467232	1.053277	3500.966288	1.0337124

Tabla 3

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependent Variable.. X

------------------------ Variables in the Equation ------------------------

Variable	B	SE B	95% Confdnce Intrvl B		T	Sig T
V2	1.192132E-07	2.05444E-07	-3.16308E-07	5.547346E-07	.580	.5698
U	.81882	1.04999E-03	.81660	.82105	779.837	.0000
UV	-4.92704E-07	1.75679E-07	-8.65127E-07	-1.20282E-07	-2.805	.0127
U2	1.069223E-07	1.91314E-07	-2.98645E-07	5.124895E-07	.559	.5840
V	.16474	9.39977E-04	.16274	.16673	175.254	.0000
(Constant)	-625.89435	1.66428	-629.42246	-622.36623	-376.075	0.0

Tabla 4

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependent Variable.. Y

------------------------ Variables in the Equation ------------------------

Variable	B	SE B	95% Confdnce Intrvl B		T	Sig T
V2	3.365132E-07	2.46414E-07	-1.85841E-07	8.588870E-07	1.366	.1909
U	-.16387	1.25938E-03	-.16654	-.16120	-130.118	0.0
UV	-3.02122E-07	2.10713E-07	-7.48914E-07	1.445699E-07	-1.434	.1709
U2	1.194132E-08	2.29466E-07	-4.74505E-07	4.983874E-07	.052	.9591
V	.81651	1.12743E-03	.81412	.81890	724.224	.0000
(Constant)	130.64268	1.99617	126.41098	134.87438	65.447	.0000

Tabla 5

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependent Variable.. X

------------------------ Variables in the Equation ------------------------

Variable	B	SE B	95% Confdnce Intrvl B		T	Sig T
V2	-1.14632E-06	3.16849E-07	-1.82067E-06	-4.69974E-07	-3.618	.0025
U2	5.476383E-07	2.25601E-07	6.635570E-08	1.028921E-06	2.428	.0284
UV	-5.65632E-07	2.37451E-07	-1.07175E-06	-5.95167E-08	-2.382	.0309
V	.15639	1.22790E-03	.15377	.15900	127.361	.0000
U	.81913	1.36845E-03	.81621	.82204	598.579	.0000
(Constant)	-613.38106	1.96403	-617.56729	-605.19483	-312.307	.0000

Tabla 6

**** MULTIPLE REGRESSION ****

Equation Number 1 Dependent Variable.. Y

------------------------ Variables in the Equation ------------------------

Variable	B	SE B	95% Confdnce Intrvl B		T	Sig T
V2	6.582640E-07	1.59494E-07	3.183103E-07	9.982178E-07	4.127	.0009
U2	4.801639E-07	1.13668E-07	2.379175E-07	7.224804E-07	4.225	.0007
UV	-2.63285E-07	1.19527E-07	-5.18051E-07	-8.51803E-09	-2.203	.0437
V	.81857	6.18098E-04	.81725	.81989	1324.334	.0000
U	-.15347	6.88847E-04	-.15493	-.15200	-222.790	.0000
(Constant)	96.77629	.98865	94.66903	98.88354	97.888	.0000

Tablas 7 y 8

Regression Analysis - Linear model: Y = a+bX **TABLA VII**

Dependent variable: C:GUADIX.RESX Independent variable: C:GUADIX.U

Parameter	Estimate	Standard Error	T Value	Prob. Level
Intercept	-0.304522	0.386334	-0.788235	0.440283
Slope	9.09926E-3	9.73507E-3	0.934686	0.361677

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	Prob. Level
Model	.7908743	1	.7908743	.8736419	.36169
Error	17.199965	19	.905261		
Total (Corr.)	17.990840	20			

Correlation Coefficient = 0.209656 R-squared = 4.40 percent
Stnd. Error of Est. = 0.951452

Regression Analysis - Linear model: Y = a+bX **TABLA VIII**

Dependent variable: C:GUADIX.RESY Independent variable: C:GUADIX.U

Parameter	Estimate	Standard Error	T Value	Prob. Level
Intercept	0.0357624	0.198654	0.180023	0.85904
Slope	-1.06665E-3	5.00581E-3	-0.213471	0.833233

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	Prob. Level
Model	.0109075	1	.0109075	.0455701	.83323
Error	4.547757	19	.239356		
Total (Corr.)	4.558664	20			

Correlation Coefficient = -0.0489151 R-squared = .24 percent
Stnd. Error of Est. = 0.48924

Tablas 9 y 10

Regression Analysis - Linear model: Y = a+bX **TABLA IX**

Dependent variable: C:GUADIX.RESXLIN Independent variable: C:GUADIX.U

Parameter	Estimate	Standard Error	T Value	Prob. Level
Intercept	-1.10869	0.638447	-1.73654	0.0986546
Slope	0.0331282	0.016088	2.05919	0.0534492

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	Prob. Level
Model	10.483147	1	10.483147	4.240263	.05345
Error	46.973459	19	2.472287		
Total (Corr.)	57.456606	20			

Correlation Coefficient = 0.427146 R-squared = 18.25 percent
Stnd. Error of Est. = 1.57235

Regression Analysis - Linear model: Y = a+bX **TABLA X**

Dependent variable: C:GUADIX.RESYLIN Independent variable: C:GUADIX.U

Parameter	Estimate	Standard Error	T Value	Prob. Level
Intercept	0.696033	0.416388	1.6716	0.110993
Slope	-0.0207978	0.0104924	-1.98218	0.0621117

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	Prob. Level
Model	4.1317132	1	4.1317132	3.9290228	.06211
Error	19.980172	19	1.051588		
Total (Corr.)	24.111885	20			

Correlation Coefficient = -0.413952 R-squared = 17.14 percent
Stnd. Error of Est. = 1.02547

5. BIBLIOGRAFIA

✔ ARDIZONE, J.A. (1990): *Análisis estadístico para la corrección geométrica de imágenes de satélite.* Memoria presentada para optar al grado de Master en Estadística por la Universidad Complutense de Madrid.

✔ BERNSTEIN, R. (1983): "Image Geometry and Rectification". *Manual of Remote Sensing*, Vol. I, pp. 873-921. American Society of Photogrametry.

✔ MATHER, P.M. (1987): *Computer processing of Remotely sensed images.* Ed. J. Wiley & Sons.

✔ PEÑA SANCHEZ DE RIVERA, D. (1989): *Estadística, modelos y métodos.* Vols. I y II, Ed. Alianza Universidad.

✔ RICHARDS, J.A. (1986): *Remote Sensing Digital. Image Analisys.* Ed. Springer-Verlag.

✔ RUIZ MAYA, L. (1986): *Métodos estadísticos de investigación.* Ed. Instituto Nacional de Estadística.

✔ SACHS, L. (1978): *Estadística aplicada.* Ed. Labor.