

Estimación de erosión de suelos utilizando sensores remotos y programación genética

Cesar Augusto Puente Montejano

UASLP, Facultad de Ingeniería, Dr. Manuel Nava # 8, Zona Universitaria poniente, C.P. 78290,
San Luis Potosí, S.L.P, México, Teléfono: 01 (444) 826-2330 ext. 2112
cesar.puente@uaslp.mx

Gustavo Olague Caballero

Carretera Ensenada-Tijuana No. 3918, Zona Playitas, C.P. 22860, Ensenada, B.C. México,
Teléfono: 01(646)175-05-00
olague@cicese.mx

Resumen

Los índices de vegetación (IVs) son ampliamente utilizados para extraer información de la vegetación a partir de imágenes satelitales. Los modelos de erosión, como la "Ecuación Universal Revisada de la Pérdida de Suelo" (RUSLE) usan IVs como insumo para estimar el factor de cobertura vegetal (C). El factor C es uno de los más importantes porque cuantifica la cobertura que actúa como capa protectora entre el suelo y los elementos atmosféricos. Sin embargo los IVs encontrados en el estado-del-arte arrojan pobres resultados, ya que la mayoría de éstos están diseñados para detectar vegetación verde y no vegetación seca; la cual es también un importante factor que contribuye al desempeño del factor C. El propósito de esta investigación es desarrollar un método basado en programación genética para sintetizar IVs que estén mejor correlacionados con el factor C. Los resultados experimentales ilustran la eficiencia de este método y su efecto en el cálculo de erosión en una zona geográfica real. Los índices sintetizados obtienen una mejor aproximación al factor C obtenido en campo que cuando se utilizan los índices reportados en el estado-del-arte.

Palabra(s) Clave(s): erosión por agua, índices de vegetación, programación genética, percepción remota, RUSLE.

1. Introducción

Cada vez existe un mayor interés por el aprovechamiento sostenible de los recursos y crece la atención por los cambios ambientales. Uno de estos cambios ambientales es la erosión del suelo. La erosión del suelo se define como el desplazamiento de suelo fértil, debido a la acción de un agente llamado erosivo [1]. La erosión causada por agua se debe a la precipitación de lluvia en zonas donde las pendientes del terreno son relativamente empinadas, la litología es suave o no consolidada y la vegetación está dispersa. Para mejorar la planeación ambiental y conservación de suelos, es importante la detección y monitoreo de las áreas sujetas a fenómenos erosivos.

Se han formulado distintos modelos para estimar erosión tanto empíricos [2, 3], como los basados en propiedades físicas formales [4, 5]. El método de estimación de erosión más utilizado es el de la Ecuación Universal de Pérdida de Suelo Revisada [3] (RUSLE, por sus siglas en inglés), el cual es un modelo empírico que calcula promedios anuales de erosión. RUSLE se compone por seis factores:

$$A = R \times K \times L \times S \times C \times P \quad (1)$$

donde A es el promedio de pérdida de suelo en $\text{Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$; R es el factor de lluvia-escorrentía que representa la energía erosiva medida en $\text{MJ} \cdot \text{mm} \cdot \text{km}^{-2} \cdot \text{h}^{-1} \cdot \text{año}^{-1}$; K es la tendencia del suelo a erosionarse ($\text{t} \cdot \text{h} \cdot \text{MJ}^{-1} \cdot \text{mm}^{-1}$). Los demás factores son factores de escalamiento adimensionales. L es el factor de longitud de la pendiente; S es el factor de inclinación de la pendiente; C es el factor de cobertura; y P es el factor de soporte a la conservación de suelos.

Uno de los parámetros más importantes del modelo RUSLE es el que estima la cobertura del suelo o factor C. La cobertura del suelo se define como las diferentes cubiertas naturales que protegen al suelo del impacto directo de los elementos

atmosféricos. Por ejemplo, en un bosque el follaje vegetal absorbe la mayor parte de la energía de impacto que posee una gota de agua al caer durante un período de lluvia; lo que evita la erosión masiva del suelo que sostiene dicho bosque.

Resulta sumamente costoso medir directamente en campo los datos necesarios para alimentar el modelo RUSLE a gran escala, por lo que se han utilizado técnicas de percepción remota que ayudan a extrapolar medidas de campo representativas sobre extensas áreas geográficas [6]. La *Percepción Remota* es la disciplina encargada de desarrollar herramientas para extraer la información con la que se estiman algunas de las propiedades biofísicas de todo aquello que se encuentre sobre la superficie terrestre. En particular, se han desarrollado varias metodologías para identificar la cobertura vegetal; como los métodos de clasificación espectral, por ejemplo en [7] o los llamados índices de vegetación, por ejemplo en [8]. Sin embargo, de acuerdo con [2], la información que requieren los modelos de erosión para medir la cobertura difiere de la información que proveen estos métodos; los cuales, generalmente han sido diseñados para otras tareas, e.g. clasificación de uso de suelo, productividad agrícola, etc. Por lo tanto, existe una importante necesidad para desarrollar IVs que estén específicamente diseñados para la estimación de la cobertura vegetal en el contexto de erosión.

En este trabajo de investigación se propone una nueva metodología basada en la *Programación Genética* (PG) con el objetivo de obtener una estimación más precisa de uno de los parámetros más importantes en el modelo de erosión RUSLE: el *factor C*. En este trabajo, se plantea el problema de calcular el factor de cobertura como un problema de optimización; donde el objetivo es encontrar el índice de vegetación que muestre una mejor correlación con datos de campo del factor C. De esta forma, el algoritmo basado en PG construye nuevos índices de vegetación mediante la recombinación iterativa de un conjunto de operadores numéricos y bandas espectrales pertenecientes a imágenes de satélite.

2. Desarrollo

En la Figura 1 se muestra un diagrama con la metodología desarrollada durante esta investigación. El objetivo de esta metodología es desarrollar un enfoque automático para crear índices de vegetación que estén bien correlacionados con el factor C de RUSLE. En esta sección se describen los pasos que conforman dicha metodología.

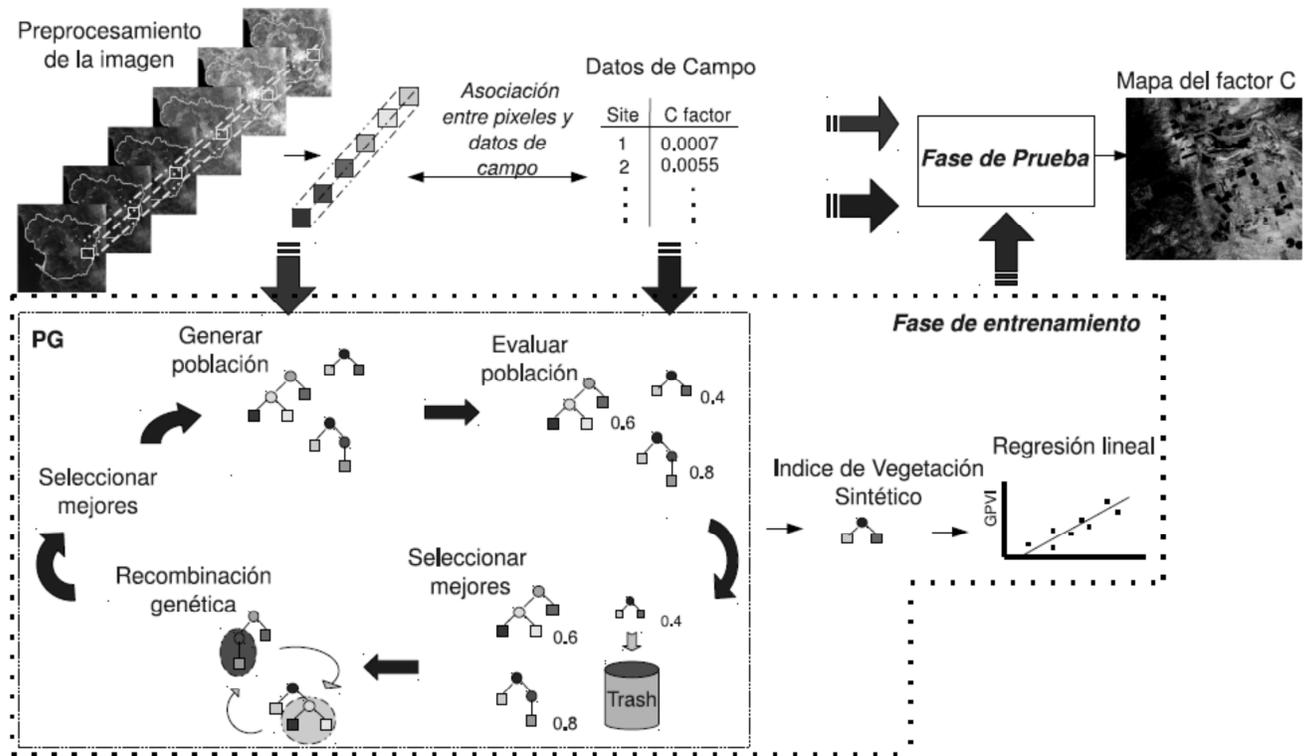


Fig. 1. Diagrama de flujo general de la metodología para estimar el factor C a partir de los IVs sintetizados por la PG.

2.1. Descripción de área de estudio

La cuenca Todos Santos se localiza al noroeste de la península de Baja California, México (ver Figura 2). Posee una topografía compleja cuya elevación oscila entre el nivel del mar y los 1876 metros sobre el nivel del mar. El promedio de pendientes es de 10°, aunque ~27% del área posee una entre los 10° y 20°; y el ~3% presenta una

inclinación de 30° o más. Las áreas aluviales (de 2°), ocupan aproximadamente el 19% de la cuenca. Existen dos grandes valles aluviales, conocidos como los valles agrícolas de Guadalupe y de Ojos Negros, En la planicie costera se encuentra asentada la mayor parte de la ciudad de Ensenada. Asimismo, destaca la presencia del valle agrícola de Maneadero al sur de dicha ciudad.

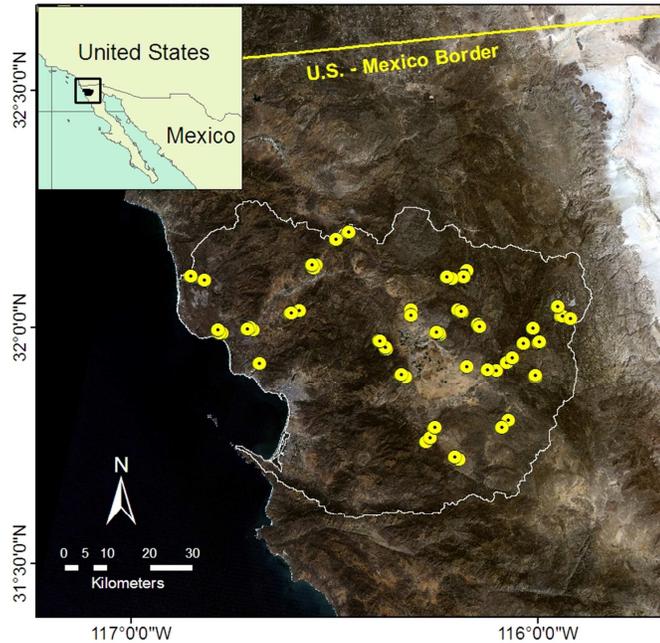


Fig. 2. Mapa de Baja California, México que muestra la localización de la cuenca Todos Santos. Los círculos amarillos representan los 67 sitios de muestreo de campo para obtener el factor C.

2.2. Muestreo de campo

El muestreo de campo implica las mediciones de diferentes parámetros para la estimación del factor C. Los autores de [9] identificaron los tres sub-factores principales que determinan la efectividad de la vegetación para limitar la erosión de suelo en ambientes áridos; a saber: *cobertura superficial del suelo* (vegetación a ras de piso: g), *cobertura vegetal aérea* (ramas suspendidas sobre el terreno: $p =$ porcentaje y $h =$ altura) y *cobertura residual* (rugosidad del suelo: r) y *subterránea* (raíces: b). Para la

campaña de muestreo de campo se realizó un análisis detallado del área de estudio con la ayuda de imágenes satelitales y mapas topográficos para identificar las zonas representativas de muestreo. La campaña se llevó a cabo entre Febrero y Mayo de 2007, adquiriendo un total de 67 sitios de muestreo (ver Figura 2). En cada uno de estos sitios se realizó un protocolo de muestreo basado en el método del transecto lineal descrito por [10, 11]. A partir de estas mediciones, el factor C para cada uno de los sitios de muestreo se define utilizando la siguiente ecuación, la cual fue derivada en el trabajo de [12] a partir de la Tabla 10 del manual del modelo RUSLE [9]:

$$C = 0.45(e^{-0.012b}) (1 - p e^{-0.328h}) (e^{-0.039g \frac{0.24}{r} 0.08}) \quad (2)$$

2.3. Pre-procesamiento de la imagen satelital

La imagen satelital utilizada para generalizar los sitios de campo en toda la cuenca pertenece al satélite Landsat 5 - TM. Cada imagen que el dispositivo Landsat TM entrega cubre aproximadamente 185 Km por lado y consta de siete bandas. Seis de ellas cubren la parte visible/infra-rojo cercano del espectro electromagnético y poseen una resolución espacial de 30 x 30 m por pixel. La banda restante cubre la parte del infra-rojo termal y tiene una resolución de 60 x 60 m (ver Tabla 1). Esta última banda no fue utilizada en este trabajo.

Banda	Rango espectral (µm)	Región espectral	Utilidad principal
1	0.45 - 0.52	Visible azul (B)	Penetración de cuerpos de agua
2	0.52 - 0.60	Visible verde (G)	Vigor de la vegetación sana
3	0.63 - 0.69	Visible rojo (R)	Absorción de clorofila
4	0.76 - 0.90	Infrarrojo cercano (NIR)	Contenido de biomasa
5	1.55 - 1.75	Térmico lejano (TIR)	Estrés de la vegetación
6	10.4 - 12.5	Infrarrojo de onda corta 1 (SWIR1)	Discriminación de rocas
7	2.08 - 2.35	Infrarrojo de onda corta 2 (SWIR2)	Absorción de agua y cantidad de hidroxilos

Tabla 1. Características de las bandas del sensor Landsat-TM.

La fecha de la imagen fue lo más cercana posible a la fecha del muestreo: el 13 de Abril de 2007. Para la aplicación que se le quiere dar en este trabajo, es necesario darle un pre-procesamiento a la imagen, el cual consiste en dos tipos de correcciones: geométricas y radiométricas. Las primeras tienen el objetivo de empatar un sistema de coordenadas de referencia con la imagen satelital; mientras que las segundas se llevan a cabo para disminuir los defectos por desgaste en el sensor y para eliminar influencias atmosféricas (ver Figura 3). El resultado de este pre-procesamiento es una nueva imagen, donde el valor numérico de cada pixel representa una magnitud física conocida como *reflectancia* (R); la cual se define como la relación entre la potencia electromagnética de la radiación incidente en una superficie con respecto a la potencia reflejada en la misma con un rango de $R: 0 \leq R \leq 1$

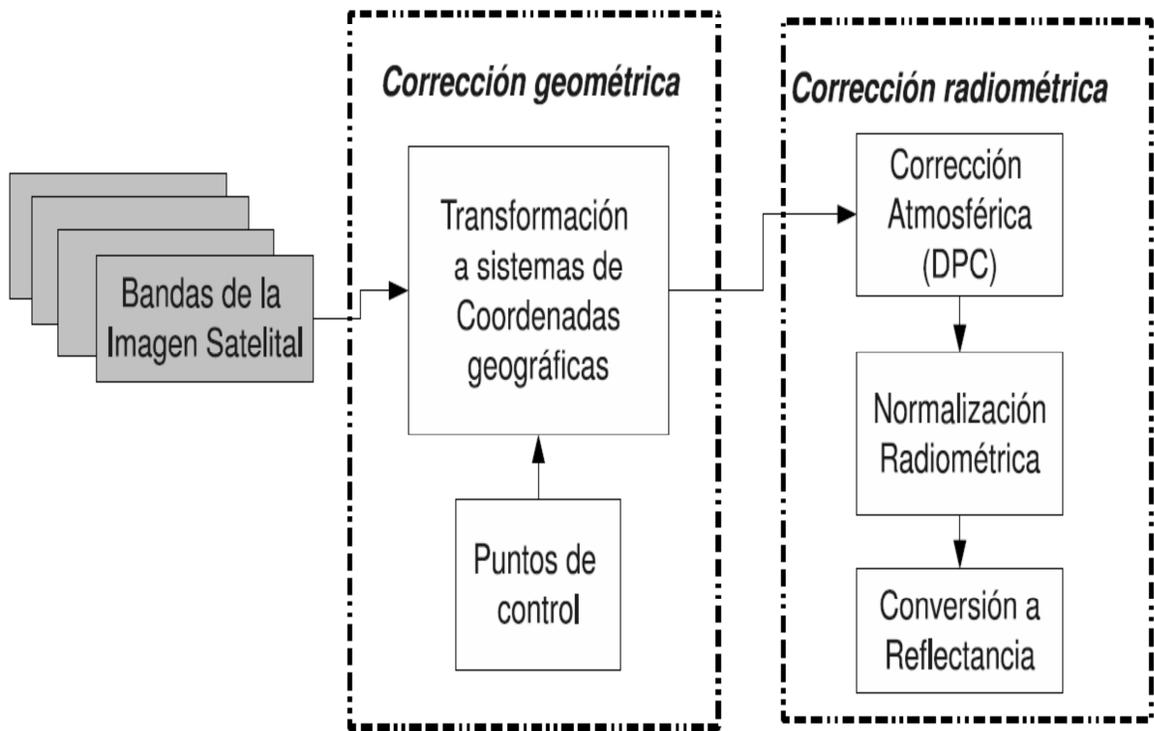


Fig. 3. Diagrama de flujo del proceso de pre-procesamiento de las imágenes satelitales utilizadas durante este trabajo.

2.4. Relación entre la imagen y los datos de campo

Una vez que se corrigieron las distorsiones en la imagen, se procedió a relacionar los datos de campo con los píxeles correspondientes en cada banda de la imagen. Se localizaron los 67 sitios de muestra en la imagen satelital. Se definió una ventana de 3×3 píxeles alrededor de cada punto identificado, ya que dichos puntos fueron escogidos pensando que cada sitio tenía una cobertura uniforme de al menos una hectárea (100×100 metros). De esta manera, por cada ventana de 9 píxeles, se extrajo la mediana, y se convirtió en el valor representativo de cada sitio de muestreo. Al final se ordenaron los datos obtenidos en una tabla de 67×6 valores de mediana. Por último, el conjunto completo de datos resultante se dividió en dos partes. El primer conjunto consta de 47×6 valores. Este grupo corresponde a los *datos de entrenamiento*. El segundo conjunto se compone de los 20×6 valores restantes y corresponde a los *datos de prueba*.

2.5. Revisión de la literatura sobre IVs existentes

Se tomaron treinta índices de vegetación extraídos de la literatura y se aplicaron sobre las tablas de píxeles de las zonas de estudio con el fin de obtener su factor de correlación con los datos de campo. De esta manera, se obtuvieron treinta valores asociados con cada valor de C de cada sitio de muestreo. Esto dio lugar a 30 matrices de 47×2 cada una. Posteriormente, por cada matriz se calculó el coeficiente de correlación ($\rho_{x,y}$) entre las dos columnas, obteniendo así 30 diferentes valores de correlación. Cada uno representa la correlación entre el factor C de los datos de campo, con uno de los índices. Es necesario hacer notar que de acuerdo a la definición del coeficiente de correlación, es posible obtener valores negativos. En este trabajo se optó por utilizar el valor absoluto $|\rho_{x,y}|$ ya que un valor de -1 indica un grado de correlación fuerte, lo mismo que un valor de 1. Por cuestiones de espacio aquí se muestra el desempeño de los mejores 10 índices (ver Tabla 2).

Índice	Descripción	Autor	$ \rho_{x,y} $
RVI4	$\frac{SWIR1}{SWIR2}$	Modificado de [13]	0.512
RVI5	$\frac{SWIR1}{R}$	Modificado de [13]	0.383
RVI1	$\frac{NIR}{R}$	Modificado de [13]	0.342
GEMI	$\mu(1 - 0.25 \mu) - \frac{(R - 0.125)}{1 - R}$ <p>donde $\mu = \frac{2(NIR^2 - R^2) + 1.5NIR + 0.5R}{NIR + R + 0.5}$</p>	[14]	0.329
RVI2	$\frac{NIR}{G}$	Modificado de [13]	0.329
NDVI	$\frac{NIR - R}{NIR + R}$	[8]	0.314
IPVI	$\frac{NIR}{NIR + R}$	[15]	0.314
SAVI2	$\frac{NIR}{R + b/a}$ Donde a y b son la pendiente y la abcisa al origen de la línea del suelo	[16]	0.309
SASI	$\beta_{SWIR1}(NIR - SWIR2)$ Donde $\beta_{SWIR1} = \cos^{-1}\left(\frac{a_2 + b_2 - c_2}{2ab}\right)$ y a, b, c son las distancias euclidianas entre NIR, SWIR1 y SWIR2	[17]	0.254
GVI	$-0.3344B - 0.3544G - 0.4556R + 0.6966NIR + 0.0242SWIR1 - 0.2630SWIR2$	[18]	0.228

Tabla 2. Los diez índices de vegetación del estado-del-arte que tuvieron un mejor rendimiento para la zona de estudio. Para más detalle sobre cada elemento que compone estos índices ver sección 2.6.1.

2.6. Desarrollo y aplicación de algoritmo basado en programación genética

En este trabajo, las soluciones candidatas que se someten al proceso de PG se codifican como árboles sintácticos, los cuales representan la fórmula matemática que define a un índice de vegetación. Por ejemplo, un IV muy utilizado en la literatura de la

Percepción Remota es el NDVI, cuya fórmula matemática se encuentra en la Tabla 2. Las variables y constantes en la fórmula: las bandas NIR y Red, son las hojas del árbol. Dentro de la nomenclatura de la PG, se les conoce como *Terminales*; mientras que a los operadores aritméticos (+, -, ÷), que son nodos internos, se les conoce como *Funciones*. El conjunto de Terminales y Funciones se define como el *conjunto de primitivas* de la PG; el cual representa el espacio de búsqueda del problema.

2.6.1. Conjunto de primitivas

- **Bandas espectrales (B, G, R, NIR, SWIR1, SWIR2).** Son los valores de reflectancias de los píxeles que representan los sitios de muestreo de cada una de las bandas de la imagen.
- **Ángulos espectrales (β_G , β_R , β_{NIR} , β_{SWIR1}).** Para conocer a detalle la manera de extraer los ángulos espectrales, sírvase revisar [17].
- **Parámetros de la línea del suelo (a, b).** La línea del suelo es la relación que existe entre las bandas R y NIR; la cuales, al ser graficadas una contra otra, agrupan los píxeles por encima de un umbral de valores. A este se le conoce como línea del suelo. El proceso que se presenta aquí para obtener los factores de la línea del suelo es el que propone [19].
- **Mejores índices convencionales (RVI4, RVI5, RVI1, GEMI, RVI2).** Se consideraron los cinco índices convencionales de mejor desempeño en la etapa anterior de esta metodología descrita en la sección 2.5.
- **Operadores aritméticos.** El conjunto de Funciones se compone primeramente por los operadores aritméticos básicos (+, -, y \times) ya que estos operadores son ampliamente utilizados para el diseño de índices de vegetación.
- **Operadores compuestos.** Representan estructuras aritméticas completas. Aquí se consideró la estructura de los índices más utilizados en la literatura de la Percepción Remota: el NDVI y el RVI. A la primera estructura se le conoce como índice espectral de diferencia normalizada (del inglés *Normalized Difference Spectral Index - NDSI*), mientras que a la segunda se le conoce como índice

espectral del cociente (del inglés *Ratio Spectral Index - RSI*), los cuales, se definen de la siguiente manera:

$$NDSI_{i,j} = \frac{R_i + R_j}{R_i - R_j}, \quad RSI_{i,j} = \frac{R_i}{R_j}, \quad (3)$$

donde R_k representa el valor de reflectancia de un pixel de la banda k .

2.6.2. Función de aptitud

La función de aptitud para este trabajo se basa en el coeficiente de correlación de Pearson $\rho_{x,y}$, el cual indica la fuerza y dirección de la relación lineal entre el factor C y cada uno de los índices sintetizados por la metodología. Se utilizó el valor absoluto del coeficiente de correlación como medida de desempeño porque mientras más cercano sea este valor a 1 ó -1, indistintamente la correlación entre las dos variables es más fuerte:

$$Q = \max(|\rho_{x,y}|) \quad (4)$$

donde x representa el factor C de RUSLE tomado en campo, mientras que y es el índice de vegetación generado por la metodología.

2.6.3. Ejecución del algoritmo de PG

Después del pre-procesamiento de la imagen (ver Figura 1), se genera la población inicial de índices de manera aleatoria, combinando los elementos de los conjuntos de Terminales y Funciones. Después, cada individuo de la población se evalúa con ayuda de la función de aptitud. El siguiente paso consiste en ordenar los índices de acuerdo a su desempeño con la función de aptitud, a fin de seleccionar los mejores candidatos para sobrevivir y descartar a los peores. Es entonces que sobreviene la recombinación genética con los candidatos elegidos. Finalmente, la siguiente generación resulta de la selección de los mejores entre los índices padres e hijos. Estos pasos son iterados sucesivamente hasta que se alcanza el máximo número de iteraciones. El índice con la

mayor correlación obtenido hasta ese momento es considerado el mejor, y constituye el nuevo *índice sintético*. A este nuevo índice se le llama GPVI_j (*Genetic Programming for Vegetation Index*), donde *j* es el número de la corrida en la que se generó dicho índice. En cuanto a los parámetros para correr el algoritmo de PG, estos serán descritos en la sección de resultados.

2.7. Evaluación de la solución

La evaluación de los resultados se lleva a cabo de acuerdo a la teoría del aprendizaje de máquina, es decir, a través de dos etapas llamadas entrenamiento y prueba. Para obtener la evaluación del desempeño de cada GPVI, se calculó la diferencia absoluta entre el coeficiente de correlación obtenido durante la etapa de entrenamiento y la etapa de prueba:

$$D = |\rho_{x,y}(\text{entren.}) - \rho_{x,y}(\text{prueba})| \quad (5)$$

Un desempeño similar en ambos conjuntos de datos implicaría una mayor estabilidad, ya que el comportamiento del índice ante diferentes valores de cobertura vegetal se mantiene consistente; obteniendo así el mejor mapa regional del factor C a partir de dicho índice.

2.8. Mapa del factor C y mapa de erosión

Finalmente, se lleva a cabo un mapeo lineal para convertir los valores numéricos de los mejores GPVI a los valores numéricos del factor C. De esta manera, los valores de C derivados de cada índice quedan en la misma escala que los valores de C derivados del muestreo de campo. Asimismo, para validar el desempeño de los nuevos índices, se aplican los nuevos mapas de C en la ecuación del modelo RUSLE para calcular la tasa de erosión. Todos los factores que componen el modelo RUSLE, excepto por C son conservados a lo largo de esta prueba. Es necesario decir que dichos factores (R, K, L, S, y P) fueron determinados usando un programa de Sistemas de Información Geográfica ArcGIS®, siguiendo la metodología descrita por [6].

3. Resultados

En esta sección se presentan los resultados de la metodología descrita en la sección anterior. La metodología se implementó utilizando MatLab®, con el módulo de programación genética conocido como GPLab, el cual está disponible con licencia libre en (<http://gplab.sourceforge.net/index.html>). La generación de los mapas de factor C y los mapas de erosión se llevaron a cabo con ArcGIS®. La Tabla 3 muestra el valor de los parámetros utilizados para el algoritmo de GP durante las pruebas experimentales. Se utilizaron valores canónicos recomendados por la literatura especializada en GP, los cuales se reportan con un buen desempeño en una gran variedad de aplicaciones. Para mayores detalles sobre los ajustes de los parámetros, sírvase revisar los trabajos de [20, 21].

Población inicial	Mitad árboles completos Mitad árboles de distribución aleatoria
Generaciones	50
Tamaño de la población	50 individuos
Probabilidad de cruzamiento	0.7
Probabilidad de mutación	0.3
Tipo de selección	Torneo de presión lexicográfica
Tipo de elitismo	Conservar el mejor de la generación
Profundidad de los árboles	Dinámica, 4 niveles

Tabla 3. Parámetros utilizados para el entrenamiento del algoritmo basado en PG.

Se llevaron a cabo treinta corridas del algoritmo de PG conservando el mejor individuo de cada una de estas corridas y fueron etiquetados como $TGPVI_j$ (*Todos santos Genetic Programming for Vegetation Index*), donde J representa el número de la corrida. Cada una de estos individuos es considerado como una solución aceptable. La Tabla 4 resume el desempeño de las soluciones obtenidas. Por razones de espacio, en

este artículo se publican los 10 mejores de esas corridas. El criterio para definir los mejores se encuentra en la sección 2.7. En general, los índices encontrados por el algoritmo basado en PG muestra un mejor rendimiento comparado con el de los índices convencionales, para ello se pueden comparar la Tabla 2 con la Tabla 4.

Índice	Descripción	ρ_{entren}	ρ_{prueba}	Diferencia
TGPVI17	$RSI(NIR, RSI(\beta_R, \beta_G)) + RSI(RSI(RVIA, a), GEMI) - RVIA)$	0.648	0.645	0.002
TGPVI27	$RSI(GEMI, RVIA - NIR) - NIR$	0.642	0.639	0.003
TGPVI19	$GEMI \times \beta_{SWIR1} - GEMI - RVIA$	0.622	0.629	0.008
TGPVI28	$\frac{GEMI}{RVIA - GEMI} - GEMI$	0.643	0.655	0.013
TGPVI8	$NDSI(NDSI(RVIA, \beta_G), RVIA) \times GEMI$	0.643	0.673	0.030
TGPVI13	$RVIA - GEMI - GEMI - G$	0.626	0.674	0.048
TGPVI16	$-RVIA - NIR$	0.626	0.679	0.053
TGPVI21	$\frac{NDSI(\beta_{SWIR1}, GEMI)}{NDSI(RVIA, GEMI)}$	0.640	0.698	0.059
TGPVI23	$RSI(B - GEMI, RVIA^2 \times b)$	0.648	0.710	0.062
TGPVI3	$NDSI(RVIA, GEMI) \times RVIA \times \beta_{NIR}$	0.619	0.681	0.062

Tabla 4. Los diez mejores índices sintetizados por la metodología basada en PG.

Para obtener la tasa de erosión en la cuenca Todos Santos, es necesario construir mapas de C a partir de uno de los índices de vegetación, ya sea de los convencionales o los sintéticos. Generalmente los IVs poseen una escala numérica diferente a la escala numérica del factor C. Por ejemplo, la escala numérica para NDVI es de -1 a 1; mientras que el del factor C es de 0 a 1. Por ello es necesario aplicar una transformación lineal para hacer la conversión entre una escala y la otra. Esta conversión se hace utilizando la función *cftool* de MatLab, la cual trata de ajustar iterativamente una línea recta al conjunto de datos. Cuando se obtiene esta línea recta, se puede transformar el rango numérico del índice al rango numérico de C y viceversa. Esta transformación lineal se

utilizó para generar mapas de C a partir de los 3 mejores índices sintetizados por la PG y a partir del mejor índice del estado del arte. La definición del mejor índice sintético se encuentra en la sección 2.7.

En seguida se procede a obtener un mapa de erosión a partir del mejor índice. Se obtuvieron mapas de erosión usando el mapa de C a partir del índice TGPVI17 y del RVI4 (ver Figura 4). Ya que la comparación entre mapas sólo puede ser cualitativa, se ha diseñado un ejercicio con el que se puede comparar cuantitativamente la tasa de erosión punto por punto de los sitios de muestreo. Para calcular la tasa de erosión de la cuenca Todos Santos, se utilizó el modelo RUSLE. Actualmente RUSLE puede ser fácilmente implementado utilizando Sistemas de Información Geográfica (SIG), pues cada uno de los coeficientes se puede manejar como una capa de datos para SIG. Además, ya que RUSLE es un modelo multiplicativo, sencillamente basta con multiplicar cada una de estas capas pixel a pixel, y el resultado es un nuevo pixel, es decir, una nueva capa que contendrá la pérdida promedio de suelo por erosión: A . Para cada uno de los mapas de erosión que se presentan a continuación, los factores R, L, S, K y P de RUSLE se calcularon como se describe en el trabajo de [6]. En cambio, el factor C va cambiando de acuerdo al mapa de C del índice en turno.

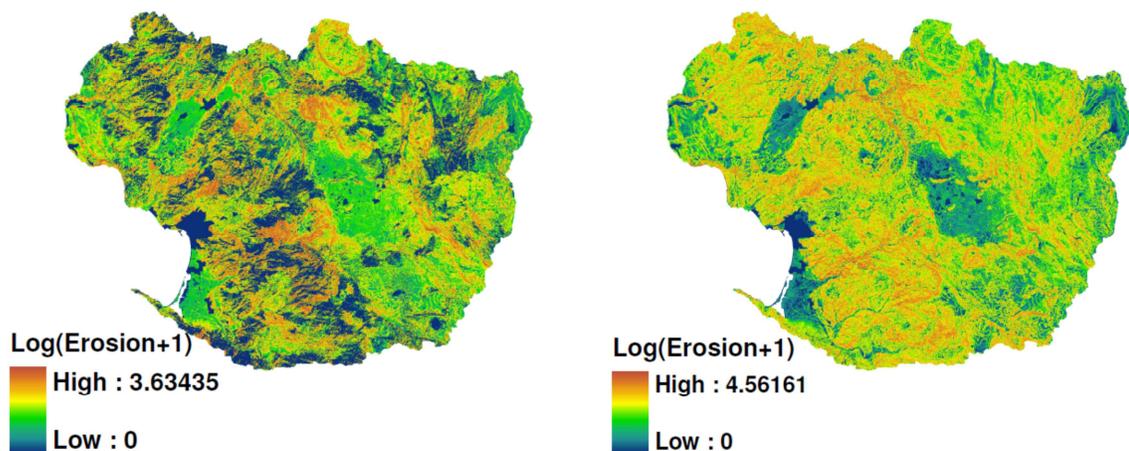


Fig. 4. Mapas del logaritmo de la tasa de erosión de la cuenca Todos Santos. a) Usado el factor C basado en el índice TGPVI17. b) Usando el factor C basado en RVI4.

Para un análisis cuantitativo es necesario observar la Tabla 5; la cual, en el primer renglón muestra el promedio de la tasa de erosión de los 67 sitios de muestreo de campo. Para este análisis, dicho renglón funge como punto de referencia o *ground truth*. De igual manera, en la misma tabla se establecen los promedios de la tasa de erosión calculados a partir de los valores de C derivados de los diferentes índices de vegetación. Nótese que las mejores aproximaciones a las mediciones de campo son las obtenidas de los índices de sintéticos. Por ejemplo, la tasa derivada del *TGPVI19* es de $76 \pm 104 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$ y muestra la mayor aproximación a la tasa de erosión obtenida usando el factor C calculado de los datos de campo; que es de $77 \pm 154 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$. Además, *TGPVI17* y *TGPVI27* también arrojan una tasa de erosión con una aproximación aceptable: 67 ± 97 y 69 ± 97 respectivamente. Finalmente, los índices tomados del estado del arte arrojan tasas de erosión promedio sobre-estimadas con respecto a los índices sintéticos y a los puntos de referencia. Por ejemplo, la tasa obtenida a partir del *RVI4* muestra una tasa de erosión aproximadamente un 50% más alta que la real: 120 ± 184 .

Método utilizado para obtener C	Erosión ($\text{Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$)
Datos de campo (<i>ground truth</i>)	76.6 ± 153.6
TGPVI17	67.1 ± 97
TGPVI19	76.5 ± 103.5
TGPVI27	68.6 ± 97.4
RVI4	119.7 ± 184.4
RVI5	193.7 ± 393.3

Tabla 5. Media y desviación estándar de la tasa de erosión obtenida para los sitios de muestreo. Cada renglón representa una mapa del factor C diferente.

4. Conclusiones

En este documento se ha descrito una nueva metodología basada en programación genética, la cual sintetiza automáticamente índices de vegetación que tienen una buena correlación con el factor C del modelo de erosión RUSLE. Como resultado, los índices

sintetizados obtuvieron un desempeño mejor al de los índices convencionales reportados en la literatura, tales como el NDVI, GVI y un índice alternativo al RVI, llamado en este trabajo como RVI4. De entre los índices sintéticos obtenidos sobresalen los índices *TGPVI17*, *TGPVI27* y *TGPVI19*. Además, destaca el hecho de que la tasa de erosión a partir del *TGPVI19* es de $76 \pm 104 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$ y fue la que más se acercó al promedio de erosión obtenido usando el factor C medido en campo: $77 \pm 154 \text{ Mg} \cdot \text{km}^{-2} \cdot \text{año}^{-1}$.

Basado en los resultados experimentales, es posible decir que el uso de programación genética es una buena alternativa para implementar análisis multi-espectral y otras aplicaciones de la percepción remota; tales como clasificadores e índices matemáticos que reconozcan y realcen la señal de diferentes rasgos en la superficie terrestre. Además, la ventaja de la PG es su propiedad de *caja blanca*; la cual consiste en que las soluciones generadas por la PG pueden ser analizadas, simplificadas e interpretadas por un científico o ingeniero experto en el área de aplicación.

El trabajo futuro está relacionado con la generalidad de la metodología desarrollada para identificar un buen índice para estimar C. En este trabajo se utilizaron datos de campo de una sola cuenca hidrográfica con características climatológicas específicas. Se pretende continuar aplicando la metodología bajo diferentes condiciones que permitan discernir sobre la generalidad de los resultados obtenidos.

Finalmente, nosotros pensamos que los resultados de este trabajo arrojan una novedosa metodología que contribuye al estado del arte en diferentes disciplinas científicas, tales como las geo-ciencias ambientales, la percepción remota y la inteligencia computacional. Esta contribución mejora la manera de obtener el factor C del modelo RUSLE y en consecuencia produce mapas de erosión más precisos, sin la necesidad de llevar a cabo extensas campañas de recolección de datos de campo; lo cual es un aliciente para continuar explorando el uso de la programación genética en aplicaciones de la percepción remota.

5. Referencias

- [1] A. Vrieling, Mapping Erosion from Space. Tesis de doctorado. 2007. Wageningen University. Wageningen, Netherlands. 152 pp.
- [2] S. de Jong, Applications of Reflective Remote Sensing for Land Degradation Studies in a Mediterranean Environment. Tesis de doctorado. 1994. Universiteit Utrecht, Netherlands. 237 pp.
- [3] K. Renard, G. Foster, G. Weesies, "Predicting Soil Erosion by Water: A Guide to Conservation Planning With the Revised Universal Soil Loss Equation". U.S. Department of Agriculture. Vol. 703. 1996.
- [4] J. Montoya, F. Francés, J. I. Vélez, P. Julien, "Desarrollo de un modelo distribuido de producción, transporte y depositación de sedimentos. Aplicación en una cuenca experimental". En XXII Congreso Latinoamericano de Hidráulica. 2006. Ciudad Guyana. Venezuela.
- [5] D. C. Flanagan, M. A. Nearing, "Water Erosion Prediction Project. Hillslope Profile And Watershed Model Documentation". USDA-ARS National Soil Erosion Research Laboratory. Vol. 10. 1995.
- [6] S. V. Smith, S. H. Bullock, A. Hinojosa-Corona, E. Franco-Viscaíno, M. Escoto.Rodríguez, T. G. Kretschmar, L. M. Farfán, J. M. Ceseña, "Soil erosion and significance for carbon fluxes in a mountainous mediterranean climate watershed". Ecological Applications. Vol. 17. No. 5. 2007. 1379-1387 pp.
- [7] R. G. Congalton, M. Balogh, C. Bell, K. Green, J. A. Milliken, R. Toman, "Mapping and monitoring agricultural crops and other land cover in the lower colorado river basin". Photogrammetric Engineering & Remote Sensing. Vol. 64. No. 11. 1998. 1107–1113 pp.

- [8] W. Rouse, R. H. Haas, J. A. Schell, D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS". Third ERTS Symposium, NASA SP-351. Vol. 1. 1973. 309-317 pp.
- [9] W. Wishmeier, D. Smith, Predicting Rainfall Erosion Losses: A Guide to Conservation Planning. U. S. Department of Agriculture. Vol. 537. 1978.
- [10] H. Bauer, "The statistical analysis of chaparral and other plant communities by means of transect samples". Ecology. Vol. 24. 1943. 45-60 pp.
- [11] B. Zippin, J. M. Vanderwier, Scrub community descriptions of the Baja California peninsula, Mexico. Madroño. 1994. 85–119 pp.
- [12] M. A. González Botello, Estimaciones de la cobertura vegetal y del suelo en el noroeste de Baja California y su aplicación a la modelación de la erosión. 2010. Tesis de maestría. Universidad Autónoma de Baja California. Facultad de Ciencias. Ensenada, Baja California.
- [13] C. F. Jordan. "Derivation of leaf area index from quality of light on the forest floor". Ecology. Vol. 50. 1969. 663-666 pp.
- [14] B. Pinty, M. M. Verstraete, "GEMI: A Non-Linear Index to Monitor Global Vegetation from Satellites". Vegetatio. Vol. 101. 1991. 15-20 pp.
- [15] R. E. Crippen, "Calculating the Vegetation Index Faster". Remote Sensing of Environment". Vol 34. 1990. 71-73 pp.
- [16] J. Qi, Y. Kerr, A. Chehbouni, "External Factor Consideration in Vegetation Index Development". En Proc. of Physical Measurements and Signatures in Remote Sensing ISPRS. 1994. 723-730 pp.
- [17] S. Khana, A. Palacios-Orueta, M. L. Whiting, S. L. Ustin, D. Riaño, J. Litago, "Development of angle indexes for soil moisture estimation, dry matter detection

and land-cover discrimination". *Remote Sensing of Environment*. No. 109. 2007. 154–165 pp.

[18] E. P. Crist, R. C. Cicone, "Application of the tasseled cap concept to simulated thematic mapper data". *Photogrammetric Engineering and Remote Sensing*. Vol. 50. 1994. 343-352 pp.

[19] F. F. Sabins, *Remote sensing: principles and interpretation*. 3° edición. 1997. Freeman, New York. U.S.A.

[20] J. R. Koza. *Genetic programming: On the programming of computers by means of natural selection*. 1992. MIT press. Cambridge, MA. USA.

[21] R. Poli, W. B. Langdon, N. F. McPhee, *A field guide to genetic programming*. Lulu. U.K. 2008.

7. Autores

Dr. Cesar Augusto Puente Montejano es profesor-investigador de tiempo completo de la Facultad de Ingeniería de la Universidad Autónoma de San Luis Potosí. Obtuvo el grado de Doctorado en Ciencias, en Ciencias de la Tierra por el Centro de Investigación Científica y de Educación Superior de Ensenada, B.C. (CICESE) en Junio de 2011. Sus intereses de investigación radican en aplicaciones del cómputo evolutivo en problemas de las ciencias de la Tierra.

Dr. Gustavo Olague Caballero es investigador científico en el Centro de Investigación Científica y de Educación Superior de Ensenada, B.C. (CICESE), en la División de Física Aplicada. Obtuvo su grado de *Diplôme de Doctorat en Imagerie, Vision et Robotique* por el Institut National Polytechnique de Grenoble, Francia.