

EVALUACIÓN DE MÚLTIPLES FUNCIONES DE MÉRITO EN LA INVERSIÓN DE PARÁMETROS BIOFÍSICOS A TRAVÉS DEL MODELO PROSAIL USANDO TABLAS DE BÚSQUEDA

J. P. Rivera⁽¹⁾, J. Verrelst⁽¹⁾, G. Leonenko⁽²⁾ and J. Moreno⁽¹⁾

⁽¹⁾Laboratorio de procesado de imagines (LPI) Parque Científico – Universidad de Valencia. C/ Catedrático José Beltrán, 2. 46980 Paterna (Valencia). Teléfono: 963544067 E-mail: rijuanpa@uv.es

⁽²⁾ Dept. Psychological Med, Cardiff University, CF14 4XN, UK

RESUMEN

La inversión por tablas de búsqueda (LUT) construidas a partir de modelos de transferencia radiativa es uno de los métodos más usados para la estimación de parámetros biofísicos. Debido a la naturaleza “mal formulada” inherente en la inversión, son necesarias estrategias de regularización. En el presente trabajo evaluamos tres estrategias de regularización para estimar del contenido de clorofila en hoja (LCC) e índice de área foliar (LAI): 1. Tipo de función de mérito, 2. Adicción de ruido a los espectros simulados con PROSAIL y 3. Promedio de múltiples soluciones. La mejor estimación de LCC se obtuvo con espectros normalizados y la función de mérito ‘L-estimate’ con un error relativo (NRMSE) de 17.6% y para la estimación del LAI con espectros sin normalizar y ‘least-squares estimator (LSE)’ con un error relativo de 15.3%.

Palabras clave: Parámetros biofísicos, inversión, tablas de búsqueda, funciones de mérito, modelos de transferencia radiativa, PROSAIL, sentinel-2

ABSTRACT

Lookup-table (LUT)-based radiative transfer model inversion is widely used but regularization strategies are needed to mitigate the drawback of ill-posedness. We evaluated various regularization options to improve leaf chlorophyll content (LCC) and leaf area index (LAI), 1) cost functions (CFs), 2) added noise, and 3) mean of multiple solutions in LUT-based inversion. Three families of CFs were compared: information measures, M-estimates and minimum contrast methods. The best LCC retrievals were obtained using a normalized 'L-estimate' function with a relative error of 17.6% while best LAI retrievals were obtained through non-normalized 'least-squares estimator' (LSE) with a relative error of 15.3%.

Keywords: Biophysical parameters; LUT-based inversion; cost functions; radiative transfer models; PROSAIL; Sentinel-2

1. INTRODUCCIÓN

La estimación de parámetros biofísicos de la vegetación como el contenido de clorofila en hoja (LCC) e índice de área foliar (LAI) son claves para estudiar las dinámicas climáticas y evaluar el estado fenológico y vigor de las plantas. El desarrollo de nuevos sensores remotos y plataformas espaciales como la familia Sentinel-2 y Sentinel-3 marcan nuevos retos para mejorar las estrategias y los algoritmos empleados para mitigar la naturaleza “mal-formulada” inherente en la inversión. Uno de los métodos más empleados por su sencillez y no necesitar entrenamiento a priori son las tablas de búsqueda (LUT) donde la función de mérito (FM) más empleada es el error medio cuadrático (RMSE). Los trabajos realizados por Claude E. Shannon y Warren Weaver a finales de la década de los cuarenta han abierto una ventana en la llamada “Teoría de la Información”, que se aplica en áreas como ecología, biología, probabilidad, economía, etc. El propósito de este trabajo es evaluar cinco FM provenientes de la

teoría de la información, junto con tres estrategias de regularización enfocadas a la normalización de los espectros simulados con los modelos de transferencia radiativa (MTR), adicción de ruido gaussiano y el promedio de las mejores soluciones para mejorar la naturaleza “mal formulada” inherente en la inversión.

2. METODOLOGÍA

2.1. Funciones de mérito (FM)

En el método de inversión por LUT, la FM permite discriminar la similitud entre los espectros simulados, medidos en campo o laboratorio (Q) y los medidos por un sensor remoto (P), donde $D[P,Q]$ representa la distancia entre dos espectros. Evaluamos cinco FM clasificadas en tres familias: *f-divergence*, *M-estimates* y *mínimo contraste*. Un listado más completo de estas funciones se pueden encontrar en el trabajo desarrollado por Leonenko *et al.*, 2013. La tabla 1 muestra las FM analizadas.

***f-divergence*:** Esta familia se basa en la minimización de la distancia entre dos funciones de

probabilidad y asume la existencia de una función de distribución conocida entre ellas. Introducidas por Kullback and Leibler en 1951 ha sido usadas en diversas ramas de la estadística, física e informática. Para ajustar las medidas de reflectividad a funciones de probabilidad se normalizan por la suma de los valores de reflectividad de todas las bandas como lo muestra la ecuación 1.

$$Q = (q_1^*, \dots, q_n^*) = \frac{B^*(\lambda_1)}{\sum_{\lambda_j \in \Lambda} B^*(\lambda_j)}, \dots, \frac{B^*(\lambda_n)}{\sum_{\lambda_j \in \Lambda} B^*(\lambda_j)} \quad (1)$$

M-estimates: Hacen parte de un conjunto estimador que presentan mayor robustez de tipo de máxima similitud o en inglés “maximum likelihood-type” de donde se deriva la letra M. A diferencia de los f-divergence, estos no asumen el conocimiento de la función de distribución de los espectros, y los datos pueden estar o no normalizados.

Mínimos contraste: Considerando el dominio espectral, las medidas de reflectividad son tomadas como funciones de densidad espectral de algunos procesos estocásticos. Se busca minimizar la distancia o “contraste” entre un modelo paramétrico y una función de densidad espectral no paramétrica.

Función de mérito	Ecuación
Kullback Leibler	$D[P, Q] = \sum_{\lambda_n} p(\lambda_i) \ln \left(\frac{p(\lambda_i)}{q(\lambda_i)} \right)$
Squared-Hellinger	$D[P, Q] = \sum_{\lambda_n} \left(\sqrt{p(\lambda_i)} - \sqrt{q(\lambda_i)} \right)^2$
LSE	$D[P, Q] = \sum_{\lambda_n} (p(\lambda_i) - q(\lambda_i))^2$
L-estimate	$D[P, Q] = \sum_{\lambda_n} p(\lambda_i) - q(\lambda_i) $
$K(x) = \log x + \frac{1}{x}$	$D[P, Q] = \sum_{\lambda_n} \left\{ \log \left(\frac{q(\lambda_i)}{p(\lambda_i)} \right) + \frac{p(\lambda_i)}{q(\lambda_i)} \right\} - 1$
$K(x) = x \log x - x$	$D[P, Q] = 1 + \sum_{\lambda_n} \frac{q(\lambda_i)}{p(\lambda_i)} \left\{ \log \left(\frac{q(\lambda_i)}{p(\lambda_i)} \right) - 1 \right\}$

Tabla 1. Funciones de mérito evaluadas

2.2. Generación de la LUT

Se construyó una LUT con datos sintéticos generados por modelos de transferencia radiativa a nivel de hoja y cubierta, empleando el programa informático ARTMO v.3.0b por sus siglas en inglés “Automated Radiative Transfer Models Operator” desarrollado en MatLab, el cual permite la generación de espectros de reflectividad a partir de MTR a nivel de hoja (e.j: PROSPECT-4, PROSPECT-5) y de cubierta (e.j.: 4SAIL, SLC, FLIGHT) (Verrelst *et al.*, 2011). ARTMO está configurado con diversos módulos para visualizar y analizar los resultados, implementar nuevos sensores, realizar inversión de parámetros biofísicos a partir de datos simulados o medidos en campo por diversos métodos (ej.: LUT, métodos estadísticos y relaciones espectrales).

La tabla 2 muestra el rango y la distribución usada en los parámetros biofísicos usados para simular reflectividades con el modelo PROSAIL. Se generaron alrededor de 5 billones de combinaciones donde se tomaron de manera aleatoria 100,000 para la construcción de la LUT.

	Parámetros	Unid.	Rango	Distribución
PROSPECT 4	Índice Estructura	n.a.	1,3-2,5	Uniforme
	Cont. clorofila	µg/cm ²	5-75	Gaussiana (µ:35/std:30)
	Cont. mat. seca	g/cm ²	0,001-0,03	Uniforme
	Cont. agua	cm	0,002-0,05	Uniforme
4SAIL	Índice área foliar	m ² /m ²	0,1-7	Gaussiana (µ:3/std:2)
	% Hum. suelo	n.a.	0-1	Uniforme
	Ang. incl. hojas	°	40-70	Uniforme
	Factor Hot Sopt	m/m	0,05-0,5	Uniforme
	Fr. rad. difusa	%	0,05	n.a
	Ang. solar zenital	°	22,3	n.a
	Ang. zenital de obs.	°	20,19	n.a
	Ang. Rel. azimutal	°	0	n.a

Tabla 2. Parámetros del modelos PORSAIL

2.3. Base de datos de validación

Los datos de validación provienen de la campaña de SPARC (SPectra bARrax Campaign) realizadas durante los años de 2003 y 2004 en la zona de Barrax, Castilla la Mancha (30°3'N, 2°6'W, 700 m.s.n.m). Las medidas espectrales se adquirieron con sensor CHRIS en modo 1 con una resolución espacial en nadir de 34m. Las medidas fueron procesadas usando ESA's CHRIS-Box disponible en VISAT/BEAM. (Alonso & Moreno, 2005; Fernández *et al.*, 2005).

2.4. Sentinel-2

Sentinel-2 (S2) es una de las cinco misiones Sentinel que desarrollará la Agencia Espacial Europea dentro del programa de Vigilancia Mundial del Medio Ambiente y de la Seguridad (GMES) (Drusch *et al.*, 2012). Un resumen de la configuración de las bandas previstas para S2 se muestra en la tabla 3. En el presente estudio se realizó un remuestreo espectral a las bandas resaltadas en la tabla 3.

ID	Banda central (nm)	Ancho banda (nm)	Res. Esp. (m)	ID	Banda central (nm)	Ancho banda (nm)	Res. Esp. (m)
B1	443	20	60	B8	842	115	10
B2	490	65	10	B8a	865	20	20
B3	560	35	10	B9	942	20	60
B4	665	30	10	B10	1375	30	60
B5	705	15	20	B11	1610	90	20
B6	740	15	20	B12	2190	180	20
B7	783	20	20				

Tabla 3. Configuración de bandas espectrales S2

2.5. Regularización

Diversos estudios han mostrado la importancia de aplicar procesos auxiliares u opciones de regularización para mitigar la naturaleza “mal formulada” del proceso de inversión (Baret *et al.*, 2007), En el presente estudio se aplicaron tres estrategias de regularización: 1. Adicción de ruido gaussiano a los datos simulados por MTR entre 0% y 50% con un paso de 2%. 2. Promediar los n mejores resultados obtenidos con la respectiva FM entre 0% (resultados con menor distancia) y 50% del total de espectros que conforman la LUT. 3. Normalización de los espectros en las familias M-estimate y Mínimo contraste. Para evaluar los resultados de la inversión se ha utilizado el estadístico NRMSE (error relativo). La ecuación 2 indica la fórmula del estadístico donde y son los valores medidos en la campaña y x los valores predichos por el método de inversión.

$$NRMSE = \sqrt{\frac{\sum_{t=1}^n (x_t - y_t)^2}{n}} \quad (2)$$

3. RESULTADOS

3.1. Impacto de las funciones de mérito y opciones de regularización

Se evaluó la inversión de LCC y LAI con cada una de las estrategias de regularización. La figura 1 muestra las matrices de error relativo donde en el eje y representa el porcentaje del número de mejores soluciones sobre el total de la LUT, y el eje x el porcentaje de ruido añadido a la LUT. La normalización de los espectros se ha identificado de la siguiente manera: *: normalizado, **: sin normalizar.

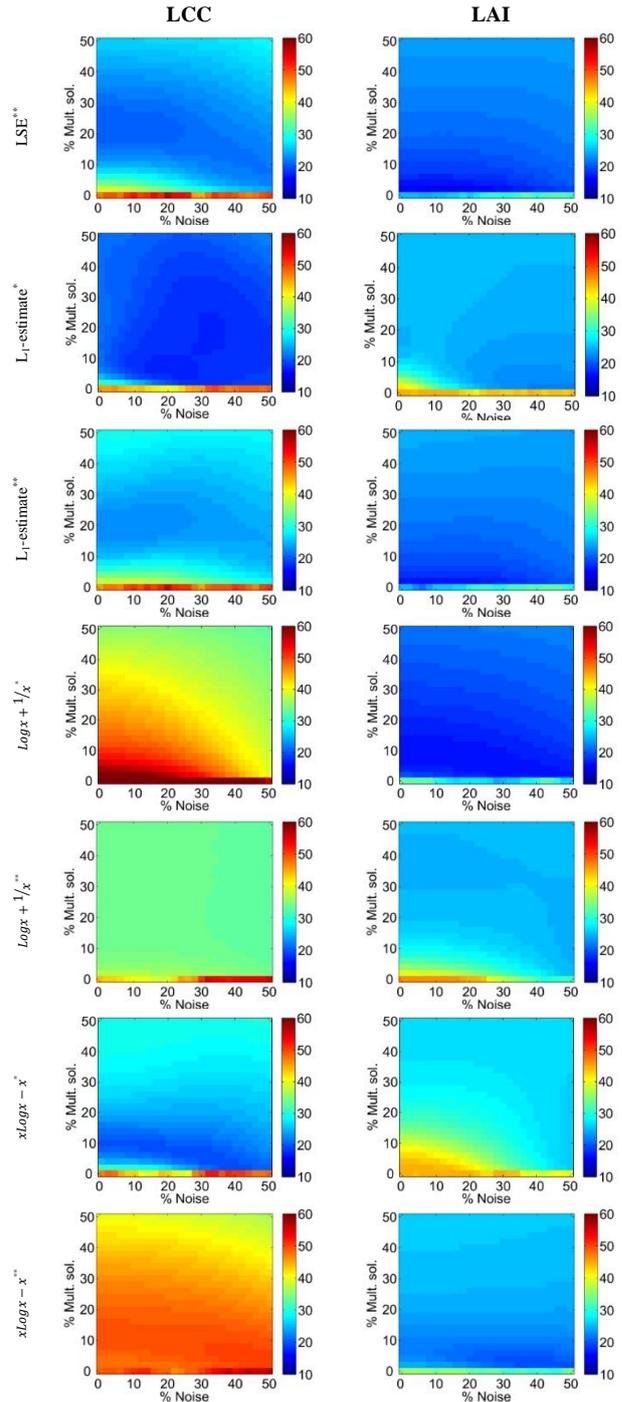
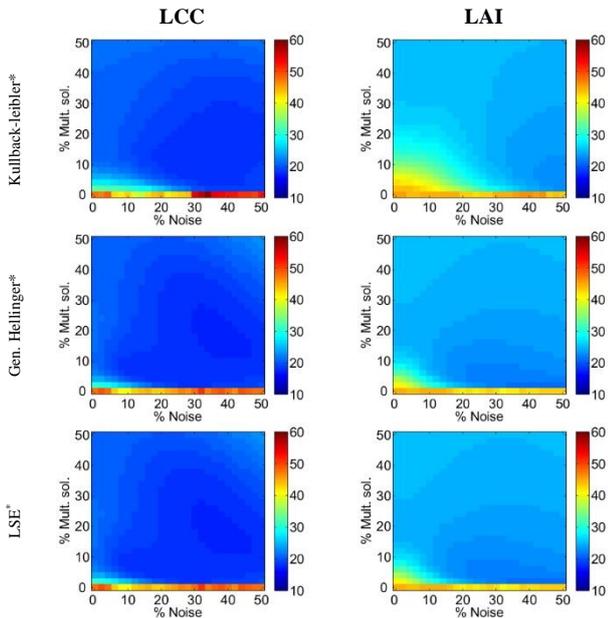


Figura 1. Conjunto de matrices de errores relativos.

La tabla 4 muestra los mejores resultados obtenidos con todas las funciones de mérito evaluadas. Las columnas A,B,C indican el porcentaje de múltiples soluciones, porcentaje de ruido y error relativo (NRMSE) respectivamente.

Función de mérito	LCC			LAI		
	A	B	C	A	B	C
Kullback leibler*	10	26	19.47	4	50	22.74
Generalised Hellinger*	20	36	17.63	2	42	21.34
LSE*	20	36	17.63	2	42	21.34

LSE**	22	0	20.27	2	14	15.32
L1-estimate*	6	18	17.59	2	50	22.25
L1-estimate**	20	12	22.52	2	12	16.57
$K(x)=\log(x)+1/x^*$	50	50	33.09	2	42	16.51
$K(x)=\log(x)+1/x^{**}$	16	50	32.91	10	50	24.57
$K(x)=x(\log(x))-x^*$	6	30	19.58	6	50	25.82
$K(x)=x(\log(x))-x^{**}$	50	50	37.06	2	40	19.56

Tabla 4. Mejores opciones de regularización por FM

Se puede generalizar que las FM que estiman mejor LCC no siempre lo hacen para el LAI, exceptuando la familia M-estimates que presentan errores relativos bajos en la estimación de ambos parámetros, pero dependen de la normalización previa de los espectros. Para LCC se obtienen errores relativos menores con la normalización, mientras que para LAI con espectros no normalizados.

Las matrices de errores relativos muestran el impacto del promedio de múltiples soluciones. Los menores errores relativos en la estimación de LCC se han obtenido con un mayor porcentaje de múltiples soluciones que para LAI con espectros sin normalizar.

La normalización de los espectros tienen un mayor impacto en la estimación de LAI. Para lograr errores relativos bajos se necesita incluir mucho ruido, mientras que para espectros no normalizados los niveles de ruido son menores.

El ruido en la señal permite mejorar la estimación de ambos parámetros biofísicos. Pero su impacto es menos relevante que el porcentaje de múltiples soluciones.

Para la estimación de LCC los mejores resultados se obtuvieron con las siguientes opciones de regularización: 6% de múltiples soluciones, 18% de ruido, con un error relativo del 17,59%. Para LAI con 2% de múltiples soluciones y 14% de ruido con un error relativo del 15,32. La figura 2 muestra el mapa de LCC(a) y LAI(b) estimados con el método de tabla de búsqueda con las anteriores opciones de regularización.

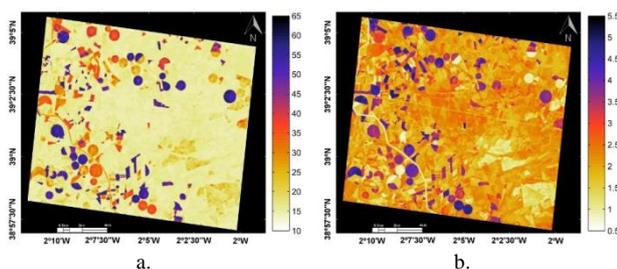


Figura 2. Mapas de estimación del LCC (a) y LAI (b) con su respectiva FM y opciones de regularización

4. CONCLUSIONES

Los resultados muestran que la inversión es un proceso complejo, donde en función del parámetro biofísico que se desea estimar la naturaleza ‘mal formulada’ inherente en la inversión se puede mejorar con la estrategia correcta de regularización.

La simulación de las bandas del Sentinel-2 muestra el gran potencial para la estimación de parámetros biofísicos para el estudio de las dinámicas de la vegetación a escala global.

Es necesario contar con herramientas informáticas que permitan de una manera intuitiva y con gran flexibilidad evaluar el impacto con las diversas estrategias de regularización para mitigar la naturaleza ‘mal formulada’ de la inversión y mejorar la estimación de los parámetros biofísicos.

AGRADECIMIENTO

Este trabajo ha sido posible gracias al Proyecto AYA2010-21432-C02-01 del Ministerio de Economía y Competitividad de España.

5. BIBLIOGRAFIA

ALONSO, L. & MORENO, J. 2005. *Advances and limitations in a parametric geometric correction of CHRIS/PROBA data*. Proceedings of the 3rd CHRIS/Proba Workshop.

BARET, F., HAGOLLE, O., GEIGER, B., BICHERON, P., MIRAS, B., HUC, M., BERTHELOT, B., NIÇO, F., WEISS, M., SAMAIN, O., ROUJEAN, J. & LEROY, M. 2007. *LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION. Part 1: Principles of the algorithm*. Remote Sensing of Environment, 110, 275-286.

DRUSCH, M., DEL BELLO, U., CARLIER, S., COLIN, O., FERNANDEZ, V., GASCON, F., HOERSCH, B., ISOLA, C., LABERINTI, P., MARTIMORT, P., MEYGRET, A., SPOTO, F., SY, O., MARCHESE, F. & BARGELLINI, P. 2012. *Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services*. Remote Sensing of Environment, 120, 25-36.

FERNÁNDEZ, G., MORENO, J., GANDÍA, S., MARTÍNEZ, B., VUOLO, F. & MORALES, F. 2005. *Statistical variability of field measurements of biophysical parameters in SPARC-2003 and SPARC-2004 campaigns*. Proceedings of the SPARC Workshop.

LEONENKO, G., NORTH, P. & LOS, S. 2013. *Statistical distances and their applications to biophysical parameter estimation: Information measures, M-estimates, and minimum contrast methods*. Remote Sensing, 5, 1355-1388.

VERRELST, J., RIVERA, J., ALONSO, L. & MORENO, J. 2011. *ARTMO: an Automated Radiative Transfer Models Operator toolbox for automated retrieval of biophysical parameters through model inversion*. Proceedings of EARSeL 7th SIG-Imaging Spectroscopy Workshop, Edinburgh, UK.