

# UN ENFOQUE BAYESIANO PARA LA ESTIMACIÓN DE HUMEDAD DEL SUELO A PARTIR DE DATOS SAR POLARIMÉTRICOS EN BANDA L

Matías Barber<sup>†</sup>, Martín Maas<sup>‡</sup>, Francisco Grings<sup>†</sup> y Haydeé Karzembraum<sup>†</sup>

<sup>†</sup>*Instituto de Astronomía y Física del Espacio (IAFE) - UBA- CONICET* [mdmaas@iafe.uba.ar](mailto:mdmaas@iafe.uba.ar); [www.iafe.uba.ar/tele](http://www.iafe.uba.ar/tele)

**Resumen:** Estimar la humedad del suelo a partir de datos de un radar de apertura sintética (SAR) polarimétrico presenta varios desafíos. Es necesario dar cuenta del scattering que produce un suelo rugoso, para lo que se dispone en la literatura de modelos analíticos aproximados y modelos semiempíricos basados en datos experimentales. Al mismo tiempo en las imágenes SAR la presencia de ruido speckle suele dificultar los procedimientos de inversión basados en minimizaciones directas. En este trabajo se presentan algunas estrategias basadas en modelos bayesianos y diferentes modelos directos conocidos. El enfoque propuesto permite tomar en cuenta el efecto del speckle y permite incorporar información a priori sobre la rugosidad que es posible obtener mediante otros sistemas. Se muestran los resultados de la aplicación de estas metodologías a imágenes obtenidas con el sistema aerotransportado SARAT de CONAE, y se contrastan los resultados con mediciones in-situ de humedad volumétrica.

**Palabras clave:** *humedad del suelo, problema inverso, SAR, métodos bayesianos.*

2000 AMS Subject Classification: 86A22, 62P12



Imagen SARAT del Centro Espacial Teófilo Tabanera (CETT).

En este trabajo estudiaremos el problema de estimar la humedad del suelo a partir de datos de un radar, que es la principal aplicación planteada para la misión SAOCOM. Más formalmente, se trata de un problema inverso para la constante dieléctrica. Estudiaremos este problema utilizando como modelos directos diferentes modelos aproximados disponibles en la literatura.

El principal desafío para resolver este problema es que la señal recibida depende tanto de la constante dieléctrica como de la geometría (rugosidad) de la superficie, que no es conocida de antemano. Como veremos más adelante, la sensibilidad a la rugosidad es tan significativa, que sin información sobre ella probablemente será imposible resolver el problema con la precisión requerida.

En el estudio de problemas inversos es importante disponer de datos reales. Los datos simulados pueden ayudar a entender mejor algunos aspectos del problema, pero las verdaderas dificultades aparecen con el uso de datos reales.

En ese sentido, se dispone de datos SAR adquiridos por el sistema aerotransportado SARAT de CONAE, y mediciones de humedad in-situ. Así mismo de información sobre la rugosidad de diferentes parcelas.

Compararemos los datos de humedad contra el resultado de la inversión de los datos de radar hechos con este esquema de inferencia, lo que nos permitirá saber si parece factible la aplicación de estimar la humedad del suelo con un radar, así como estimar el peso relativo entre los diferentes factores que limitan la precisión de las estimaciones en la etapa actual.

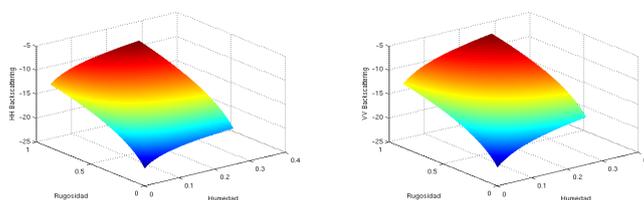
## 1. ANÁLISIS DE SENSITIVIDAD DEL PROBLEMA DIRECTO.

El modelo de Oh [5], utilizado para modelar el scattering en microondas de un suelo descubierto, es el siguiente:

$$HV = 0,11mv^{0,7} \cos(\theta)^{2,2} (1 - e^{-0,32ks^{1,8}})$$

$$VV = \frac{HV}{0,095 * (0,13 + \sin(1,5\theta))^{1,4} (1 - e^{-1,3ks^{0,9}})}$$

$$HH = VV * \left(1 - \left(\frac{2\theta}{\pi}\right)^{0,35mv^{-0,65}} e^{-0,4ks^{1,4}}\right)$$



En la inversión de datos reales - que no fueron generados por el modelo - hay que tener en cuenta el efecto que produce el 'ruido' en la inversión. En el caso de los problemas lineales, una herramienta que permite caracterizar estos efectos es la descomposición en valores singulares (SVD). Por lo tanto, en primer lugar linealizamos el problema y aplicamos la SVD.

**Propagación de las incertezas en los datos a la solución** Analicemos por ejemplo el modelo semiempírico de Oh, linearizandolo y usando la descomposición en valores singulares (SVD).

Si tenemos  $Kf = d + \epsilon$ , la solución a  $Kx = d$  será

$$x = f + \sum_k \frac{u_k^t \epsilon}{\sigma_k} v_k$$

En nuestro caso la SVD de la matriz es  $K = U\Sigma V$ , con

$$\Sigma = \begin{pmatrix} 0,0760 & 0 \\ 0 & 0,0046 \\ 0 & 0 \end{pmatrix}, V = \begin{pmatrix} -0,4883 & -0,8727 \\ -0,8727 & 0,4883 \end{pmatrix}$$

Al invertir este modelo, el 'ruido' se proyecta más hacia la dirección de la humedad que de la rugosidad. Este es un inconveniente que habrá que superar, pero antes veamos que tipo de ruido está presente en las imágenes de radar.

## 2. INCERTEZAS

### 2.1. RUIDO SPECKLE

En principio analicemos un tipo de ruido clásico presente en todas las imágenes SAR, conocido como 'Speckle'. Qué es el 'ruido' y que es 'la señal' es algo que depende del caso, y que le corresponde determinar al investigador.

La causa del 'ruido' llamado Speckle es que cada píxel de una imagen SAR (típicamente del orden de  $10 \text{ m}^2$ ) es, en los hechos, distinto de los otros por más que para el aplicador se trate de una misma zona homogénea. Veamos: es correcto considerar la geometría de un medio natural (por ejemplo una superficie) como un proceso aleatorio. Ese proceso está caracterizado por parámetros (ej. la desviación estándar de las alturas), y son éstos parámetros los que pueden considerarse constantes a lo largo de un área, que para el aplicador se tratará de un área constante ('un suelo de rugosidad  $x$ '). Pero sin embargo, al ser sistemas de alta resolución y muy sensibles a las características geométricas, estaremos viendo las diferencias entre las distintas 'realizaciones' de ese proceso.

## 2.2. CAMINATAS AL AZAR

El modelo más común para el speckle es el de la caminata al azar. Esto surge de la siguiente manera: podemos pensar que la energía que vemos en cada píxel de la imagen está compuesta por la suma de una cantidad de dispersores independientes. Es decir

$$z = \sum_i^N A_i e^{-2\pi i \theta_i}$$

Donde  $A_i$  y  $\theta_i$  son variables aleatorias, la primera típicamente normal y la segunda uniforme. Esto constituye una caminata al azar en el plano complejo.

Veremos que la intensidad (módulo) sigue una distribución Rayleigh bajo condiciones bastante generales. Sin embargo, en la práctica, suele usarse la distribución gamma, debido a que una imagen SAR pasa por varias etapas de procesamiento para reducir su varianza que no trataremos en el presente trabajo.

La variable gamma de alguna manera 'generaliza' a la distribución de Rayleigh, ya que

En la práctica se usa la distribución gamma, que 'generaliza' a la Rayleigh y tiene dos parámetros libres.

$$\sum_{i=1}^N \Gamma(k_i, \theta) \sim \Gamma\left(\sum_{i=1}^N k_i, \theta\right)$$

$$c\Gamma(k, \theta) = \Gamma(k, c\theta)$$

Hay expresiones simples para el promedio de gammas, que se utilizan para disminuir la incerteza a costa de perder resolución espacial.

El scattering se modela como  $Z = X.Y$ , donde  $Y$  es el ruido (de media uno y la varianza un parámetro libre), y  $X$  la predicción de los modelos directos. La varianza del speckle se representa con un parámetro  $N$  que se conoce en la literatura de radar como el 'número de looks', que vendría a indicar cuantas adquisiciones fueron promediadas para obtener el valor de cada píxel.

## 2.3. CÓMO SURGE LA DISTRIBUCIÓN RAYLEIGH

La deducción rigurosa a partir de un modelo físico - por más que contenga simplificaciones importantes - de una distribución estadística para los valores de una imagen SAR, es de gran valor porque permite interpretar mucha de la información contenida en las mismas, que de otro modo no sería posible. De hecho, un problema recurrente en el modelado de imágenes SAR consiste en que las distribuciones que ajustan bien los datos observacionales obtenidas mediante 'generalizaciones', agregan parámetros libres que no tienen un correlato físico claro, y por lo tanto tienen un significado muy vago.

Es interesante entonces el estudio probabilístico de 'dinámicas' (modelos físicos muy simples), que den lugar a diferentes distribuciones de probabilidad para los datos y que al mismo tiempo permitan asignar algún significado a los parámetros involucrados. En esta sección daremos una demostración moderna del resultado clásico antes citado de la distribución Rayleigh.

De la caminata al azar a la que nos referimos, la magnitud que mide el sistema es la intensidad (módulo al cuadrado), y es posible deducir su distribución de diversas formas. En Beckmann-Spizziccino [7], esto se hace trabajosamente mediante el teorema central de límite, y llama la atención una nota al pie sobre cómo había hecho esta cuenta originalmente Lord Rayleigh en el siglo XIX:

Rayleigh, who was not familiar with today's routine methods of probability theory, used an ingenious method of adding one more vector to the sum and transforming the resulting difference equation to a PDE, which turns out to be the equation of heat conduction (...) [Rayleigh, 1896, Section 42a]

Esto no puede dejar de llamarle la atención a cualquier persona familiarizada con la teoría moderna de difusiones y el cálculo de Itô.

Siguiendo a [2] (e indirectamente a Rayleigh), podemos usar la fórmula de Itô y el teorema de Feynmann-Kac, para obtener una PDE para la intensidad, cuya solución estacionaria resulta

$$P(z) = \frac{\alpha}{2} \exp\left(\frac{-\alpha z}{2}\right)$$

## 2.4. GAMMAS MULTIVARIADAS

En el caso de un sistema polarimétrico, los datos resultarán multivariados.

En dimensión 2 (que utilizaremos para tratar datos co-polarizados ignorando los cross-polarizados), hay una expresión analítica para una de tales distribuciones, conocida como la gamma de Kibble

$$pdf(x, y) = \frac{(\lambda_1 \lambda_2)^v}{(1 - \rho)\Gamma(v)} \left(\frac{xy}{\rho \lambda_1 \lambda_2}\right)^{\frac{v-1}{2}} \exp\left(-\frac{\lambda_1 x + \lambda_2 y}{\rho \lambda_1 \lambda_2}\right) I_{v-1}\left(\frac{2\sqrt{\rho \lambda_1 \lambda_2 xy}}{1 - \rho}\right)$$

(con  $I_{v-1}$  la función modificada de Bessel de tipo 1)

Si  $n > 2$ , no es posible obtener expresiones analíticas para la distribución, por lo menos para generalizaciones de la gamma de Kibble. Únicamente se puede escribir esta distribución como una transformada inversa de Laplace.

## 2.5. DESBALANCES Y CROSS-TALK

En un sistema SAR real hay otras clases de ruido además del Speckle, debido a desbalances y cross-talk, que los identificaremos como el error de ingeniería. Se pueden modelar (según [10]), como

$$\mathbf{M} = A \begin{pmatrix} 1 & \delta_1 \\ \delta_2 & f_1 \end{pmatrix} \begin{pmatrix} S_{hh} & S_{vh} \\ S_{hv} & S_{vv} \end{pmatrix} \begin{pmatrix} 1 & \delta_3 \\ \delta_4 & f_2 \end{pmatrix} + \begin{pmatrix} n_{hh} & n_{vh} \\ n_{hv} & n_{vv} \end{pmatrix}$$

Los valores de los parámetros se obtienen con una calibración con reflectores, y a partir de los errores residuales se estima el error de ingeniería. Sin embargo, no disponemos actualmente de una función de error caracterizada en términos estadísticos.

## 3. ESTIMADORES BAYESIANOS

Debido a las dificultades para lidiar con distribuciones gammas multivariadas antes mencionadas, y a las actuales limitaciones en nuestro conocimiento de las características estadísticas del error de ingeniería, lo más razonable parece ser considerar al error una variable normal, con una covarianza que es la suma de la covarianza del Speckle, más parámetros que caracterizarían el error de ingeniería. La hipótesis de normalidad para el Speckle se puede justificar a resoluciones bajas (mucho multi-looking) mediante el Teorema Central del Límite. De hecho, a continuación estudiaremos la resolución óptima del sistema y veremos que a altas resoluciones, el ruido Speckle genera demasiadas incertezas en la obtención de humedad, por lo que de todas formas estaremos en el caso en el que la hipótesis de normalidad se cumple aproximadamente.

Entonces, la función de verosimilitud podemos modelarla como:

$$\text{Likelihood}(Z|p) = \text{Norm}(Z - f(p), \Sigma)$$

- Z Mediciones
- p Parámetros del Modelo
- f Modelo Directo
- $\Sigma$  Covarianza del speckle + error de ingeniería

La distribución a posteriori, via el teorema de Bayes, resulta:

$$\text{Post}(p/Z) = \frac{\text{Likelihood}(Z/p)\text{Prior}(p)}{\int_{p \in P} \text{Likelihood}(Z/p)}$$

Y aquí hay dos estimadores a considerar: la media de la distribución a posteriori, y el máximo de la misma. El primer método es el estimador óptimo en la teoría bayesiana y el segundo está muy emparentado con los métodos de minimización.

Calcular los momentos de la distribución a posteriori involucra integrales de dimensión igual al número de parámetros ( $\dim(p)$ ), lo que puede resultar un proceso complejo en caso de que esta dimensión sea alta. En el caso del modelo de Oh y del IEM con  $\frac{s}{l}$  parametrizado, esta dimensión es 2, por lo que estos momentos pueden ser calculados fácilmente mediante integración numérica. Implementamos una regla de gauss de orden 100 en cada dimensión, con muy buenos resultados, y que además sus nodos y pesos pueden ser reutilizados todas las veces que sea necesario.

### 3.1. ANÁLISIS DE LA VARIANZA Y RESOLUCIÓN ÓPTIMA

La humedad la podemos considerar como un campo aleatorio. Su heterogeneidad depende de la escala espacial, y el clima, entre otros.

Según Famiglietti [4], la varianza de la humedad en un terreno depende de la media según

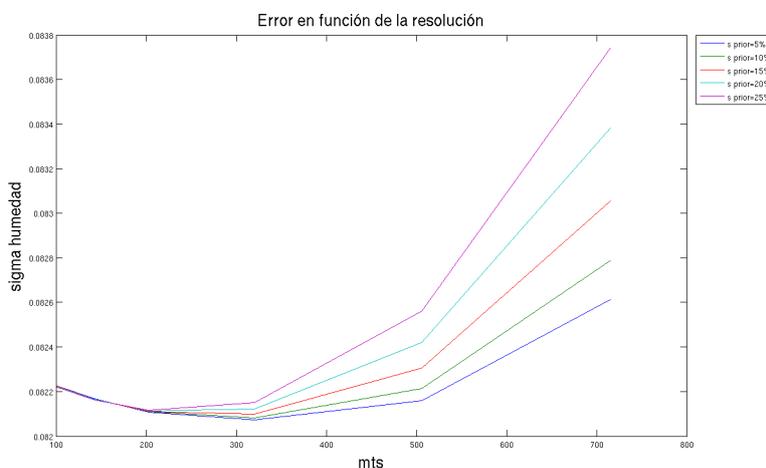
$$\sigma = k_1 \cdot \mu \cdot \exp^{-k_2 \mu}$$

donde  $k_1$  y  $k_2$  dependen de la escala espacial

escala	$k_1$	$k_2$
2.5 m	0.7803	9.0607
16 m	0.7287	7.3796
100 m	0.8941	8.0774
800 m	0.8840	5.8070
1.6 km	1.2070	7.1128
50 km	1.0429	5.2212

Hacemos retrieval de una 'constante efectiva', pero que será menos representativa si la varianza del campo aumenta. Al mismo tiempo, a escalas mayores disminuye el speckle porque lo promediamos. Suponiendo que la distribución a posteriori de la constante efectiva es independiente de la heterogeneidad, la varianza se suma linealmente.

$$\text{Var}(H) = \text{Var}(C_{ef}) + \text{Var}(Het)$$



#### 4. MÉTODO BAYESIANO PARA EL PROBLEMA INVERSO

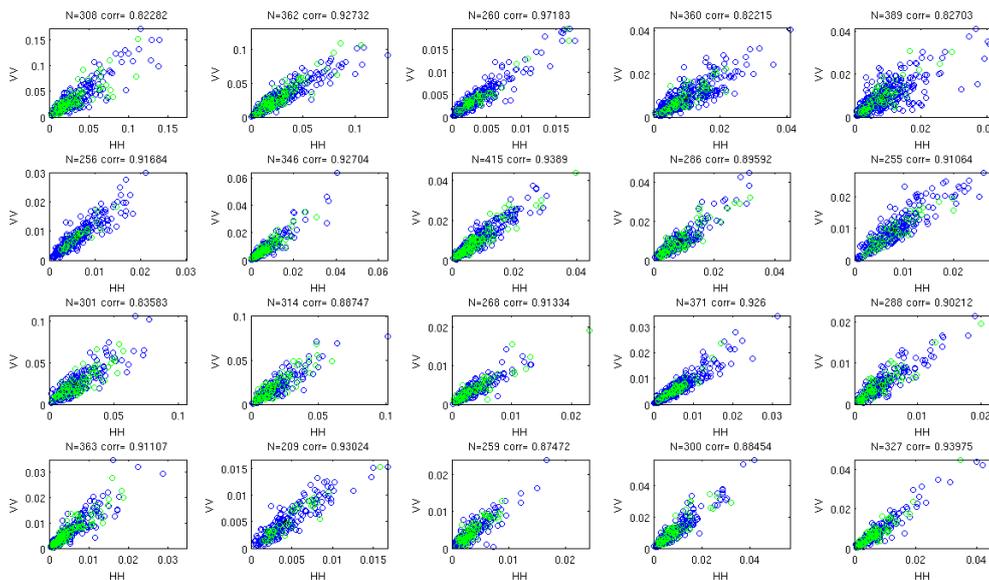
El enfoque estadístico (bayesiano) con el que trabajamos nos provee un marco que permite:

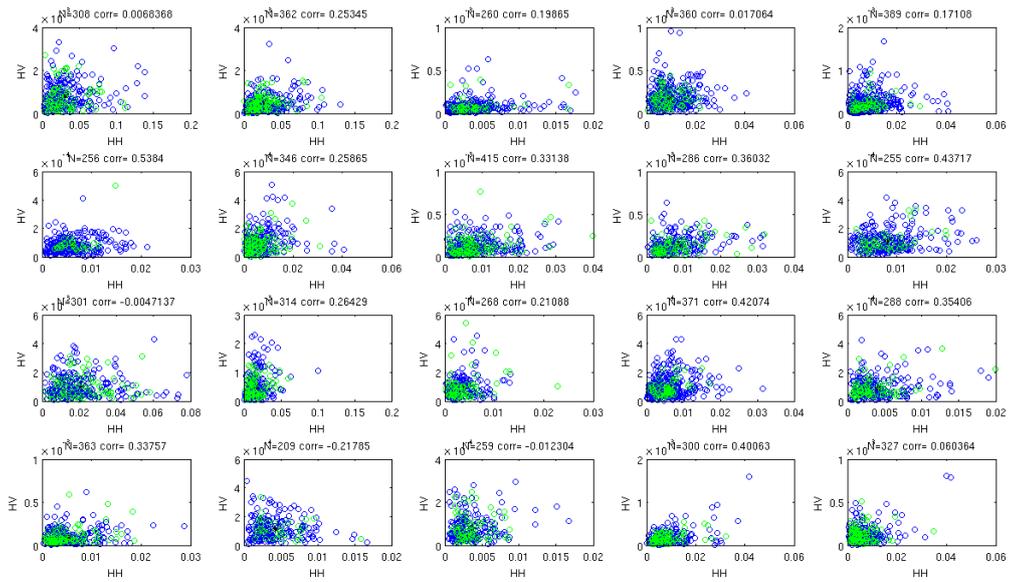
- Incorporar información a priori sobre la rugosidad, que no es sólo una ventaja sino una necesidad para la aplicación propuesta.
- Incorporar las incertezas inherentes a cualquier sistema SAR, y analizar fácilmente el efecto de las mismas en las estimaciones de humedad
- A partir de este análisis, determinar una resolución espacial óptima para el sistema
- Para pocos parámetros a obtener, como es el caso, estos estimadores se pueden implementar muy eficientemente con integración numérica.

#### 5. INVERSIÓN DE DATOS SARAT

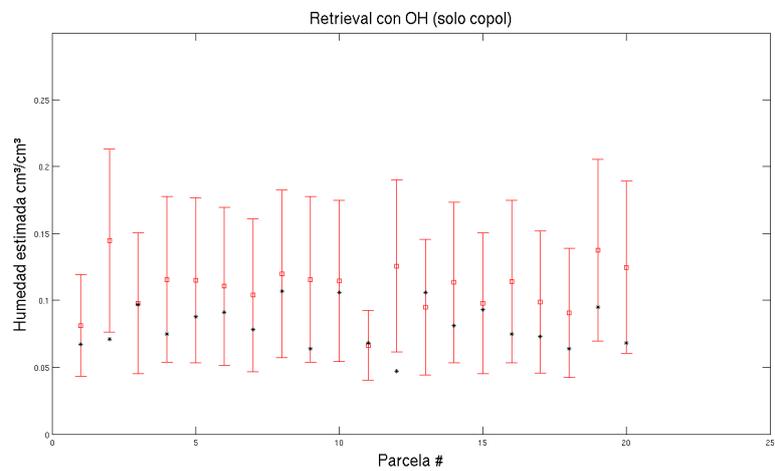
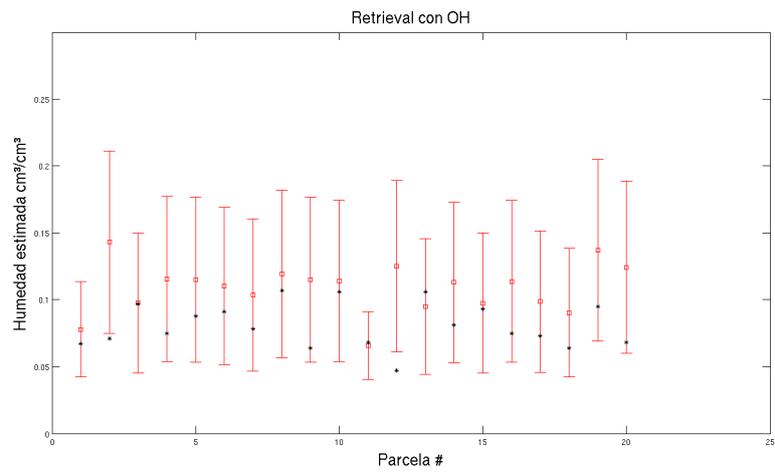
La implementación del esquema de inversión es la siguiente. En principio, dado que los datos tienen una correlación espacial, se procede a decorrelacionarlos de una manera eficiente, que es tomar promedios en cajas de tamaño creciente, hasta que la auto-correlación haya disminuido suficientemente. Los datos iniciales se muestran en azul, y los decorrelacionados en verde. Posteriormente se procede a estimar los parámetros del ruido speckle. Como estamos en valores grandes de  $N$ , puede usarse la aproximación normal. Luego se construyen los likelihoods sumando el error de medición propiamente dicho, y se calculan las integrales para los diferentes modelos forward.

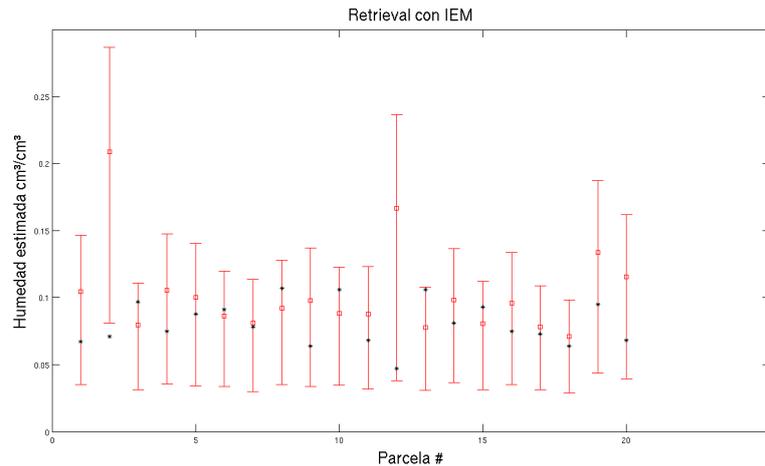
##### 5.1. DATOS





## 5.2. RESULTADOS





## 6. CONCLUSIONES Y PERSPECTIVAS

- Con información sobre la rugosidad, *es posible obtener buenas estimaciones de humedad con un radar*, al menos para suelos secos.
- También parece posible obtener buenas estimaciones *utilizando solamente datos co-pol*.
- Al mismo se analizaron los errores y se exploró *la resolución óptima* del sistema.

En principio, si bien no parece afectar al retrieval considerablemente, en los datos se observa que polarizaciones (HH y VV) son prácticamente iguales, mientras que los modelos predicen  $VV > HH$ . Falta explicar cuál es el efecto que causa esta anomalía. Además, se tiene conocimiento de que esta anomalía es aún más pronunciada en longitudes de onda más pequeñas, donde puede ocurrir  $HH > VV$ .

Esto marca la dificultad de un problema de gran interés, que es obtener información sobre la rugosidad con banda X (en la que opera el sistema italiano CosmoSkymed, parte de la misma constelación que el SAOCOM). En este caso, el problema de scattering es más desafiante. Por lo tanto, el esfuerzo futuro se centrará en resolver estos problemas de scattering con métodos numéricos eficientes.

A su vez, los modelos de back-scattering desarrollados en la literatura no dan cuenta de la correlación entre canales, por lo que este observable ha tenido que ser tomado como un parámetro libre en la distribución de speckle, sin ninguna relación con características de la superficie, lo que indica otro de los posibles caminos de mejora de dichos modelos.

## REFERENCIAS

- [1] M.Barber, Grings, F. ; Perna, P. ; Piscitelli, M. ; Maas, M. ; Bruscantini, C. ; Jacobo-Berlles, J. ; Karszenbaum, H. *Speckle Noise and Soil Heterogeneities as Error Sources in a Bayesian Soil Moisture Retrieval Scheme for SAR Data* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Junio 2012. Vol:5, issue:3.
- [2] Timothy R. Fields *Electromagnetic Scattering from Random Media*. Oxford University Press, 2009.
- [3] Albert Tarantola *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.
- [4] Famiglietti, J. S., D. Ryu, A. A. Berg, M. Rodell, and T. J. Jackson, *Field observations of soil moisture variability across scales*, Water Resour. Res., 44, W01423, 2008.
- [5] Y. Oh, *Quantitative retrieval of soil moisture content and surface roughness from multipolarized radar observations of bare soil surfaces*, IEEE Trans. Geosci. Remote Sens., vol. 42, no. 3, pp. 596–601, Mar. 2004.
- [6] Y. Oh, K. Sarabandi, F. T. Ulaby, *An empirical model and an inversion technique for radar scattering from bare soil surfaces* IEEE Trans. Geosci. Remote Sens., vol. 30, no. 2, pp. 370–381, Mar. 1992
- [7] P.Beckmann, A.Spizzichino. *The scattering of electromagnetic waves from rough surfaces* Pergamon Press, 1963.
- [8] J. L. Alvarez-Perez *An extension of the IEM/IEMM surface scattering model*, Waves in Random Media, vol. 11, pp.307 -329. 2001
- [9] A. K. Fung, W. Y. Liu, K. S. Chen, and M. K. Tsay, *An improved IEM model for bistatic scattering from rough surfaces*, J. Electromagn. Waves Applicat., vol. 16, no. 5, pp.689 -702. 2002.
- [10] A.Freeman, M. Alves, B. Chapman, J. Cruz, Y. Kim, S. Shaffer, J. Sun, E. Turner K. Sarabandi. *SIR-C Data Quality and Calibration Results* IEEE Transactions on Geosciences and Remote Sensing. 1995. Vol 33 no 4.