

UNIVERSIDAD NACIONAL DEL LITORAL



DOCTORADO EN INGENIERÍA

**Desagregación de Datos de Microondas Pasivas y Activas
para la Estimación de Humedad Superficial del Suelo en
Áreas de Llanura de la Región Pampeana**

Gabriel Agustín García

FICH

FACULTAD DE INGENIERÍA
Y CIENCIAS HÍDRICAS

INTEC

INSTITUTO DE DESARROLLO TECNOLÓGICO
PARA LA INDUSTRIA QUÍMICA

Tesis de Doctorado **2018**

Doctorado en Ingeniería
Mención Recursos Hídricos

Título de la obra:

Desagregación de datos de microondas pasivas y activas para la estimación de humedad superficial del suelo en áreas de llanura de la región Pampeana

Autor: Gabriel Agustín García

Lugar: Santa Fe, Argentina

Palabras Clave:

Humedad de suelo, Regresión lineal múltiple, Redes neuronales, Sentinel-1, SMAP, GPM, Áreas de llanura, Desagregación



UNIVERSIDAD NACIONAL DEL LITORAL
Facultad de Ingeniería y Ciencias Hídricas
Instituto de Desarrollo Tecnológico para la Industria Química

**DESAGREGACIÓN DE DATOS DE MICROONDAS
PASIVAS Y ACTIVAS PARA LA ESTIMACIÓN DE
HUMEDAD SUPERFICIAL DEL SUELO EN ÁREAS
DE LLANURA DE LA REGIÓN PAMPEANA**

Gabriel Agustín García

Tesis remitida al Comité Académico del Doctorado
como parte de los requisitos para la obtención
del grado de
DOCTOR EN INGENIERÍA
Mención Recursos Hídricos
de la
UNIVERSIDAD NACIONAL DEL LITORAL

2018

Secretaría de Posgrado, Facultad de Ingeniería y Ciencias Hídricas, Ciudad Universitaria, Paraje: "El
Pozo", S3000, Santa Fe, Argentina.



UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Ingeniería y Ciencias Hídricas

Instituto de Desarrollo Tecnológico para la Industria Química

**DESAGREGACIÓN DE DATOS DE MICROONDAS
PASIVAS Y ACTIVAS PARA LA ESTIMACIÓN DE
HUMEDAD SUPERFICIAL DEL SUELO EN ÁREAS
DE LLANURA DE LA REGIÓN PAMPEANA**

Gabriel Agustín García

Lugar de trabajo:

Centro de Estudios Hidro-Ambientales (CENEHA)

Facultad de Ingeniería y Ciencias Hídricas (FICH)

Universidad Nacional del Litoral (UNL)

Directora:

VENTURINI, Virginia

FICH-UNL

Codirectora:

RODRÍGUEZ, Leticia

FICH-UNL

Jurado evaluador:

Del Valle, Héctor

CENPAT-CONICET

Sione, Walter

CEREGeo-FCyT/UADER

Scavuzzo, Marcelo

CONAE

Thibeault, Marc

CONAE

2018



ACTA DE EVALUACIÓN DE TESIS DE DOCTORADO

En la sede de la Facultad de Ingeniería y Ciencias Hídricas de la Universidad Nacional del Litoral, a los un día del mes de marzo del año dos mil dieciocho, se reúnen los miembros del Jurado designado para la evaluación de la Tesis de Doctorado en Ingeniería titulada **“Desagregación de datos de microondas pasivas y activas para la estimación de humedad superficial del suelo en áreas de llanura de la región pampeana”**, desarrollada por el Ing Gabriel Agustín GARCÍA, DNI N.º 32.582.191 . Ellos son: Dr. Héctor del Valle, el Dr. Walter Sione, el Dr. Marc Thibeault y el Dr. Marcelo Scavuzzo.

Luego de escuchar la Defensa Pública y de evaluar la Tesis, el Jurado resuelve:

La tesis constituye un aporte al conocimiento en una temática actual y relevante como el de Humedad de suelo en la región Pampeana. Es original y en el manuscrito se incluyeron las observaciones y sugerencias realizadas por el jurado evaluador.

Al momento de la defensa, se presenta con solvencia y responde de manera apropiada las preguntas del jurado.

*El jurado asigna una calificación de 9 (nueve) **Distintivo**.*

Sin más, se da por finalizado el Acto Académico con la firma de los miembros del Jurado al pie de la presente.

.....
Dr. Walter Sione

.....
Dr. Marcelo Scavuzzo

.....
Dr. Héctor del Valle (*)

.....
Dr. Marc Thibeault (*)

Los Doctores Héctor del Valle y Marc Thibeault participaron por videoconferencia.

Dr. JOSÉ LUIS MACOR
SECRETARIO DE POSGRADO
Facultad de Ingeniería y Cs. Hídricas

Universidad Nacional del Litoral
Facultad de Ingeniería y Ciencias Hídricas
Secretaría de Posgrado

Ciudad Universitaria
C.C. 217
Ruta Nacional Nº 168 - Km. 472,4
(3000) Santa Fe
Tel: (54) (0342) 4575 229
Fax: (54) (0342) 4575 224 5
E-mail: posgrado@fich.unl.edu.ar

DECLARACIÓN LEGAL DEL AUTOR

La presente tesis ha sido remitida ante la comisión de Posgrado de la *Universidad Nacional del Litoral (UNL)* como parte de los requisitos para la obtención del grado académico de *Doctor en Ingeniería – Mención Recursos Hídricos*. Una copia de la misma permanecerá depositada en la biblioteca de la *Facultad de Ingeniería y Ciencias Hídricas (FICH)*, su consulta queda supeditada a la normativa legal vigente por el reglamento de la mencionada biblioteca.

Para fines educativos y de investigación, no se requiere autorización especial alguna en lo concerniente a reproducción, copia, distribución o citaciones de la presente tesis.

Cualquier sugerencia o comentario, sobre los tópicos desarrollados y discutidos en este trabajo de investigación, será altamente apreciada.

Gabriel Agustín García

*A mi hermana Florencia Camila.
A mi viejo Gabriel Carlos.
Conmigo siempre...*

Agradecimientos

Deseo expresar mi agradecimiento a todas las personas que de una manera u otra contribuyeron para que este trabajo de Tesis pudiera ser llevado a cabo, especialmente:

A mi directora, Dra. Virginia Venturini, por brindarme la oportunidad de iniciarme en la investigación científica. Por todo su tiempo dedicado a guiarme en este largo camino.

A mi madre, Marta, por apoyarme con la decisión a la hora de comenzar el doctorado y por acompañarme día a día. Sobre todo, por tenerme paciencia, por comprender mis tardes-noches de soledad en el altillo y aligerar mis momentos de desazón.

A Emiliano Pedro López, director de mi proyecto final de grado y luego mi compañero de trabajo. Gracias a él descubrí que la investigación era lo que buscaba para mi desarrollo profesional. Gracias por sus charlas enriquecedoras y sus consejos.

Finalmente, quiero agradecer a las siguientes instituciones:

- A la Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral y, en particular, al Centro de Estudios Hidro-Ambientales, por brindarme el lugar para realizar esta tesis.
- Al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), por la beca de formación de posgrado otorgada para realizar las actividades del doctorado.
- A la Comisión Nacional de Actividades Espaciales (CONAE), por proporcionarnos amablemente los datos *in situ* que pertenecen a su red de estaciones automáticas.

Gabriel Agustín García
Santa Fe, Noviembre 2017.

Índice general

Agradecimientos	IX
Resumen	XVI
Abstract	XVIII
Índice de figuras	XXI
Índice de tablas	XXIV
Listado de siglas	XXV
1. Introducción	29
1.1. Objetivos	36
1.1.1. Objetivo general	36
1.1.2. Objetivos específicos	36
1.2. Motivación	36
1.3. Organización de la tesis	37
2. Metodología	39
2.1. Ecuación de balance hídrico	39
2.2. Regresión lineal múltiple	41
2.3. Redes neuronales artificiales	43
2.4. Secuencia metodológica	48
2.4.1. Índices de rendimiento	48
3. Área de estudio y datos	51
3.1. Localidad de Ataliva - Provincia de Santa Fe	51
3.1.1. Ubicación	51
	XI

3.1.2. Datos	52
3.2. Localidad de Monte Buey - Provincia de Córdoba	54
3.2.1. Area de estudio	54
3.2.2. Datos	55
4. Estimación de HS con datos observados	61
4.1. Localidad de Ataliva	61
4.1.1. Método MLR	62
4.2. Localidad de Monte Buey	65
4.2.1. Calibración de los métodos MLR y MLP	73
4.2.2. Validación de los métodos	78
4.2.3. Discusión de los resultados	79
4.2.4. Análisis de incertidumbre del modelo	81
4.2.5. Mapas de humedad del suelo	82
5. Desagregación de mapas de HS de SMAP	89
5.1. Desagregación de HS-SMAP con PP , T_s , σ^0 y HR	95
5.1.1. Calibración método MLR	95
5.1.2. Calibración del método MLP	97
5.1.3. Contraste HS estimada versus HS de otras fuentes	98
5.2. Desagregación de HS-SMAP con PP , T_s , σ^0 y ET	100
5.2.1. Calibración del método MLR	100
5.2.2. Calibración método MLP	102
5.2.3. Contraste HS estimada versus HS de otras fuentes	103
5.2.4. Análisis de la variabilidad regional	106
6. Conclusiones y trabajos futuros	111
6.1. Conclusiones	111
6.2. Trabajos Futuros	114
Bibliografía	116
Anexos	133
A. Análisis de la consistencia de los datos	135
B. Implementación en Python	141

<i>Índice general</i>	XIII
C. Remuestreo mapas NDVI utilizando transformada ondita	153
D. Análisis de los productos satelitales utilizados	157
E. Soil moisture estimation over flat lands in Argentinean Pampas region using Sentinel-1A data and non-parametric methods - <i>International Jour- nal of Remote Sensing</i>	171

Resumen

La humedad de suelo (HS) es una variable de estado que influye en la redistribución del agua dentro del ciclo hidrológico. Durante décadas, su estimación a escala regional ha cobrado relevancia para resolver problemas hidrológicos, meteorológicos, agronómicos, climatológicos de diversa índole. En este sentido, el sensoramiento remoto brinda observaciones de esta variable a gran escala y de forma instantánea. Las microondas han potenciado el monitoreo de diferentes ambientes terrestres bajo condiciones adversas y han permitido desarrollar diferentes metodologías para obtener mapas de HS aprovechando la información complementaria que proveen tanto los sensores de microondas activos como pasivos. En este contexto, el objetivo de la tesis es desarrollar un procedimiento para estimar la HS bajo cobertura vegetal moderada en zonas de llanura, introduciendo la desagregación de información de microondas pasivas, radar y variables hidro-meteorológicas en el balance de agua. Para ello, se seleccionaron dos áreas de estudio y se desarrolló un modelo basado en la ecuación simplificada de balance hídrico que representa los procesos que influyen en la variabilidad de HS. El modelo tiene en cuenta los procesos de entrada y salida de agua del sistema suelo, y los representa con diferentes variables hidroambientales y datos radar. Para la aproximación de la ecuación de balance se seleccionaron las metodologías estadísticas de regresión lineal múltiple (MLR) y la red neuronal del tipo perceptrón multicapa (MLP). Los modelos resultantes fueron obtenidos conjuntamente con datos puntuales de precipitación, temperatura del aire y humedad relativa y, con datos radar de la misión satelital Sentinel-1. Los valores estimados presentaron un error medio cuadrático similar para ambos modelos, sugiriendo que cualquiera de ellos puede ser utilizado, aunque MLR permite analizar y cuantificar los errores introducidos por las variables. En base a los resultados parciales obtenidos, se aplicaron las mismas metodologías y supuestos para desagregar productos de HS de la misión SMAP. En este caso, se utilizaron datos de estaciones meteorológicas y productos satelitales para realizar la primer desagregación de HS-SMAP y posteriormente se utilizaron sólo

productos satelitales. En el primer caso, el modelo fusiona datos de humedad relativa (observaciones puntuales), precipitación de la misión satelital GPM, temperatura de superficie de la misión SMAP y datos radar. En el segundo caso, se utilizaron únicamente datos satelitales, reemplazando las observaciones de humedad relativa por el producto de evapotranspiración de la misión TERRA y AQUA de la NASA. Entre los principales resultados, se encontró que todos los métodos disminuyen el error base de SMAP en la zona de estudio, sin embargo, el método MLR parece ser más robusto que MLP.

Palabras clave:

Humedad de suelo, Regresión lineal múltiple, Redes neuronales, Sentinel-1, SMAP, GPM, Áreas de llanura, Desagregación

Abstract

Soil moisture (SM) is a state variable that influences the redistribution of water within the hydrological cycle. For decades, its estimation has been studied on a regional scale to solve hydrological, meteorological, agronomic and climatological problems. In this sense, remote sensing provides instantaneously observations of this variable on a large scale. Microwaves have enhanced the monitoring of different terrestrial environments under adverse conditions and have allowed the development of different methodologies to obtain SM maps taking advantage of the complementary information provided by both active and passive microwave sensors. In this context, the thesis objective is to develop a procedure to estimate SM under moderate vegetation cover in plain areas, introducing the disaggregation of passive microwave information, radar and hydro-meteorological variables in the water balance. For this, two study areas were selected and a model was developed based on the water balance equation that represents the processes that influence in SM variability. The model takes into account water input and output processes of the soil system, and represents them with different hydro-environmental variables and radar data. For the resolution of the balance equation, multiple linear regression (MLR) and multilayer perceptron (MLP) statistical methodologies were selected. The resulting models were obtained with precipitation, air temperature and relative humidity observations and with radar data from the Sentinel-1 satellite mission. Similar mean square error were reached for both models, suggesting that any of them can be used, although MLR allows analyzing and quantifying the errors introduced by the variables. Based on the partial results obtained, the same methodologies and assumptions were applied to disaggregate SM SMAP product. In this case, data from meteorological stations and satellite products were used to carry out the first SM SMAP disaggregation and later only satellite products were used. In the first case, the model merges relative humidity (point observations), precipitation data from the GPM satellite mission, SMAP surface temperature and radar data. In the second case, only satellite data were used, replacing

relative humidity observations with evapotranspiration product of NASA's TERRA and AQUA mission. Among the main results, it was found that all the methods reduce the SMAP base error in the study area, however, the MLR method seems to be more robust than MLP.

Keywords:

Soil moisture, Multiple linear regression, Neural networks, Sentinel-1, SMAP, GPM, flat lands, Disaggregation

Índice de figuras

2.1. Estructura de una neurona: a) Biológica; b) Artificial	44
2.2. Arquitectura de la red neural MLP utilizado para la predicción	45
2.3. Diagrama de flujo del procedimiento propuesto para la estimación de HS.	48
3.1. Ubicación del área de estudio Ataliva	52
3.2. Área de estudio Monte Buey. Izquierda: Ubicación; Derecha: Puntos de medición <i>in situ</i> ; datos meteorológicos (rombo); datos de HS (círculos)	55
4.1. Modelo MLR con retrodispersión.	64
4.2. Modelo MLR con retrodispersión+precipitación.	64
4.3. Modelo MLR Retrodispersión + Precipitación + Humedad relativa del aire.	65
4.4. Integración de resoluciones espaciales.	66
4.5. Variación temporal de las variables.	69
4.6. Matriz de correlación entre HS CONAE, T_a , PP, σ^0 , e_a , HR y T_s .	71
4.7. Clasificación de la cobertura con el producto NDVI MOD13Q1 a 30m: agua (Azul); suelo desnudo (Gris); vegetación escasa (Amarillo); vegetación densa (Verde).	72
4.8. Relación entre PP y HS.	74
4.9. Comparación entre los valores de HS observados y estimados que resultan en la fase de calibración para los métodos MLR (Círculos) y MLP (Triángulos).	78
4.10. Comparación entre los valores de HS observados y estimados que resultan en la fase de validación para los métodos MLR (Círculos) y MLP (Triángulos).	79

4.11. Imágenes de clases para cuatro fechas de Landsat 8. Negro: píxeles inutilizables; Blanco: píxeles donde se aplicó el modelo.	84
4.12. Mapas de HS sobre el área de estudio: (a) HS obtenida con MLR, (b) HS utilizando MLP; HS (escala de Verdes); Cuerpos de agua (Azul); Ciudades (Rojo); Píxeles enmascarados donde no se aplican los modelos (Blanco).	86
5.1. Comparación entre PP-INTA y PP-GPM en milímetros acumulados en siete días.	90
5.2. Variabilidad temporal de PP-INTA y PP-GPM en milímetros acumulados en siete días. PP-INTA (Verde); PP-GPM (Azul).	91
5.3. Integración de las resoluciones espaciales de las imágenes HS-SMAP y T_s -SMAP, PP-GPM, σ^0 -Sentinel y ET -Modis y NDVI-Modis.	92
5.4. Matriz de correlación entre HS-SMAP, T_s -SMAP, ET , HR, PP-GPM y σ^0	94
5.5. Gráficos de dispersión entre HS de SMAP y HS estimada utilizando HR con los métodos MLR (Círculos) y MLP (Triángulos) durante las fases de: (a) calibración y (b) validación.	97
5.6. Evolución temporal de PP-GPM (Barras rojas) y HS-CONAE (Líneas de punto Negro). Las líneas de trazo de colores corresponden a la diferencia entre la HS observada y la estimada utilizando HR por los métodos MLR (Línea de trazo y punto Azul) y MLP (Línea de trazo Verde).	100
5.7. Gráficos de dispersión entre HS de SMAP y HS estimada utilizando ET con los métodos MLR (Círculos) y MLP (Triángulos) durante las fases de: (a) calibración y (b) validación.	102
5.8. Evolución temporal de PP-GPM (Barras rojas) y HS-CONAE (Líneas de punto Negro). Las líneas de colores corresponden a la diferencia entre la HS observada y la estimada utilizando ET por los métodos MLR (Línea de trazo y punto Azul) y MLP (Línea de trazo Verde).	104
5.9. Diagrama de Taylor para el desempeño de las metodologías. SMAP (cuadrado); CONAE (estrella); MLR con HR (círculo); MLP con HR (triángulo hacia arriba); MLR con ET (triángulo hacia abajo); MLP con ET (diamante).	105

5.10. HS para la localidad de Monte Buey (Córdoba), comparación entre	
(a) SMAP, desagregada utilizando ET (b) con MLR, (c) con MLP, (d)	
resultado del modelo VIC.	108
A.1. Boxplot de la variación de HR de INTA respecto a la media regional	136
A.2. Boxplot de la variación de T_a de INTA respecto a la media regional	136
A.3. Boxplot de la variación de e_a de INTA respecto a la media regional	137
A.4. Boxplot de la variación de PP de INTA respecto a la media regional	138
A.5. Boxplot de la variación de HS de CONAE respecto a la media regional	139
A.6. Boxplot de la variación de T_s de CONAE respecto a la media regional	139
C.1. Esquema transformada ondita discreta en dos dimensiones (DWT -2D)	154
C.2. Rescalado con transformada ondita inversa (IDWT)	154

Índice de tablas

3.1. Inventario de las imágenes ópticas y SAR	54
3.2. Inventario de las imágenes Sentinel-1, GPM, HS observado, VIC, SMAP y MODIS MOD / MYD 16.	59
4.1. Coeficientes, errores, estadístico t y p-valor de la ecuación final obtenida con MLR.	62
4.2. Estadísticos de las variables de entrada: mínimo (Mín), máximo (Máx), media y S.	70
4.3. Estadísticas de las variables de entrada seleccionadas.	75
4.4. Parámetros de los métodos de optimización de pesos del MLP.	76
4.5. Pesos sinápticos (W_{ij}) para el método MLP con arquitectura 4-3-4-1.	77
4.6. Resumen de las metodologías para la estimación de HS a partir de datos SAR.	80
4.7. Media y S de las variables de entrada para el análisis de primer orden.	82
4.8. Inventario de las imágenes para generar mapas de HS.	83
4.9. Análisis estadístico de los mapas de HS	87
5.1. Estadísticas de PP-INTA y PP-GPM en milímetros.	90
5.2. Estadísticas de las variables de entrada seleccionadas del primer conjunto de datos.	96
5.3. Pesos sinápticos (W_{ij}) para el método MLP con arquitectura 4-5-4-1.	98
5.4. Estadísticas de las variables de entrada seleccionadas del segundo conjunto de datos.	101
5.5. Pesos sinápticos (W_{ij}) para el método MLP con arquitectura 4-3-6-1.	103
5.6. RMSE de los modelos MLR y MLP con <i>HR</i> y <i>ET</i> durante las fases de calibración, validación y prueba.	106
5.7. RMSE y SSIM entre los mapas obtenidos con MLR y MLP, y los mapas de SMAP y VIC.	110

C.1. Inventario de las imágenes NDVI MODIS y Landsat 8	155
C.2. RMSE y R^2 para distintas familias onditas	155

Listado de siglas

AMSR-E Advanced Microwave Scanning Radiometer for EOS

ANN Artificial Neural Network

CONAE Comisión Nacional de Actividades Espaciales

DS Desvío estándar

EOS Earth observing system

ESA European Space Agency

ET Evapotranspiración

e_a Tensión de vapor

FIV Factor de Inflación de la Varianza

GD Gradiente Descendiente

GOM Geometrical Optics Model

GPM Global Precipitation Measurement

GRD Ground Range Detected

HS Humedad de Suelo

HR Humedad Relativa

IEM Integral Equation Model

IDW Inverse Distance Weight

INTA Instituto Nacional de Tecnología Agropecuaria

- IMERG** Integrated multi-satellite retrievals for GPM
- JAXA** Japan Aerospace Exploration Agency
- LAFDM** Latin American Flood and Drought Monitor
- MODIS** Moderate Resolution Imaging Spectroradiometer
- MLR** Multiple Linear Regression
- MLP** Multilayer perceptron
- NASA** National Aeronautics and Space Administration
- NDVI** Normalized Difference Vegetation Index
- PP** Precipitación
- PDF** Probability Distribution Function
- SAR** Synthetic Aperture Radar
- SAOCOM** Satélite Argentino de Observación COon Microondas
- SMOS** Soil Moisture and Ocean Salinity
- SMAP** Soil Moisture Active and Passive
- SVR** Support Vector Regression
- RELU** Rectified linear unit
- RMSE** Root Mean Square Error
- SNAP** Sentinel application platform
- SPM** Small Perturbation Model
- SLC** Single Look Complex
- TB** Temperatura de brillo radiométrica
- T_a** Temperatura del aire
- T_s** Temperatura del suelo

TRMM Tropical Rainfall Measuring Mission

VIC Variable Infiltration Capacity model

Capítulo 1

Introducción

El ciclo del agua, también denominado ciclo hidrológico, describe el movimiento continuo y cíclico del agua en el planeta Tierra. En este contexto se investiga cómo el agua se mueve en sus diferentes fases a través de la atmósfera, hacia abajo y a través de la litósfera, hacia el océano y de regreso a la atmósfera (Brutsaert, 2005).

El ciclo hidrológico está dominado por procesos como la escorrentía, la infiltración y la evaporación-*evapotranspiración* (*ET*). Parte del agua que llega a la superficie terrestre en forma de precipitación (*PP*) escurrirá por el terreno hasta llegar a un río, un lago o el océano, a este fenómeno se lo conoce como escorrentía. Otro porcentaje del agua se filtrará a través del suelo y eventualmente llegará a los acuíferos, en particular a la capa freática. Este proceso es la infiltración. Tarde o temprano, toda este agua volverá nuevamente a la atmósfera, debido a la *ET*, definida como la pérdida de humedad de la superficie por evaporación junto con la pérdida de agua por transpiración de la vegetación (Brutsaert, 2005). La humedad del suelo (*HS*) es una variable de estado clave en este ciclo porque influye en la redistribución del agua entre estos procesos.

La *HS* también juega un rol vital en el funcionamiento de los ecosistemas, ya que no solo influye en el ciclo meteorológico sino que también es una componente fundamental para el rendimiento de los cultivos (Behari, 2005). Estrictamente, la *HS* puede definirse como un almacenamiento temporal dentro de una capa poco profunda de la tierra que generalmente se limita a la zona de aireación (Engman, 1991), y depende de la profundidad, de la textura y de la estructura del suelo. Este almacenamiento puede oscilar aproximadamente entre 30-50 mm en suelos arenosos y superficiales hasta más de 200 mm en suelos profundos, arcillosos y franco-arcillosos (Verheyne, 2004).

La considerable variabilidad espacio-temporal de *HS* implica un desafío para su precisa estimación y predicción. Dicha variabilidad se manifiesta tanto en dirección

vertical como lateral, generada por la *PP* y la *ET*, e influenciada por la topografía, la textura del suelo y la vegetación. Si bien las variaciones espaciales a escala pequeña están afectadas por la textura del suelo, las escalas más grandes están dominadas por la precipitación y la evaporación (Pandey and Pandey, 2010; Del Valle et al., 2011), siendo la topografía más importante en períodos húmedos. Sin embargo, durante épocas secas, la distribución de la HS depende sobre todo de las características del suelo (Grayson et al., 1997; Western, 1999).

La estimación de HS es importante para aplicaciones hidrológicas, meteorológicas y agrícolas, por ejemplo, en la predicción de desastres naturales tales como inundaciones y sequías. No obstante, a pesar de que la HS constituye una variable indispensable para tener en cuenta en los estudios medioambientales, su determinación a escala regional y periódica es difícil a través de mediciones puntuales, las cuales son escasas, complejas y costosas (Mallick et al., 2009).

El sensoramiento remoto ha tomado gran interés en las últimas décadas, mostrando un gran potencial en la estimación de HS a escala regional y global de forma instantánea (Jackson and Schmugge, 1991). En este sentido, el uso de las microondas ha potenciado el monitoreo de diferentes ambientes terrestres en condiciones climáticas adversas, tal como se puede apreciar en el siguiente sitio web SAREDU (<https://saredu.dlr.de/>, último acceso Febrero 2017). En efecto, en dicho portal se encuentran valiosos aportes metodológicos para el monitoreo de diversos ambientes, así como también herramientas de procesamiento y datos.

Los datos remotamente obtenidos con sensores de microondas activos (radar) y pasivos (Radiómetro) han probado ser los mejores instrumentos para cuantificar la HS en los primeros centímetros de suelo (Ulaby et al., 1982; Jackson and Schmugge, 1991; Engman and Chauhan, 1995). Los radiómetros detectan la radiación en las frecuencias de microondas emitida naturalmente por la superficie terrestre y obtienen la temperatura de brillo radiométrica (TB). Estos dispositivos poseen una óptima sensibilidad a la HS, inclusive en presencia de moderada vegetación. Sin embargo, presentan la desventaja de tener una resolución espacial del orden de las decenas de kilómetros debido a la baja cantidad de energía de microondas emitida por la superficie de la tierra. Un antecedente muy importante para el desarrollo de metodologías para determinar HS con radiómetros ha sido el sensor Advanced Microwave Scanning Radiometer (AMSR-E) a bordo del satélite EOS-Aqua. Las frecuencias que cubre este sensor comprenden el rango 6.9 – 89 GHz (Njoku et al., 2003). Paralelamente, la tecnología de microondas pasivas en banda L (1.4 Ghz – 20 cm) ha demostrado su capacidad para obtener ma-

pas de HS a escala regional con una alta resolución temporal (Schmugge and Jackson, 1994; Njoku and Entekhabi, 1996; Jackson et al., 1999). Además, dicha banda está menos afectada por la atmósfera y la cobertura vegetal (Kerr et al., 2001). Los sensores modernos, como el de la misión Soil Moisture and Ocean Salinity (SMOS) y la misión Soil Moisture Active Passive (SMAP), operan en banda L y poseen una resolución espacial de aproximadamente 40 km (Entekhabi et al., 2010; Kerr et al., 2010). Esta resolución espacial gruesa limita la aplicabilidad de los datos a una escala menor que la escala de la cuenca.

En contraste, los sensores activos como el radar cuentan con su propia fuente de microondas que ilumina la superficie terrestre y registran la diferencia entre la onda electromagnética emitida y la recibida, denominada retrodispersión (σ^0). Esta capacidad permite el monitoreo de variables tales como la HS, a una resolución espacial de metros a decenas de metros (Ulaby et al., 1982; Fung, 1994). La sensibilidad de estos sensores a la HS, se ve reducida por la influencia de la rugosidad de la superficie, la estructura de la vegetación y su contenido de agua.

La profundidad de penetración de estas tecnologías, pasivas y activas, es del orden de un décimo de la longitud de onda, dependiendo del perfil de humedad del suelo (Ulaby et al., 1986), es decir, en longitudes de onda más largas se exploran capas de suelo más profundas.

Por un lado, en los últimos diez años la evolución tecnológica de los sensores radar denominados Synthetic Aperture Radar (SAR), ha permitido el desarrollo de varias misiones satelitales que proporcionan información en diferentes frecuencias de microondas. Por ejemplo, las misiones TerraSAR-X y Cosmo-SkyMed operan en banda X (9 GHz - 3 cm); las misiones ERS-1, ERS-2, Envisat, Sentinel-1, RADARSAT-1 y RADARSAT-2 operan en banda C (6 GHz - 5 cm) y ALOS, ALOS-2 y SMAP operan en banda L (1,4 GHz - 20 cm). Por otro lado, la principal limitación para obtener HS con estos sensores es la complejidad de la señal retrodispersada (Loew and Mauser, 2006; Mattia et al., 2006; Shi et al., 1997). Ulaby et al. (1996) definieron la retrodispersión total co-polarizada (σ_{pp}) desde una superficie, que es la señal que recibe el radar, como la suma de tres componentes:

$$\sigma_{pp} = t^2 * \sigma_{soil} + \sigma_{veg} + \sigma_{int} \quad (1.1)$$

donde el primer término es la dispersión del suelo, σ_{soil} , modificada por la atenuación de la capa de vegetación t^2 . El segundo y tercer término representan la dispersión desde el volumen de vegetación, σ_{veg} , y la interacción entre la vegetación y la superficie del

suelo, σ_{int} , respectivamente (Ulaby et al., 1996). Para suelos desnudos o con poca vegetación ($t^2 \approx 1$), la contribución de σ_{soil} domina la señal recibida por el radar, siendo σ_{pp} determinada por la HS y la rugosidad del terreno (R), debido a que $\sigma_{soil} = f(R, HS)$. En caso de superficies con vegetación ($t^2 \approx 0$), σ_{pp} es determinada en gran parte por la dispersión del volumen de vegetación por lo que, para estimar la HS a partir de σ_{pp} se deben analizar los términos σ_{veg} y σ_{int} .

Sobre suelos desnudos se han desarrollado modelos para estimar HS basados en la ecuación de retrodispersión y en métodos de inversión. Los modelos denominados de retrodispersión permiten establecer una relación directa entre los parámetros de la superficie y la radiación retrodispersada, y son muy útiles para entender la sensibilidad de la respuesta del radar a los cambios de esas variables biofísicas (Barrett et al., 2009). Estos métodos de inversión también son usados para determinar parámetros de la superficie potencialmente desconocidos a partir de la señal retrodispersada registrada por el satélite (Kornelsen and Coulibaly, 2013).

En general, los modelos de retrodispersión se pueden clasificar en modelos teóricos, empíricos y semi-empíricos (Barrett et al., 2009). Los modelos de retrodispersión teóricos simulan la retrodispersión derivada de una configuración de radar y una geometría particular. Los mismos están basados en respuestas aparentes a una determinada longitud de onda y resolución del radar. Estos modelos pueden utilizarse en una amplia variedad de condiciones, debido a que hacen pocas suposiciones *a priori* sobre las características de la superficie en comparación con los modelos empíricamente derivados.

Los modelos empíricos relacionan el coeficiente de retrodispersión del radar con HS mediante relaciones lineales. Es decir, la relación empírica entre el coeficiente de retrodispersión del radar (σ^0), medida en decibeles (dB), y la humedad volumétrica del suelo (HS_{vol}) para las superficies con suelo desnudo está dada por (Oh et al., 1992; Zribi et al., 2005; Holah et al., 2005; Baghdadi et al., 2012):

$$\sigma^0 = \delta HS_{vol} + \xi \quad (1.2)$$

donde el coeficiente δ depende de los parámetros del sensor SAR (longitud de onda del radar, ángulo de incidencia y polarización), mientras que el coeficiente ξ está controlado por los parámetros del sensor SAR y la rugosidad de la superficie. Si bien las relaciones derivadas permiten determinar adecuadamente la HS a partir de la retrodispersión radar, estos modelos empíricos son difíciles de generalizar para diferentes suelos debido a que los coeficientes que describen la relación encontrada a menudo

son diferentes de una cuenca a otra y también de un año a otro (Dubois and Engman, 1995; Le Hegarat-Masclé et al., 2002).

Los modelos semi-empíricos combinan la simplicidad de los modelos empíricos con una base teórica que proporciona un compromiso entre la representación teórica y la física de los procesos. Estos modelos son ampliamente aplicables ya que se derivan de múltiples mediciones de la superficie. Dichas mediciones se realizan empleando un dispersómetro, o difusómetro de radar, bajo diversas configuraciones del radar y condiciones de superficie (Oh et al., 1992; Dubois and Engman, 1995). La dificultad que se suele encontrar en el uso de modelos semi-empíricos es la parametrización de la rugosidad superficial y la vegetación (Dubois and Engman, 1995; Baghdadi et al., 2006; Merzouki et al., 2011). Dentro de estos modelos, la regresión lineal múltiple (MLR, por su sigla en inglés) es uno de los algoritmos más directos y simples para relacionar el coeficiente de retrodispersión con HS. La base del modelo MLR es encontrar la relación entre varias variables independientes o predictoras y una variable dependiente o criterio (Weisberg, 2005). Los modelos simples de MLR han sido aplicados desde 1982 por Ulaby et al. (1982) para relacionar las propiedades físicas del suelo con la de la señal radar. Otros autores describieron la relación entre HS y la temperatura de brillo de los radiómetros (Narayan et al., 2004; Paloscia et al., 2006; Sobrino et al., 2012; Sánchez et al., 2014; Al-Yaari et al., 2016; Al-Yaari et al., 2017).

Por otra parte, dentro de los métodos de inversión se encuentran modelos basados en técnicas de detección de cambios, de inversión estadística y de redes neuronales, entre otros. Las técnicas de detección de cambios se emplean para eliminar el efecto de la rugosidad superficial en los modelos semi-empíricos. Estos métodos suponen la rugosidad superficial invariante en el tiempo, pero no fijan *a priori* la escala y la distribución espacial de la misma. La eliminación de la rugosidad en las imágenes de radar permite establecer una relación empírica entre la señal retrodispersada y la HS (Moran et al., 2000; Zribi et al., 2005; Kim and van Zyl, 2009; Pathe et al., 2009). Además, a menudo se supone que el crecimiento de la vegetación no cambia en el intervalo de tiempo entre las adquisiciones (Joseph et al., 2008), de lo contrario la vegetación se explica explícitamente en el modelo (Wickel et al., 2001).

La inversión estadística se utiliza para crear una función de distribución de probabilidad (PDF, por su sigla en inglés) a partir de la cual se puede determinar la probabilidad estadística de un determinado valor de HS en base a las propiedades del suelo y de la señal retrodispersada. Las PDF se suelen asociar a los métodos bayesianos, cuando se usa un modelo directo para encontrar la función de probabilidad (Notarnicola et al.,

2008).

Dentro de los modelos de inversión también se utilizan las redes neuronales artificiales (ANNs, por su sigla en inglés). Las ANNs vinculan un conjunto definido de entradas a través de una matriz de peso con una capa oculta compuesta de neuronas artificiales. Se considera que una ANN es un aproximador universal de cualquier función no lineal y se ha utilizado con éxito para invertir modelos de HS (Said et al., 2008; Paloscia et al., 2010; Merzouki et al., 2011; Kseneman et al., 2012).

Aunque los sensores pasivos y activos muestran capacidades complementarias útiles para la cuantificación de HS, el principal inconveniente para fusionar ambas fuentes de datos es la disparidad de sus resoluciones espaciales y temporales (Del Valle et al., 2011). Para aprovechar las bondades de los radiómetros y los radares, se han propuesto diferentes metodologías de fusión. La combinación de datos pasivos y activos proporciona estimaciones de HS precisas y con alta resolución espacial, listas para ser asimiladas por modelos regionales hidrológicos y agronómicos.

Como se mencionó anteriormente, los datos de microondas pasivos se adquieren usualmente con una resolución espacial más gruesa que los datos SAR; para fusionarlos es necesario reducir su resolución espacial a resoluciones más finas, operación usualmente denominada desagregación. Según Peng et al. (2017), hay tres grupos generales de métodos propuestos para fusionar imágenes de sensores activos y pasivos para obtener HS. El primer grupo incluye la desagregación de HS derivada de radiómetro con datos de retrodispersión SAR (Narayan et al., 2006; Piles et al., 2009; Das et al., 2011). El segundo grupo incluye la desagregación de TB con datos de retrodispersión y su inversión a HS (Das et al., 2014). Finalmente, el tercer grupo comprende la fusión de productos HS de sensores activos y pasivos (Montzka et al., 2016). Dentro de estos grupos, se encuentran algoritmos tales como detección de cambios (Njoku et al., 2002), métodos bayesianos (Zhan et al., 2006), ANNs, algoritmo lineal ponderado (LWA), MLR (Srivastava et al., 2013) y filtros de Kalman. Varios estudios han intentado reducir la escala de los productos HS de microondas pasivas con la ayuda de información sobre la cubierta vegetal y la temperatura de la superficie, así como otros parámetros de superficie obtenidos a partir de datos ópticos y térmicos. Merlin et al. (2008), Piles et al. (2011), Merlin et al. (2012), Piles et al. (2014), Molero et al. (2016) y Chen et al. (2017) han propuesto diferentes metodologías para desagregar SMOS utilizando estos datos auxiliares. Por otro lado, Das et al. (2011) han desarrollado un algoritmo de detección de cambios para desagregar HS basado en radiómetros con retrodispersión radar.

La misión SMAP, dedicada explícitamente a proporcionar mapas globales de HS, posee a bordo un radar y un radiómetro, ambos en banda L, que exploran la superficie terrestre (Entekhabi et al., 2010). Esta característica de llevar a bordo simultáneamente ambos sensores permitiría generar un producto HS con una resolución mejorada, de 36 a 9 y 3 km, respectivamente, con una cobertura global. El fallo del radar en Das et al. (2016) y Entekhabi et al. (2016) el satélite SMAP indujo a buscar otras frecuencias de microondas disponibles para obtener una estimación mejorada de HS. propusieron el uso de datos Sentinel-1 en el algoritmo activo-pasivo de SMAP, obteniendo de esta manera HS desagregado. Santi et al. (2016) han evaluado el uso de SMAP, AMSR2 y Sentinel-1 como entradas de un algoritmo basado en ANNs para monitorear HS. Por otro lado, Lievens et al. (2017) investigaron la asimilación de observaciones Sentinel-1 en el modelo GEOS-5 del Sistema de Observación de la Tierra Goddard para simulaciones hidrológicas, ya sea por separado o simultáneamente con observaciones del radiómetro SMAP. Alexakis et al. (2017) han investigado el potencial de ANNs para determinar HS con Sentinel-1 y Landsat-8. Recientemente, diferentes métodos de regresión han sido utilizados para la fusión temperatura de superficie (T_s) de diferentes fuentes de datos (Bisquert et al., 2016; Olivera-Guerra et al., 2017). Sin embargo, en la literatura consultada no se encontraron aplicaciones de MLR para la fusión HS con PP, como se propone en esta tesis.

La mayoría de los modelos disponibles para estimar HS requieren de información de las propiedades físicas del suelo y su composición mineral, las que son escasas y no siempre se encuentran en formato digital en la Argentina, y en muchas otras áreas remotas del mundo. En oposición, los datos meteorológicos son accesibles desde diferentes agencias gubernamentales. Los algoritmos de inversión están más enfocados en explotar la información de retrodispersión (es decir, intensidad y textura) que en considerar datos de variables meteorológicas. Por otra parte, desde el punto de vista hidrológico, la variabilidad de HS se explica mediante las entradas y salidas de agua hacia y desde el sistema suelo (Del Valle et al., 2011). Los modelos hidrológicos cuantifican los procesos hidro-meteorológicos para estimar un balance entre los ingresos de agua al sistema, por medio de la PP y la nieve, y los egresos de agua por medio de la ET , el escurrimiento (Q) y la recarga de agua subterránea. De hecho, los modelos de balance hídrico relacionan la variabilidad de HS con los procesos hidro-meteorológicos mediante relaciones no lineales (Alley, 1984; Liang et al., 1994; Orth et al., 2013; Brocca et al., 2014). Hasta el presente, diversos modelos hidrológicos han utilizado productos derivados de sensores ópticos, radiómetros y radar, tales como HS, T_s , ET ,

Q , índices de vegetación, entre otras, para resolver el balance hídrico (Al-Shrafany et al., 2014; Bertoldi et al., 2014).

Hasta el momento no se han utilizado las imágenes de σ^0 para representar las propiedades del suelo en el balance de agua a nivel de superficie. Es por esto que es importante desarrollar una metodología para estimar HS a nivel superficial que responda a los patrones regionales de PP e involucre las propiedades del suelo implícitas en las imágenes radar.

1.1. Objetivos

1.1.1. Objetivo general

Por lo expuesto en la introducción, el objetivo general de la tesis es mejorar o desarrollar un procedimiento para estimar la HS bajo cobertura vegetal moderada en zonas de llanura, introduciendo la desagregación de información de microondas pasivas, radar y variables hidro-meteorológicas en el balance de agua.

1.1.2. Objetivos específicos

Del objetivo general se desprenden los siguientes objetivos específicos:

- Desarrollar un modelo operativo para estimar HS que tenga en cuenta la entrada y salida de agua al sistema suelo, aplicando metodologías existentes.
- Investigar el problema de desagregar datos con diferentes escalas espaciales, propias de las tecnologías radar y radiómetros pasivos.
- Mejorar y/o desarrollar una metodología operativa que permita obtener mapas de HS, con una resolución espacial y precisión compatibles con aplicaciones agro-hidrológicas.

1.2. Motivación

En Argentina el financiamiento para la adquisición imágenes SAR con múltiple polarización es escaso, en este marco toda imagen SAR gratuita adquiere importancia. Por lo que, el desarrollo de productos derivados de imágenes gratuitas motivó esta tesis. El programa Copernicus, con sus satélites Sentinel, administrado por la Agencia

Espacial Europea ofrece una oportunidad única de disponer de imágenes SAR gratuitas para la comunidad científica, con la limitante que estas sólo se distribuyen en polarización simple VV o en el mejor de los caso VV-VH en nuestra región. Por esto, se estableció como un objetivo de esta tesis desarrollar una metodología que permita explotar este recurso gratuito.

1.3. Organización de la tesis

Esta tesis se encuentra organizada en seis capítulos. El Capítulo 1 introduce el concepto de la variable humedad de suelo y su importancia en el ciclo hidrológico, como así también la necesidad de su cuantificación a escala local y global. Asimismo, se revisan las principales metodologías desarrolladas para su estimación, analizando sus potencialidades y limitaciones.

En el Capítulo 2 se detallan las metodologías a desarrollar en la presente tesis. Primero se describe el procedimiento empleado para la obtención de HS a partir de datos satelitales y datos meteorológicos. Luego, se describe la metodología empleada para la fusión de datos.

En el Capítulo 3 se describen las áreas de estudio y datos disponibles en la localidad de Ataliva, provincia de Santa Fe, y en la ciudad de Monte Buey, provincia de Córdoba, en las cuales se implementaron las metodologías propuestas.

En los Capítulos 4 y 5 se presentan los resultados obtenidos de aplicar la metodología para la estimación de HS y la metodología para la fusión de datos, respectivamente. Los resultados se encuentran organizados según la descripción metodológica y las áreas de estudio.

Finalmente, en el Capítulo 6 se discuten los resultados encontrados junto a las principales conclusiones de la presente tesis. Por último, se exponen recomendaciones y futuras líneas de investigación.

Capítulo 2

Metodología

Como se mencionó anteriormente, el objetivo de esta tesis es desarrollar un modelo para estimar HS. En la Introducción se presentó la necesidad de cuantificar las entradas y salidas de agua del sistema suelo mediante un modelo de balance hídrico de manera de representar todos los procesos que influyen en la variabilidad de HS. En este capítulo se describe el modelo de balance hídrico empleado y su relación con la HS. Luego se detallan los métodos estadísticos propuestos para estimar la HS.

2.1. Ecuación de balance hídrico

Los modelos de balance hídrico del suelo expresan la variación temporal de HS como una suma algebraica de la entrada PP , y las salidas ET y el escurrimiento Q , como se muestra en la Ecuación 2.1 (Orth et al., 2013).

$$\frac{\Delta HS}{\Delta t} = PP - ET - Q \quad (2.1)$$

donde Q representa genéricamente tanto la infiltración como el escurrimiento superficial. Todas las variables están expresadas en mm/día o equivalente.

El movimiento vertical del agua en el sistema suelo - atmósfera ocurre de manera espontánea por gradientes de energía desde el suelo al aire circundante. El flujo de calor latente, que representa la ET , al igual que el flujo de calor sensible, que representa el aumento de T_s , dependen de los gradientes de humedad. La incorporación de vapor de agua en el aire es dependiente del calor latente o ET y está cuantificado mediante la humedad relativa (HR) (Merlin et al., 2016).

Las relaciones entre cada proceso y HS pueden adoptar formas diferentes, sin em-

bargo Orth et al. (2013) y Brocca et al. (2014) coinciden en el hecho de que estas relaciones son no lineales de la forma de la ecuación 2.2:

$$Z = \alpha(HS)^\gamma \quad (2.2)$$

donde Z representa ya sea ET o Q . Por lo tanto, la Ecuación 2.2 puede ser expresada en forma lineal como:

$$\log(Z) \approx \gamma \log(HS) \quad (2.3)$$

Por otro lado, se han desarrollado diversos métodos para calcular ET bajo diferentes condiciones de humedad de la superficie y escalas espacio-temporales. Entre los métodos más utilizados se pueden mencionar Thornthwaite (Thornthwaite, 1948), Penman–Monteith (Penman, 1948), Priestley and Taylor (Priestley and Taylor, 1972), entre otros. Estas metodologías requieren de datos de temperatura de la superficie (T_s), temperatura del aire (T_a), presión de vapor de agua (e_a), radiación neta (R_{net}), HR , entre otras variables (Venturini, 2007; Orth et al., 2013). Esto permite deducir que el logaritmo de las variables involucradas en ET está relacionado con el segundo miembro de la ecuación 2.3. Además, los flujos varían con la rugosidad superficial, las propiedades del suelo y la PP por lo que dichas variables expresan la relación entre Q y HS (Orth et al., 2013). En consecuencia, cualquiera sean las ecuaciones utilizadas para estimar ET y Q , todas conducen a un balance de agua donde la variabilidad de la HS queda expresada como una superposición de funciones no lineales que involucran las variables antes mencionadas.

Para resolver la ecuación de balance de agua, en esta tesis se asume que no hay cambio en HS durante el tiempo de adquisición de los productos satelitales, es decir, $\frac{\Delta HS}{\Delta t} = 0$. Además, el balance de agua se resuelve dentro de un píxel, donde se supone que los movimientos horizontales de entrada y salida de agua, son despreciables y prevalecen los movimientos verticales. Por otra parte, Orth et al. (2013), argumentan que los flujos de agua dentro del volumen de control son una función potencial de HS modificada por PP (Ecuación 2.4); esta relación se obtiene reemplazando las funciones potenciales entre Q y HS , y ET y HS , propuestas por los autores, en la ecuación 2.1. La ecuación 2.4 define la relación entre HS y las variables propuestas en esta tesis.

$$0 = PP(1 - HS^\alpha) - \beta HS^\gamma \quad (2.4)$$

Alley (1984) describe algunos de los modelos usualmente utilizados para efectuar el balance; los modelos T, P, abcd, entre otros. Estos modelos difieren en el tratamiento

de la *ET*, la contabilidad de HS y la estimación de la recarga al acuífero. Estos métodos requieren la calibración de parámetros de base física, tarea que puede ser compleja para grandes áreas (Kornelsen and Coulibaly, 2013). Por esto, en esta tesis se propone simplificar la ecuación de balance utilizando métodos estadísticos que permiten establecer las relaciones entre la HS y las variables que describen *ET*. La razón fundamental de esta elección es la necesidad de desarrollar soluciones aplicables con datos de distribución gratuita y de fácil acceso para usuarios no expertos.

Los métodos propuestos en este trabajo son MLR y ANNs. Por un lado, MLR es un método bien conocido que encuentra los coeficientes que relacionan todas las variables de entrada con la variable incógnita como una combinación lineal simple. Aunque este enfoque no es innovador, es más flexible que los métodos numéricos (que no son más recientes que la regresión lineal) para resolver este tipo de relaciones y no requiere de usuarios expertos. Al-Yaari et al. (2016) y Al-Yaari et al. (2017) aplican una metodología similar a la aquí propuesta para estimar HS a partir de radiómetros. Por otro lado, las ANNs ofrecen una forma más precisa de resolver ecuaciones no lineales con reglas complejas (Ali et al., 2015).

2.2. Regresión lineal múltiple

El modelo MLR es un método estadístico utilizado para modelar linealmente la relación entre múltiples variables. Es un método bien conocido, que se ha utilizado en diferentes estudios para diversas aplicaciones. A pesar de ser un método simple, en teledetección se ha utilizado recientemente para la estimación de HS (Ahmad et al., 2010; Albergaria et al., 2014; Ali et al., 2015; Al-Yaari et al., 2016; Al-Yaari et al., 2017), y para la fusión de datos (Srivastava et al., 2013; Peng et al., 2017), entre otros. El MLR calcula la variable dependiente (Y) usando una combinación lineal de variables independientes (X_i). La regresión lineal tiene la forma de la ecuación (Weisberg, 2005):

$$Y_i = \beta_0 + \sum_{k=1}^n \beta_k X_{k,i} + \varepsilon \quad (2.5)$$

donde β_0 , $\beta_k (k > 0)$, y ε son términos constantes que representan la intersección, los coeficientes de regresión y el error del modelo, respectivamente.

Los coeficientes β_k son estimados por mínimos cuadrados, buscando minimizar la

suma de los residuos cuadrados (S) definidos en la ecuación 2.6:

$$S = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})^2 \quad (2.6)$$

Para aplicar el criterio de mínimos cuadrados en el modelo MLR, se calcula la primera derivada de S con respecto a cada β_j en la expresión 2.6:

$$\begin{aligned} \frac{\partial S}{\partial \beta_0} &= 2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})(-1) \\ \frac{\partial S}{\partial \beta_1} &= 2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})(-X_{1,i}) \\ &\vdots = \vdots \\ \frac{\partial S}{\partial \beta_k} &= 2 \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})(-X_{k,i}) \end{aligned} \quad (2.7)$$

Los estimadores de mínimos cuadrados se obtienen al igualar a 0 las derivadas anteriores:

$$\begin{aligned} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})(-1) &= 0 \\ \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})(-X_{1,i}) &= 0 \\ &\vdots = \vdots \\ \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1,i} - \beta_2 X_{2,i} - \dots - \beta_k X_{k,i})(-X_{k,i}) &= 0 \end{aligned} \quad (2.8)$$

Finalmente, en este sistema hay k ecuaciones con $k + 1$ incógnitas ($\beta_0, \beta_1, \beta_2, \dots, \beta_k$), que pueden encontrarse utilizando álgebra matricial.

El modelo de regresión requiere que las variables de entrada (X_i) expliquen la variable de salida (Y), y además que cada variable de entrada aporte información diferente, es decir, no debe ser colineal. Por lo tanto, el proceso de selección de las variables significativas dentro del conjunto de variables de entrada disponibles, es importante para obtener buenos resultados. En esta tesis se utiliza una combinación de las metodologías propuestas por Weisberg (2005) y Harrell (2001). Primero, se aplica la metodología de eliminación hacia atrás utilizando el estadístico p-valor para determinar la significancia de cada variable (Weisberg, 2005). Posteriormente, una vez que se selecciona un

conjunto de variables de entrada representativas del comportamiento de la variable de salida, se analiza la correlación entre ellas. Esta tarea se realiza para investigar la linealidad de las variables recurriendo al factor de inflación de la varianza (FIV) (Fox and Monette, 2002; Harrell, 2001). Valores de FIV superiores a cuatro sugieren la existencia de multicolinealidad, valores de FIV mayores de diez indican la presencia de multicolinealidad severa (Montgomery and Peck, 1992; Fox and Monette, 2002; Cohen et al., 2003). Otros autores sugieren que el límite de FIV es entre cinco y diez (Marquardt, 1970). Ambos estadísticos son relevantes para la selección de variables del modelo, en este trabajo se aplica una combinación de p-valor y el criterio FIV. Por lo tanto, las variables con elevado p-valor (mayor a 0.05) y colineales (VIF mayor de 5) se descartan.

Una vez seleccionadas las variables se procede a calibrar el modelo, es decir se calculan los coeficientes β_k . Finalmente el modelo encontrado es el que mejor describe la variable Y en función de las variables X_i seleccionadas, sin embargo se deben validar los resultados calculados con el modelo para cuantificar los errores involucrados.

2.3. Redes neuronales artificiales

Las ANNs son un paradigma computacional basado en un modelo matemático inspirado en el comportamiento biológico de las neuronas y en la estructura del cerebro. El cerebro puede considerarse un sistema altamente complejo, con más de cien mil millones de neuronas. La clave para el procesamiento de la información son las conexiones entre neuronas, denominadas sinapsis.

La estructura biológica de las neuronas está compuesta por dendritas, el axón y el soma (Figura 2.1 -a). Las neuronas recogen señales procedentes de otras neuronas a través de las dendritas y emiten impulsos de actividad eléctrica mediante el axón. Las extremidades del axón llegan hasta otras neuronas y establecen conexiones llamadas sinapsis, mediante la liberación de unas sustancias llamadas neurotransmisores. Los neurotransmisores liberados en las sinapsis pueden tener un efecto excitatorio o inhibitorio sobre la neurona receptora (Rubiolo, 2014).

La neurona toma un cierto nivel de activación, en función de la estimulación recibida, que se traduce en breves impulsos nerviosos que propaga hacia las neuronas con las cuales se conecta. De esta manera la información se transmite de unas neuronas a otras y va siendo procesada a través de las conexiones sinápticas y las propias neuronas. El aprendizaje de las redes neuronales se produce mediante la variación de la efectividad

de las sinapsis.

Las ANNs están compuestas por un conjunto de neuronas artificiales interconectadas entre sí, y emulan las conexiones sinápticas y la liberación de neurotransmisores con efectos excitatorios o inhibidores. Cada neurona artificial se representa como una unidad compuesta de conexiones de entrada, un núcleo central de proceso y una salida (Figura 2.1-b) (Rumelhart et al., 1988).

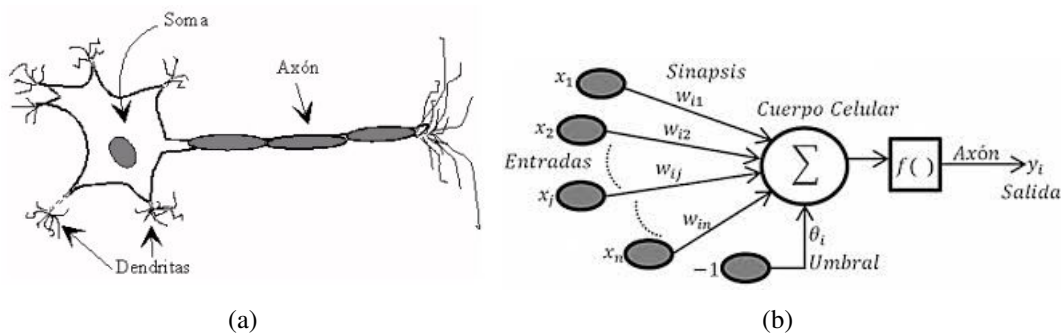


Figura 2.1: Estructura de una neurona: a) Biológica; b) Artificial

El Perceptrón simple, inventado por Rumelhart et al. (1988) fue la primera red neuronal descrita algorítmicamente. Es un modelo simple de neurona basado en el modelo de McCulloch and Pitts (1943) y en una regla de aprendizaje basada en la corrección del error. El Perceptrón está constituido por un conjunto de sensores de entrada que reciben los patrones de entrada a reconocer o clasificar, y una neurona de salida que se ocupa de clasificar los patrones de entrada en dos clases. Esta red es usada para clasificación de patrones que pertenecen a dos clases linealmente separables.

El Perceptrón multicapa (MLP, por su sigla en inglés) es una generalización del Perceptrón simple, que surge como respuesta a las limitaciones de dicha arquitectura, es decir, al problema de la separabilidad no lineal. Minsky and Papert (1969) mostraron que la combinación de varios perceptrones simples podría resultar en una solución adecuada para tratar ciertos problemas no lineales. Sin embargo, la regla de aprendizaje del Perceptrón simple no puede aplicarse en este escenario. Rumelhart et al. (1988) presentaron una manera de retropropagación de los errores medidos en la salida de la red hacia las neuronas ocultas, dando lugar a la llamada regla delta generalizada o retropropagación.

Autores como Cybenko (1989) y Hornik (1991) han demostrado que el MLP es un aproximador universal, es decir, cualquier función continua en un espacio R^n puede aproximarse por un MLP con al menos una capa oculta de neuronas. Es por esto que el

MLP es un modelo matemático útil a la hora de aproximar o interpolar relaciones no lineales entre datos de entrada y salida.

La arquitectura del MLP se caracteriza por contar con sus neuronas agrupadas en capas de diferentes niveles. Se distinguen tres tipos diferentes de capas: la capa de entrada, las capas ocultas y la capa de salida. La Figura 2.2 representa la arquitectura de la red neural MLP.

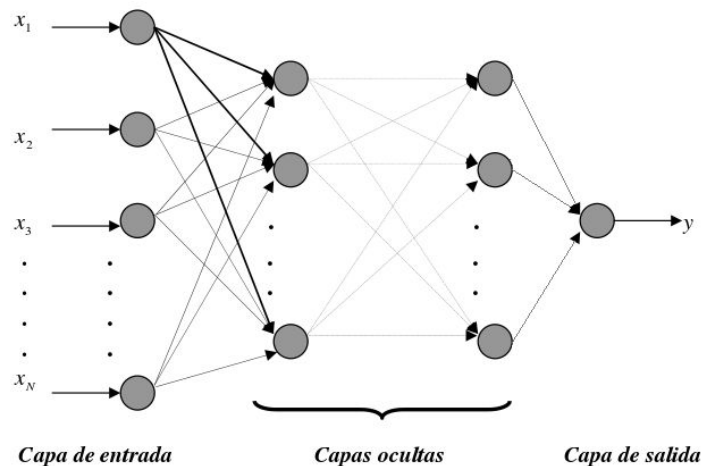


Figura 2.2: Arquitectura de la red neural MLP utilizado para la predicción

El modelo MLP puede formularse matemáticamente como:

$$\hat{Y} \approx F \left(\sum_{j=1}^m w_{kj} F \left(\sum_{i=1}^n w_{ji} X_i + B_j \right) + B_k \right) \quad (2.9)$$

donde w_{kj} son los pesos entre las capas ocultas y de salida; w_{ji} son los pesos entre la entrada y las capas ocultas; X_i son variables de entrada; m es el número de neuronas en una capa oculta; n es el número de neuronas en la capa de entrada; B_j y B_k son los valores bias de las neuronas en las capas ocultas y de salida, respectivamente; F es la función de transferencia; e \hat{Y} es la salida.

El MLP es entrenado por un gran conjunto de muestras de entrenamiento de acuerdo a un procedimiento de aprendizaje supervisado. El método de aprendizaje básico para el MLP se denomina de retropropagación o propagación hacia atrás de errores (Rumelhart et al., 1988). El proceso de entrenamiento de la red con este algoritmo involucra tres fases: la fase del cálculo hacia adelante con los patrones de entrada, el cálculo del error cometido y la propagación hacia atrás de ese error con el ajuste correspondiente de los pesos del modelo.

El MLP encuentra una relación entre las variables de entrada y las variables de salida de la red, propagando hacia adelante los valores de las variables de entrada. Para ello, cada neurona de la red procesa la información recibida por sus entradas y produce una respuesta o activación que se propaga, a través de las conexiones correspondientes, hacia las neuronas de la siguiente capa.

Para implementar dicha relación, la primera capa (de entrada) tendrá tantos sensores como componentes tenga el patrón de entrada, es decir, N ; la capa de salida tendrá tantas unidades de proceso como componentes tengan las salidas deseadas, es decir, M , y el número de capas ocultas y su tamaño dependerán de la dificultad de la correspondencia a implementar.

Como las entradas a las unidades de proceso de una capa son las salidas de las unidades de proceso de la capa precedente, el MLP con sólo una capa oculta implementa la siguiente función:

$$y_i = g_1 \left(\sum_{j=1}^L w_{i,j} s_j \right) = g_1 \left(\sum_{j=1}^L w_{i,j} \left(g_2 \left(\sum_{r=1}^N t_{j,r} x_r \right) \right) \right) \quad (2.10)$$

donde $w_{i,j}$ es el peso sináptico de la conexión entre la unidad de salida i y la unidad de proceso j de la capa oculta; L es el número de unidades de proceso de la capa oculta; g_1 es la función de transferencia de las unidades de salida, que puede ser una función logística, una función tangente hiperbólica o la función identidad; t_{jr} es el peso sináptico que conecta la unidad de proceso j de la capa oculta con el sensor de entrada r y g_2 es la función de transferencia de las unidades de la capa oculta, que puede ser también una función logística, una función tangente hiperbólica o la función identidad.

La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las variables de salida (Ecuación 2.11). Las señales de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. El proceso trata de determinar los pesos de las conexiones sinápticas entre las unidades de proceso de manera que las salidas de la red coincidan con las salidas deseadas, o por lo menos, sean lo más próximas posibles. Es decir, se trata de determinar los pesos de manera que el error total sea mínimo:

$$E = \sum_{k=1}^P \sum_{i=1}^M (z_i(k) - y_i(k))^2 \quad (2.11)$$

siendo, $z_i(k)$ y $y_i(k)$ la salida deseada i y la salida i obtenida por el MLP para la iteración k , respectivamente.

El algoritmo de retropropagación utiliza el método del descenso por el gradiente del error y realiza un ajuste de los pesos comenzando por la capa de salida, según el error cometido, y se procede propagando el error a las capas anteriores, de atrás hacia delante, hasta llegar a la capa de las unidades de entrada (Rumelhart et al., 1988).

Supongamos que se está en la iteración k , donde se ha introducido el patrón cuya salida de unidad i es $y_i(k)$ y la salida deseada $z_i(k)$, siendo los pesos sinápticos $w_{ij}(k)$ y $t_{j,r}(k)$ con $i=1,2,\dots, M$, $j=1,2,\dots, L$ y $r=1,2,\dots, N$. Entonces la regla de modificación de los pesos sinápticos de la capa de salida será:

$$w_{i,j}(k+1) = w_{i,j}(k) + \Delta w_{i,j}(k) \quad (2.12)$$

Es decir, al peso sináptico $w_{i,j}(k)$ se le aplica una corrección $\Delta w_{i,j}(k)$, que es proporcional al gradiente instantáneo:

$$\Delta w_{i,j}(k) = -\eta \frac{\partial E}{\partial w_{i,j}(k)} \quad (2.13)$$

donde η es una constante que determina la tasa o velocidad de aprendizaje. El signo menos en la Ecuación 2.13 denota el descenso por el gradiente del error en el espacio de los pesos. De acuerdo a la regla de la cadena, es posible expresar este gradiente como:

$$\frac{\partial E}{\partial w_{i,j}(k)} = \eta [z_i(k) - y_i(k)] g_1(h_i) s_j(k) \quad (2.14)$$

donde s_j es la entrada a la neurona y g_1 es la función de transferencia evaluada en h_i que se calcula como:

$$h_i = \sum_{j=i}^L w_{i,j}(k) s_j(k) \quad (2.15)$$

Luego de obtener el error para cada neurona de salida, este valor de error es transmitido hacia atrás en la red, para todas las capas intermedias. Los pesos sinápticos de las capas intermedias se modifican según el error y los valores de las salidas de las unidades de proceso precedentes, multiplicados por sus pesos sinápticos.

Para asegurar un correcto funcionamiento del algoritmo de retropropagación, las entradas deben estar normalizadas. Además, con el objetivo de acelerar el proceso de aprendizaje, las entradas no deben estar correlacionadas y se deben escalar para que sus covarianzas sean aproximadamente iguales, asegurando así que los diferentes pesos

sinápticos en la red aprendan aproximadamente a la misma velocidad (Haykin, 2009).

2.4. Secuencia metodológica

La Figura 2.3 presenta el procedimiento completo propuesto para la estimación de HS utilizando las metodologías antes descritas. En el diagrama de flujo se incluyen las siguientes etapas: procesamiento de imágenes satelitales, preparación de los datos, calibración y validación de las metodologías propuestas, y generación de los mapas de HS.

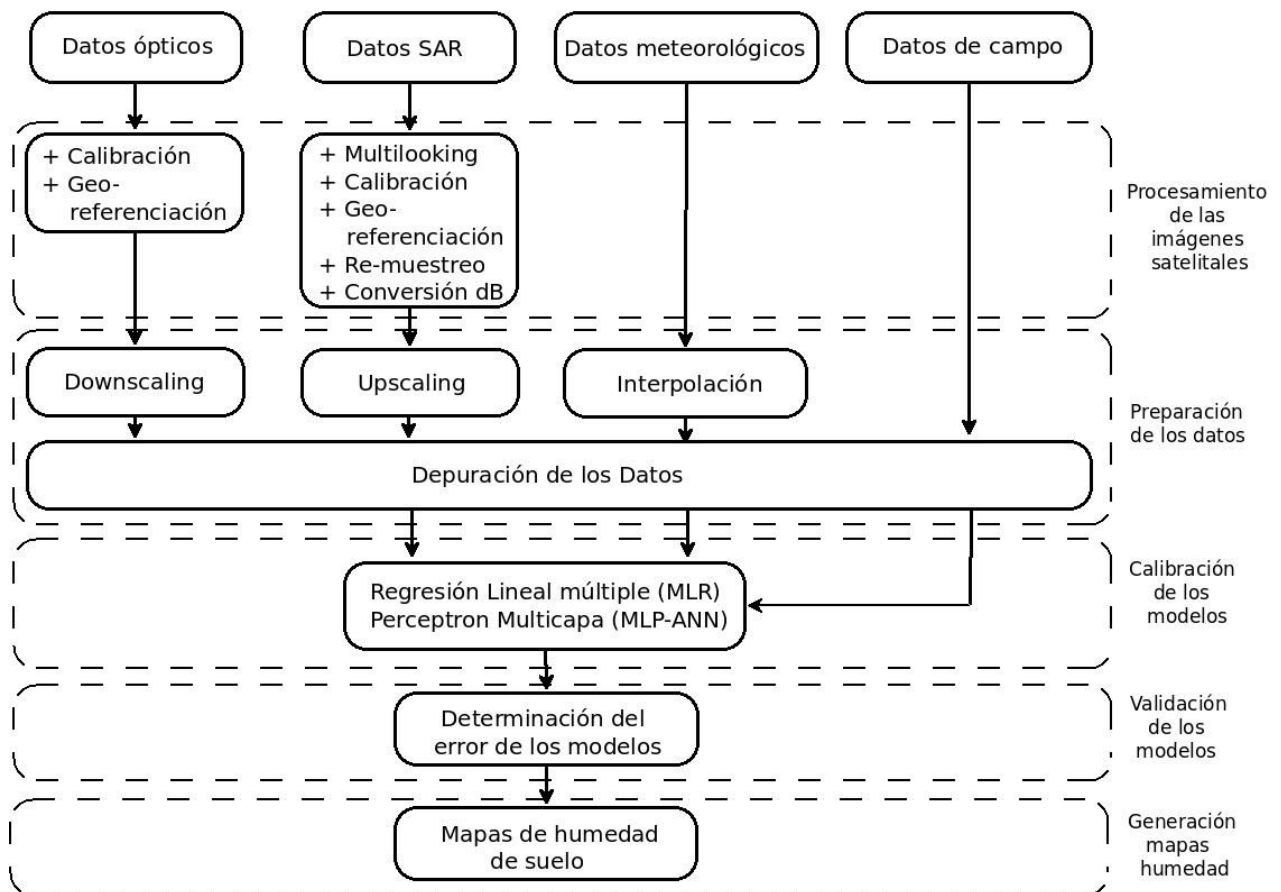


Figura 2.3: Diagrama de flujo del procedimiento propuesto para la estimación de HS.

2.4.1. Índices de rendimiento

Las metodologías propuestas requieren de una etapa de calibración, donde se obtienen los parámetros, coeficientes, pesos, etc, de los modelos, y otra de validación donde

se evalúa la bondad de las estimaciones. Por lo tanto, es necesario evaluar los resultados de los modelos mediante diferentes estadísticos que cuantifican la performance de los mismos. El coeficiente de determinación (R^2), el error medio cuadrático (RMSE), el desvío estándar (S) y el sesgo se utilizan para cuantificar los errores de los modelos. Las ecuaciones para obtener estos estadísticos son:

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.16)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (2.17)$$

$$S = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad (2.18)$$

$$Sesgo = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)}{n} \quad (2.19)$$

donde Y_i e \hat{Y}_i son la HS observada y estimada, respectivamente, \bar{Y} es la HS media observada y n es el número de muestras.

Para analizar la calidad de los mapas de HS obtenidos, es necesario contrastarlos con productos de HS elaborados con otras técnicas y fuentes de datos. El índice de similitud estructural (SSIM, por su sigla en inglés), es un método que intenta modelar el cambio percibido en la información estructural de la imagen (Wang et al., 2004). El SSIM se puede interpretar como una medida de calidad de una de las imágenes que se comparan, mientras que la otra imagen se considera de calidad perfecta. La ecuación del índice es la siguiente:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.20)$$

donde μ es la media de las intensidades de los píxeles en la dirección x e y ; σ es la varianza de las intensidades en la dirección x e y ; y σ_{xy} es la covarianza. El valor de SSIM varía entre -1 y 1, donde 1 indica una similitud perfecta.

Capítulo 3

Área de estudio y datos

La metodología propuesta se aplicó sobre dos áreas de estudio, primero en la localidad de Ataliva y luego, en la de Monte Buey, ambas en la región Pampeana Argentina.

El motivo de la selección de estos sitios fue la disponibilidad de datos de campo para la calibración y validación de los modelos. La fuente de los datos disponibles, así como las imágenes SAR utilizadas en cada caso, son diferentes. De esta forma se plantea la ecuación de balance con diferente cantidad y calidad de datos, aunque la metodología propuesta es la misma en ambos casos. Por ello, ambas aplicaciones no pueden ser consideradas como una réplica experimental. Por el contrario, son dos estudios distintos en los que se determina HS con distintas imágenes SAR y datos hidro-meteorológicos.

3.1. Localidad de Ataliva - Provincia de Santa Fe

3.1.1. Ubicación

Se ubica en la Provincia de Santa Fe, Argentina, con centro en la localidad de Ataliva ($-30^{\circ} 59'$, $-61^{\circ} 25'$) (Figura 3.1). Esta provincia integra la región agrícola-ganadera e industrial de Argentina, donde se concentra gran parte de la actividad del sector agrícola y de la población del país. Desde el punto de vista climático, la zona de estudio se caracteriza por un clima templado; la isoterma media anual es de 18°C y la precipitación media anual es de 950 mm. Las precipitaciones se concentran en el semestre noviembre–abril, con un 75 % del total anual. Los suelos de la zona pertenecen a asociaciones de Argiudoles Típicos. La localidad de Ataliva se caracteriza fundamentalmente por la producción agrícola y tambera, y en menor medida por la ganadera. Los

principales cultivos son soja, maíz, alfalfa y trigo.

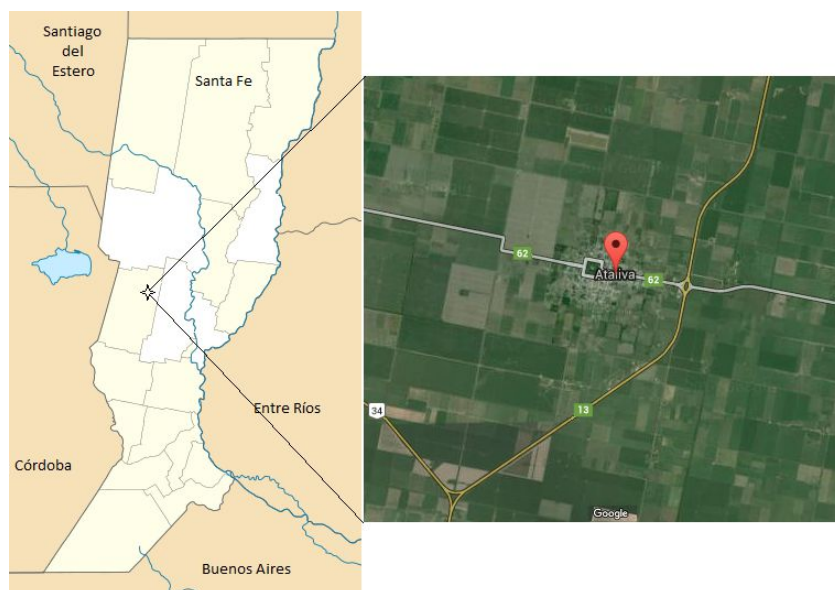


Figura 3.1: Ubicación del área de estudio Ataliva

3.1.2. Datos

El conjunto de datos utilizado abarca el período comprendido entre septiembre de 2012 y mayo de 2013. Se compone de: datos *in situ* de HS, datos meteorológicos y datos satelitales de sensores ópticos y radar.

Datos *in situ*

Los datos de campo fueron obtenidos en sucesivas campañas realizadas por el Centro de Estudios Hidro-Ambientales de la Universidad Nacional del Litoral-UNL, en el marco de proyectos financiados por CONAE y la UNL. Los valores de HS corresponden a los días 03/09/2012, 14/09/2012, 25/09/2012, 30/11/2012, 22/04/2013 y 14/05/2013, fechas en las que se extrajeron muestras de suelo a una profundidad de 3-5 cm para análisis gravimétrico en laboratorio. En cada fecha se muestrearon en promedio 10 parcelas cuyos tamaños varían entre 1.5 ha a 40 has, resultando en 63 registros. Mayor información de las campañas está disponible en [Brogioni et al. \(2014\)](#).

Durante las campañas se identificaron los siguientes tipos de cobertura: rastrojo (soja, sorgo, maíz, trigo), barbecho, pastos (verdeo, alfalfa). Algunos campos se encontraban arados y otros sembrados con soja en diferentes estados fenológicos.

Datos meteorológicos

La única estación meteorológica con datos utilizables se encuentra en la localidad de Rafaela, Santa Fe, perteneciente a la red de estaciones experimentales agropecuarias del Instituto Nacional de Tecnología Agropecuaria (INTA). La estación meteorológica automática instalada en esta localidad es del tipo NIMBUS THP (Termo Higro Pluviométricas), que registra T_a a 1.5 m de altura, T_s a 10 cm de profundidad, HR y PP , mediante sensores de alta precisión y con frecuencia de diez minutos (Belloni et al., 2011).

Rafaela está ubicada aproximadamente a 32 km al sur de Ataliva, por lo que previo a su utilización se verificó la homogeneidad de las precipitaciones medias mensuales en la zona mediante la base de datos CRU-TS 3.2 (Krepper and Venturini, 2009).

Los datos meteorológicos utilizados fueron PP , T_a , T_s y HR , los cuales fueron pre-procesados para obtener valores acumulados para el caso de PP y valores promedio para las restantes variables. Se utilizó una ventana de 7 días previos a la fecha de recolección de las muestras.

Datos satelitales ópticos y radar

Como variables de entrada al método MLR se utilizó la información que se detalla a continuación: por un lado se trabajó con imágenes del satélite TerraSAR-X que posee un sensor SAR en la longitud de onda X de 31 mm. Los productos adquiridos son en modo single look Stripmap en polarización HH, con ángulo de incidencia del centro de la escena de 24° y fecha de adquisición coincidente con los datos de campo. Los productos de TerraSAR-X fueron adquiridos con anterioridad a esta tesis y se utilizaron para aproximar un balance hídrico con una imagen con polarización HH, adecuada para estimar HS.

Otra variable de entrada fue el Índice de Vegetación Normalizado (NDVI), obtenido a partir del sensor MODIS (Moderate Resolution Imaging Spectroradiometer) a bordo de la misión satelital Terra (EOS AM-1) de la NASA, el cual proporciona cada dieciséis días el producto MOD13A1 que posee el índice de vegetación con una resolución espacial de 500 m. Los detalles sobre los datos satelitales SAR y ópticos, tales como la fecha de adquisición y el tipo de producto disponibles para este trabajo, se proporcionan en la Tabla 3.1.

Tabla 3.1: Inventario de las imágenes ópticas y SAR

Fechas (mm/dd/yyyy)	
Terra-MOD13A1	TerraSAR-X
08/28/2012	09/03/2012
09/13/2012	09/14/2012
09/29/2012	09/25/2012
12/02/2012	11/30/2012
04/23/2013	04/22/2013
05/09/2013	05/14/2013

3.2. Localidad de Monte Buey - Provincia de Córdoba

3.2.1. Área de estudio

Está ubicada en el sureste de la provincia de Córdoba, Argentina, en torno a una pequeña localidad llamada Monte Buey ($-32^{\circ} 55'$, $-62^{\circ} 27'$) (Figura 3.2). Parte de esta provincia se está inserta en la región Pampeana, un área de aproximadamente 840000 km² caracterizada por pendientes inferiores a 0.5 % (Bedano et al., 2016).

La región tiene un clima continental con una temperatura media anual de 17.4 °C. La temperatura media mensual oscila entre 24.3 °C en enero y 10.5 °C en julio. La PP media anual es de 920 mm (1960-2010), acumulada principalmente de septiembre a abril (Thibeault et al., 2015). Según Duval et al. (2016) el suelo se clasifica como Argiudol típico. El área de Monte Buey se caracteriza principalmente por la producción agrícola y láctea y, en menor medida, por la ganadería. Los cultivos predominantes son la soja y el maíz. A veces, son reemplazados con sorgo, alfalfa y pasturas (Thibeault et al., 2015). Se trata de una zona plana en la que durante los años hidrológicos normales, predominan los flujos verticales de agua sobre los horizontales.

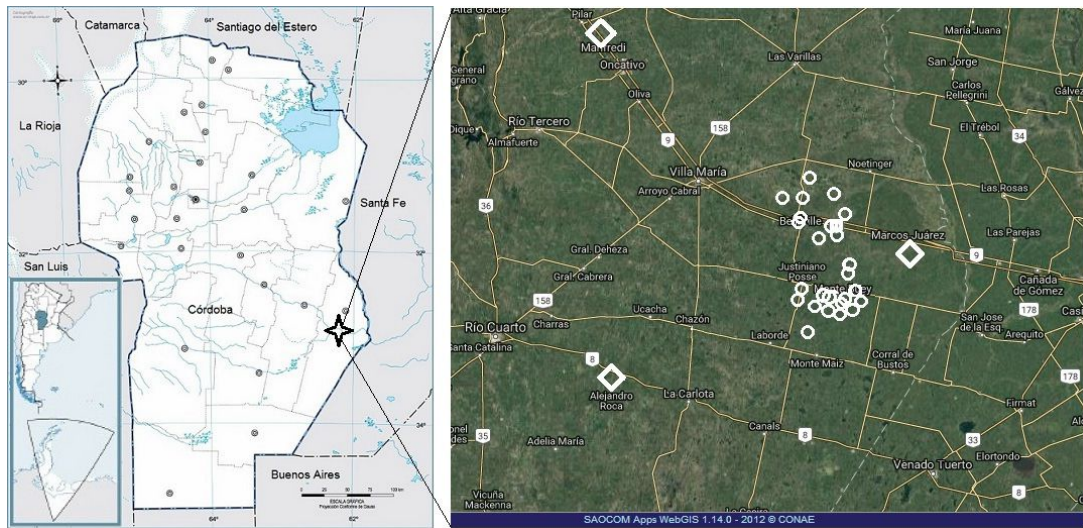


Figura 3.2: Área de estudio Monte Buey. Izquierda: Ubicación; Derecha: Puntos de medición *in situ*; datos meteorológicos (rombo); datos de HS (círculos)

3.2.2. Datos

El conjunto de datos utilizado en este caso abarca el período comprendido entre diciembre de 2014 y abril de 2016, y se compone de: mediciones *in situ* de HS y T_s , datos meteorológicos, datos satelitales provenientes de sensores ópticos, SAR, y microondas pasivas, y mapas de HS modelada.

Datos *in situ*

Las observaciones de HS y T_s fueron proporcionadas por la Comisión Nacional de Actividades Espaciales (CONAE). CONAE opera una red de medición implementada para ayudar en el proceso de calibración de los mapas de HS producidos por las misiones actuales como SMAP y futuras, como la misión Argentina denominada Satélite Argentino de Observación Con Microondas (SAOCOM). La red se encuentra en los campos cercanos a la ciudad de Monte Buey, donde cada sitio está equipado con sensores HydraProbe II (Stevens Water Monitoring System, Inc.) instalados para el monitoreo continuo de humedad (% Volumétrico), temperatura y salinidad del suelo a una profundidad de 5 cm (Thibeault et al., 2015). Las estaciones han sido adecuadamente distribuidas en el área de estudio como se muestra en la Figura 3.2 (círculos) con el objetivo de cubrir la variabilidad espacial de los parámetros superficiales de la zona.

Datos meteorológicos

Los datos meteorológicos disponibles pertenecen a estaciones meteorológicas del INTA. Las estaciones más cercanas son: Alejandro Roca, Manfredi y Marcos Juárez, todas en la provincia de Córdoba, marcadas con rombos en la Figura 3.2. Las estaciones meteorológicas automáticas instaladas en estas localidades son del tipo NIMBUS THP (Belloni et al., 2011). Los datos provistos por las estaciones de INTA incluyen T_a , e_a , PP y HR . El valor acumulado de siete días de (PP) y el promedio de siete días de T_a , e_a y HR se utilizaron como variables de entrada del modelo.

Datos satelitales ópticos y radar

Las áreas sin vegetación o con escasa vegetación se identifican con datos satelitales ópticos. Para llevar a cabo esta tarea se utilizó el producto TERRA-MODIS MOD13Q1 que proporciona datos de NDVI a una resolución espacial de 250 m cada dieciséis días. Este producto está fácilmente disponible en contraste con los productos de otras misiones que tienen una mejor resolución espacial pero una peor resolución temporal.

Además, se utilizaron los productos de ET provistos por las misiones satelitales TERRA y AQUA de la NASA. La misión AQUA, al igual que TERRA, tiene a bordo el sensor MODIS. La órbita del satélite TERRA alrededor de la tierra va de norte a sur cruzando el Ecuador por la mañana, mientras que el satélite AQUA va de sur a norte cruzando el Ecuador por la tarde. En este trabajo, se usaron los productos MODIS-TERRA MOD16A2 y MODIS-AQUA MYD16A2, ambos proporcionan datos de ET real a una resolución espacial de 500 m cada ocho días, donde ET es la pérdida de agua total acumulada de 8 días. La ET es estimada sobre la base de los productos MODIS y la ecuación de Penman-Monteith (Mu et al., 2007, 2011).

Los datos SAR se obtienen de la misión Sentinel-1 que tiene a bordo un sensor SAR que opera en banda C. Esta misión proporciona productos gratuitos en modo Single Look Complex (SLC) y Ground Range Detected (GRD), ambos en modo Interferometric Wide Swath para la polarización VV. Cabe destacar que la Agencia Espacial Europea-ESA no distribuye en forma gratuita imágenes polarización HH en la zona de estudio utilizada. Si bien esta polarización no es la óptima para la determinación de HS (Holah et al., 2005; Paloscia et al., 2010), en esta tesis se explora el uso de la misma para estimar HS, considerando que es un producto gratuito a disposición de todos los usuarios finales que requieran de esta variable, y no dispongan de recursos económicos para comprar imágenes SAR de mejor polarización.

Las imágenes fueron procesadas con la herramienta SNAP (Sentinel Application Platform) proporcionada por la ESA. Los productos SLC se procesaron realizando el multilooking para reducir el ruido de Speckle, proceso que modificó la resolución espacial original a 30 m; luego, se calibró para obtener la intensidad de retrodispersión. Esta imagen de intensidad fue georreferenciada utilizando un sistema geográfico y un modelo de elevación digital. En este trabajo, se seleccionó el sistema de coordenadas geográficas WGS84 (son las siglas en inglés de World Geodetic System 84) y un modelo de elevación provisto por la misión Shuttle Radar Topography Mission con una resolución plana de 90 m. La imagen de intensidad de retrodispersión se convirtió a decibeles (σ^0). Los productos GRD se procesaron con el mismo software realizando las siguientes operaciones: multilooking, calibración, geocodificación y conversión a decibeles (σ^0).

Por otra parte, se utilizan datos de HS y T_s satelitales provistos por la misión SMAP. Estos datos son obtenidos del producto L3_SM_P con una resolución espacial de 36 km. Este producto es calculado utilizando un algoritmo basado en el modelo tau-omega (Kerr and Njoku, 1990), que emplea la temperatura de brillo, obtenida con el radiómetro en la banda L, junto a información adicional. La HS así obtenida posee una precisión de $0.04 \text{ m}^3/\text{m}^3$ (Entekhabi et al., 2010).

Se utilizan además, mapas de PP pertenecientes a la misión Global Precipitation Measurement (GPM). GPM es una misión conjunta entre la Agencia Japonesa de Exploración Aeroespacial (JAXA) y la NASA, así como otras agencias espaciales internacionales, para hacer observaciones frecuentes (cada 2-3 horas) de la precipitación de la Tierra. El proyecto proporciona mapas mundiales de precipitación para ayudar a la comunidad científica a mejorar la predicción de eventos extremos, estudiando el clima global y agregando capacidades actuales para usar esos datos de satélite para beneficiar a la sociedad. GPM se basa en los notables éxitos de la Misión de Medición de las Precipitaciones Tropicales (TRMM), que también fue una actividad conjunta de la NASA-JAXA. Los mapas de PP utilizados son productos del algoritmo denominado Integrated Multi-satellite Retrievals for GPM (IMERG). Este algoritmo está destinado a intercalibrar, combinar e interpolar las estimaciones de todos los satélites de precipitación de microondas, junto con estimaciones de satélites infrarrojos y otros estimadores de precipitación en escalas de tiempo y espacio fina para los programas TRMM y GPM en todo el mundo.

En la Tabla 3.2 se incluyen las fechas de adquisición de las imágenes satelitales disponibles para este trabajo.

Mapas de HS disponible

Por otro lado, también se utilizaron mapas de HS modelada a 25 km, provistos por el grupo de investigación de hidrología terrestre de la Universidad de Princeton (<http://hydrology.princeton.edu/models.php>, último acceso Febrero 2017), para contrastarlos con los mapas de HS generados con la metodología propuesta. Este grupo provee información sobre las condiciones meteorológicas, indicadores de sequía e inundación, índices de humedad de suelo, tasas de flujo, pronósticos, entre otros, para América del Sur, México y el Caribe, mediante el sitio web denominado “Monitoreo de Inundaciones y Sequía en América Latina y el Caribe (LAFDM)”.

Los mapas de HS modelada son calculados con el modelo de Capacidad de Infiltración Variable o VIC (Liang, 1994) que es un modelo de hidrología semi-distribuido a macro-escala que resuelve los balances de agua y energía completos. Este modelo ha sido extensamente utilizado en estudios de análisis de recursos hídricos, interacciones atmósfera-suelo y cambio climático (Hamlet and Lettenmaier, 1999; Beyene et al., 2010).

El modelo VIC realiza los balances de masas y energías a través de celdas que conforman un espacio grillado; mientras que, los procesos que ocurren en menor escala que el tamaño de la celda, se incluyen de manera estadística. Las características que distinguen al modelo VIC de otros modelos son: variabilidad en las clases de vegetación a nivel de celda; variabilidad a nivel de celda de la capacidad de almacenamiento de la humedad del suelo; drenaje desde la zona de humedad del suelo hacia una recesión no lineal (flujo base); e inclusión de la topografía que permite precipitación orográfica y cambios en la temperatura, resultando en una hidrología más realista en zonas montañosas. Las fechas de los mapas VIC son coincidentes con las fechas de las imágenes SMAP (Tabla 3.2).

Tabla 3.2: Inventario de las imágenes Sentinel-1, GPM, HS observado, VIC, SMAP y MODIS MOD / MYD 16.

Sentinel-1/GPM/ <i>in situ</i>	SMAP/VIC	MODIS MOD/MYD 16
Fecha (mm/dd/yyyy)		
04/08/2015	04/11/2015	04/06/2015
05/02/2015	05/02/2015	05/02/2015
05/12/2015	05/10/2015	05/12/2015
05/26/2015	05/26/2015	05/25/2015
06/05/2015	06/03/2015	06/05/2015
06/19/2015	06/19/2015	06/18/2015
07/23/2015	07/24/2015	07/20/2015
08/16/2015	08/17/2015	08/13/2015
08/30/2015	08/30/2015	08/29/2015
09/09/2015	09/10/2015	09/06/2015
09/23/2015	09/23/2015	09/22/2015
10/27/2015	10/28/2015	10/24/2015
11/10/2015	11/13/2015	11/09/2015
11/20/2015	11/21/2015	11/17/2015
12/14/2015	12/18/2015	12/11/2015
12/28/2015	12/28/2015	12/27/2015
01/07/2016	01/08/2016	01/09/2016
01/21/2016	01/19/2016	01/17/2016
01/31/2016	01/27/2016	01/25/2016
02/14/2016	02/14/2016	02/10/2016
02/24/2016	02/25/2016	02/26/2016
03/09/2016	03/12/2016	03/13/2016
03/19/2016	03/20/2016	03/21/2016
04/02/2016	04/02/2016	04/06/2016
04/12/2016	04/13/2016	04/14/2016
04/26/2016	04/24/2016	04/22/2016
05/20/2016	05/20/2016	05/20/2016

La presencia de valores atípicos en los datos indica errores en la medición o la ocurrencia de un fenómeno inesperado y previamente desconocido. Es importante verificar

los valores atípicos en cada variable, ya que tienen impacto en el análisis estadístico. Es por esto que, antes de comenzar a utilizar los datos, mediciones *in situ*, se aplicó un análisis de consistencia para realizar ciertas verificaciones sobre los valores de las variables. Los resultados detallados se presentan en el Anexo [A](#).

Capítulo 4

Estimación de HS con datos observados

En este capítulo se presentan los resultados obtenidos luego de aplicar las metodologías propuestas a los datos disponibles. Como se mencionó en el capítulo anterior, se trabajó en dos áreas de estudio, primero en la provincia de Santa Fe en el entorno de la localidad de Ataliva donde se disponía de datos SAR en banda X y datos de HS *in situ* obtenidos por gravimetría y luego, en la provincia de Córdoba en el entorno de la localidad de Monte Buey donde se contaba con datos SAR en banda C y mediciones volumétricas de HS obtenidas por la red de sensores operada por CONAE. La ecuación de balance fue resuelta utilizando los métodos estadísticos MLR y MLP, los que se implementaron utilizando el lenguaje de programación Python. Los programas desarrollados se encuentran listados en el Anexo [B](#).

4.1. Localidad de Ataliva

La ecuación de balance propuesta en la metodología fue resuelta para el área de estudio de Ataliva. En este caso los datos disponibles eran escasos por lo que se decidió solo utilizar el método MLR. Como se explicó en el Capítulo 2, el método MLP requiere de un volumen importante de datos para el entrenamiento de la red. En contraste, el MLR puede ser aplicado con menor cantidad de muestras.

Por otra parte, [Holah et al. \(2005\)](#) encontraron que en suelos desnudos la señal radar aumenta en forma lineal con el aumento de la humedad. Considerando este resultado, se propuso que la HS podía aproximarse como una superposición de funciones lineales, sin que esto represente una pérdida importante de la precisión.

4.1.1. Método MLR

Como se describió en el Capítulo 3, las variables disponibles para las fechas indicadas eran σ^0 (dB), NDVI, PP (mm), T_a ($^{\circ}\text{C}$), T_s ($^{\circ}\text{C}$) y HR (%). Un análisis teórico previo de las variables puso de manifiesto la relación existente entre T_a y T_s (Parton, 1984). Por lo tanto, para evitar problemas de colinealidad entre estas variables, se decidió tomar el diferencial de temperatura ($\Delta T = T_a - T_s$).

Con el conjunto de datos de las variables σ^0 , NDVI, PP , ΔT y HR , se generó un modelo considerando todas las variables posibles con un R^2 de 0.631. Luego, en base al análisis del p-valor y FIV se discriminaron las variables menos significativas. En la Tabla 4.1 se presentan los coeficientes de las variables explicativas, el error estándar, el estadístico-t y el p-valor para cada una de ellas. El modelo final se obtiene con las siguientes variables: σ^0 , PP y HR , arrojando un R^2 de 0.630.

Tabla 4.1: Coeficientes, errores, estadístico t y p-valor de la ecuación final obtenida con MLR.

Parámetro/Variable	Coeficientes	Error Estándar	Estadístico-t	p-valor
Constante	-24.6724	11.022	-2.238	0.029
Retrodispersión	1.8289	0.529	3.456	0.001
Precipitación	6.6740	0.84	7.943	0.000
Humedad relativa del aire	0.555	0.114	4.871	0.000

Los coeficientes obtenidos dependen de las unidades de cada variable, ya que las variables no fueron normalizadas. Por lo que un mayor valor del coeficiente no representa un mayor peso en el modelo. La importancia relativa de las variables está dada por el estadístico-t, que asigna el mayor valor a PP , seguido por HR , y por último σ^0 .

Una vez establecido el modelo, se continuó con el análisis de multicolinealidad mediante el parámetro FIV. Las variables PP , HR y σ^0 presentan valores de FIV similares siendo 1.045, 1.074 y 1.075, respectivamente, lo que indica que las mismas no son co-lineales.

Posteriormente, con las variables PP , HR y σ^0 se procedió a calcular los coeficientes del modelo final. Para esto se determinó la cantidad óptima de datos para entrenar y validar el modelo mediante el análisis de la curva de aprendizaje, que utiliza el error de entrenamiento y el error verdadero de predicción. El error de entrenamiento se obtiene comparando los valores simulados de HS con los datos de HS utilizados para

obtener los coeficientes de la regresión; mientras que el error verdadero es el error de validación, es decir el que se obtiene contrastando los resultados de las simulaciones obtenidas con valores observados de HS no utilizados para definir los coeficientes del modelo. Por consiguiente se fue variando la cantidad de datos utilizados para estimar los coeficientes y los errores resultantes a fin de obtener un porcentaje óptimo de datos para cada proceso, calibración/entrenamiento y validación.

Como resultado de la aplicación del procedimiento de entrenamiento, surgió que el cálculo de los coeficientes se realizó con 30% de los datos (18 registros), y el 70% restante se utilizó para realizar la validación del mismo. Los coeficientes de determinación de entrenamiento y validación fueron 0.65 y 0.60, respectivamente, siendo el RMSE de entrenamiento igual a 5.68 %Gravimétrico, mientras que el de validación fue de 5.41 %Gravimétrico. Los valores de RMSE que arrojó el modelo son comparables con los publicados con el radar TRMMPR (Puri et al., 2011). Kseneman et al. (2012) aplican un modelo de redes neuronales para obtener HS con imágenes de TerraSAR-X reportando valores de RMSE de 13 %. Utilizando imágenes TerraSAR-X con diferentes ángulos de observación, Baghdadi et al. (2012) obtienen errores RMSE de aproximadamente 4 %.

Finalmente, el modelo queda estructurado como sigue:

$$\widehat{HS} = -36.744 + 1.426 \sigma^0 + 6.655 PP + 0.667 HR \quad (4.1)$$

Esta ecuación permite estimar los valores de HS en la zona en estudio, utilizando información meteorológica y SAR en banda X.

Luego, se efectuó un análisis del aporte que cada variable *per se* hace a la dinámica de HS con el objetivo de comprobar el supuesto de que HS puede aproximarse como la superposición de funciones lineales mencionado anteriormente. Para esta tarea, se calibraron y validaron tres modelos en donde cada uno incorpora una variable de entrada adicional. El primer modelo configurado consideró solo los valores de σ^0 como variable de entrada, debido a la conocida relación lineal entre la retrodispersión y la HS para suelos desnudos (Holah et al., 2005). Con este modelo se obtuvo un RMSE de validación igual a 7.68 %Gravimétrico [que representa 19.7% del valor de HS máximo] (Figura 4.1) y un R^2 igual a 0.201. Esto demuestra que la retrodispersión es una variable relativamente significativa, pero no explica la HS por si sola.

El segundo modelo configurado incorpora como variable de entrada, además de la retrodispersión, la PP. En este análisis, la incorporación de PP redujo el error de validación RMSE a 6.48 %Gravimétrico [16.62% de HS máx.]. La Figura 4.2 mues-

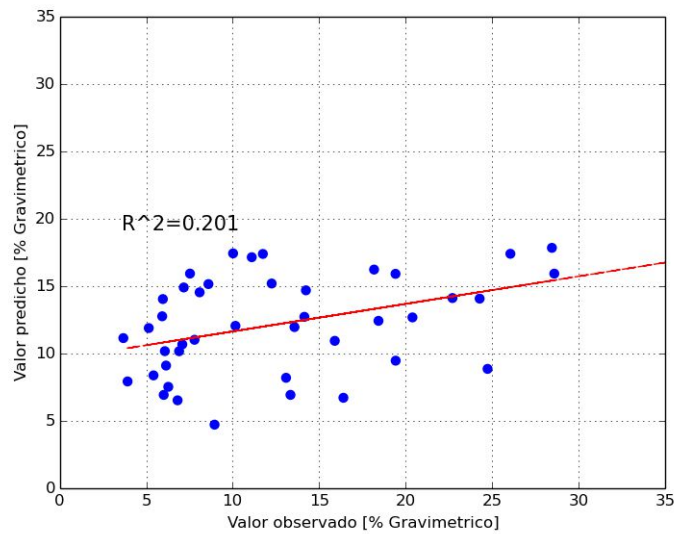


Figura 4.1: Modelo MLR con retrodispersión.

tra la validación de este nuevo modelo que arrojó un R^2 igual a 0.431. Sin duda la precipitación introduce información sustancial pero sigue siendo insuficiente.

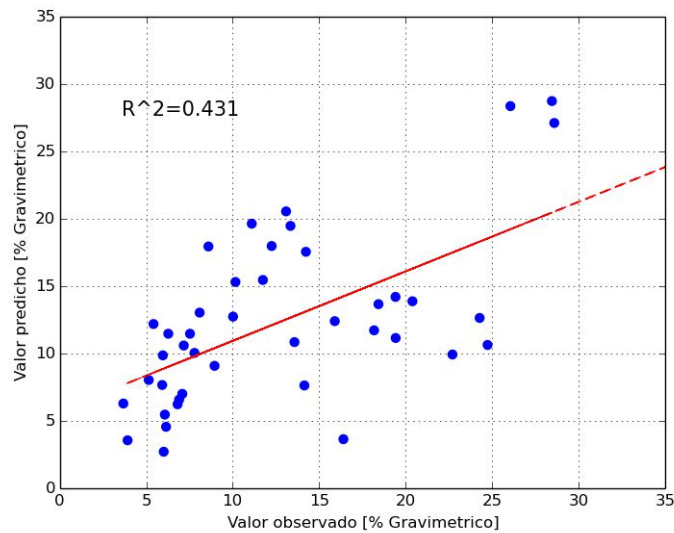


Figura 4.2: Modelo MLR con retrodispersión+precipitación.

Por último, se incorporó la variable HR y se validó el modelo con todas las variables. En este caso, el contraste con los datos de campo arroja un $RMSE=5.41$ %Gravimétrico [13.89% del valor de HS máx.] y un R^2 igual a 0.603 (Figura 4.3). Utilizando HR , el $RMSE$ se reduce y además el R^2 mejora considerablemente.

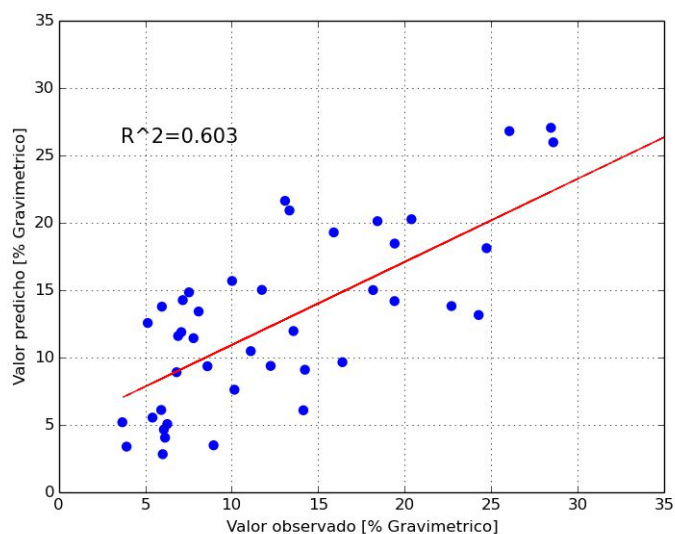


Figura 4.3: Modelo MLR Retrodispersión + Precipitación + Humedad relativa del aire.

Este análisis mostró que la HS puede ser representada como una combinación de funciones lineales, y que la retrodispersión no explica por sí sola la HS. Además, puso en evidencia que representar el ingreso o egreso de agua solamente no explican la HS, por el contrario es necesario representar ambos procesos simultáneamente.

Se debe destacar que los datos meteorológicos utilizados en este trabajo, podrían estar subestimando la variabilidad climática espacial de la zona en estudio. Es por esto que, de disponer una mayor distribución espacial de la precipitación y de la humedad relativa así como también de un mejor índice de verdor, se esperaría que los resultados mejoren. Sin embargo, en este trabajo se mostró que es factible construir un modelo, aplicando un principio hidrológico simple a las imágenes SAR, que representa la dinámica de la HS en una zona de llanura. También de este estudio surgió la necesidad de explorar nuevas misiones SAR cuyas longitudes de onda sean más sensibles a la HS.

4.2. Localidad de Monte Buey

Como se mencionó antes, la cantidad de muestras disponibles es fundamental para poder aplicar los métodos seleccionados, por lo que se buscó una región donde se dispusiera de observaciones de HS con una resolución espacial y temporal adecuada para este estudio. En la Sección [3.2.2](#), se explicó que CONAE administra una red de

estaciones en la localidad de Monte Buey que forman parte del sitio experimental para la futura misión satelital SAOCOM. Esta red de monitoreo consta de 20 estaciones que se distribuyen en las cercanías de Monte Buey donde se registra la HS, entre otras variables, cada una hora.

Con las mediciones *in situ* y los datos satelitales disponibles para este área de estudio se resolvió la ecuación de balance descrita en la metodología. Cabe señalar que la Ecuación 1.1 requiere del conocimiento de la cubierta de vegetación sobre el punto de medición para poder cuantificar la HS correctamente. Dado que no se contaba con la posibilidad de determinar t^2 de la Ecuación 1.1, se aplicó el modelo para suelo desnudo solamente.

A continuación se explica en detalle la aplicación de cada uno de los pasos definidos en el diagrama de flujo de la Figura 2.3. El conjunto de datos de entrada está compuesto por datos meteorológicos (T_a , e_a , PP y HR), observaciones satelitales (σ^0 y $NDVI$) y datos *in situ* (T_s y HS). El primer paso del procedimiento es procesar las imágenes satelitales para realizar la calibración y geo-referenciación de la información.

Luego, los datos deben ser preparados para su procesamiento, ya que cada grupo: meteorológico, satelital o *in situ*, se registra con diferente resolución espacial. Para obtener datos satelitales y meteorológicos compatibles se aplican técnicas de re-escalado. Las diferentes fuentes de información con sus resoluciones espaciales y operaciones de transformación utilizadas se presentan en la Figura 4.4.

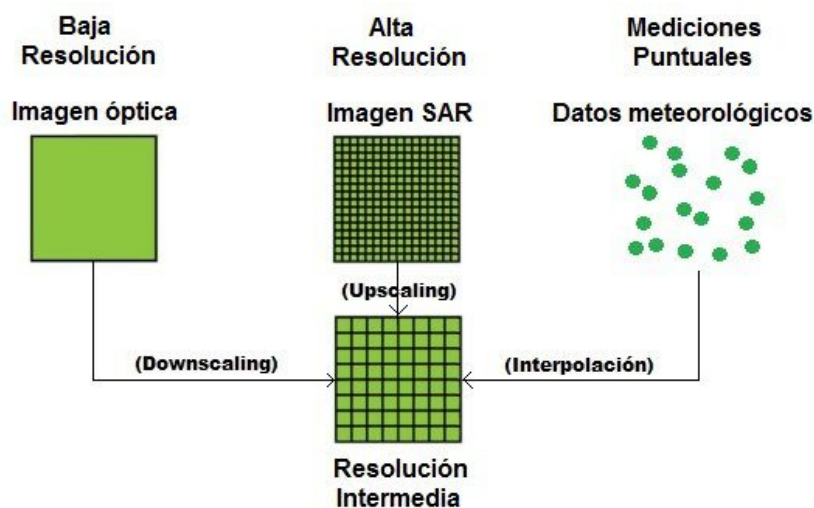


Figura 4.4: Integración de resoluciones espaciales.

Los datos de satelitales obtenidos con sensores ópticos suelen tener una resolución espacial más gruesa que los datos de sensores SAR. Por lo tanto, para poder superponer

ambas fuentes de datos, se debe modificar la resolución espacial de uno de ellos o ambos. En este trabajo, la resolución espacial de los datos ópticos se modificó mediante el proceso conocido como “downscaling”, usando la transformada ondita discreta 2-D (DWT-2D, por sus siglas en inglés), implementada a través de una extensión del algoritmo piramidal propuesto por Mallat (Mallat, 1989). Este algoritmo consiste en un operador de sobremuestreo junto con operaciones de filtrado sobre las filas y columnas de la imagen, los filtros empleados son paso bajo y paso alto en una dimensión (1-D). En este proceso, las imágenes MODIS NDVI con resolución espacial original de 250 m se redujeron a aproximadamente 30 m. Se emplearon diferentes familias de onditas (es decir, *Daubechies*, *Coiflet*, *Symlets* y *Haar*) con orden dos y cuatro y se analizaron los errores de las imágenes resultantes. Las imágenes NDVI obtenidas fueron contrastadas con imágenes NDVI del satélite Landsat 8. El RMSE más pequeño, aproximadamente 18 % del valor máximo de NDVI, se logró con la familia *Coiflets* de orden dos. La técnica utilizada y los resultados obtenidos se detallan en el Anexo C. Por otro lado, en este trabajo, se ha utilizado el producto MODIS NDVI con la fecha de adquisición más cercana a la de Sentinel-1. Luego, la resolución espacial de las imágenes SAR se modificó a 30 m, mediante el proceso conocido como “upscaling”, aplicando la técnica denominada “multilooking”.

Los datos meteorológicos se obtienen de estaciones meteorológicas distribuidas espacialmente en forma muy dispersa. Esta información es representativa de un área limitada alrededor de la estación. Por lo tanto, se debe interpolar para obtener una distribución espacial sobre el área de interés acorde a las imágenes. Para esto, se utiliza el método de interpolación de distancia inversa (IDW, por sus siglas en inglés) (Watson and Philip, 1985). Para estimar el valor de la variable en una ubicación donde no se la mide, el método hace uso de los datos observados en estaciones próximas a la ubicación donde debe predecirse el valor. IDW asume que cada punto medido tiene un efecto local que disminuye con la distancia. En consecuencia, los valores medidos más cercanos a la ubicación de predicción tienen más influencia sobre el valor interpolado que los que están más lejos.

Los datos transformados se organizan en una matriz de entrada donde cada fila es un conjunto de variables relacionadas con una única medición de HS, y cada columna es una variable distinta. Como resultado se obtiene una matriz de 248 muestras y seis columnas.

Una vez conformada la matriz, se aplica una operación de depuración a los datos de entrada que consiste en eliminar la fila de la matriz cuando cualquier variable está fuera

de un rango establecido. Para definir el rango de variabilidad de HS y σ^0 en este trabajo se ha utilizado conocimiento previo de cada una de ellas. El rango de variabilidad de HS se definió analizando los datos de campo. El rango de HS observado en el área de estudio por CONAE varió de 5 a 45 %Vol; por lo tanto ese fue el rango válido adoptado para la simulación (Thibeault et al., 2015). Para definir el rango válido de σ^0 se utilizó un análisis de la variación de σ^0 para diferentes coberturas de terreno, el rango seleccionado fue entre -18 y -4 dB (Liu et al., 2016). El criterio utilizado para eliminar los valores atípicos en las otras variables se basó en el estadístico percentil, que permite seleccionar el bloque de datos más representativo de cada variable (Helsel and Hirsch, 2002). Aquí, la depuración se realizó con el percentil 95 % de T_a , e_a , PP , RH y T_s . Después de aplicar esta operación, el número de muestras se reduce a 184.

La variación temporal de cada variable se presenta en la Figura 4.5, donde cada punto representa el valor promedio en el área de estudio. La serie temporal de PP (PP acumulada durante siete días antecedentes a la fecha de la imagen) muestra dos estaciones: una estación húmeda de septiembre a abril y una estación seca de mayo a agosto. La estacionalidad también se observa en la serie de HS, en contraste, T_a , T_s y e_a parecen estar correlacionadas.

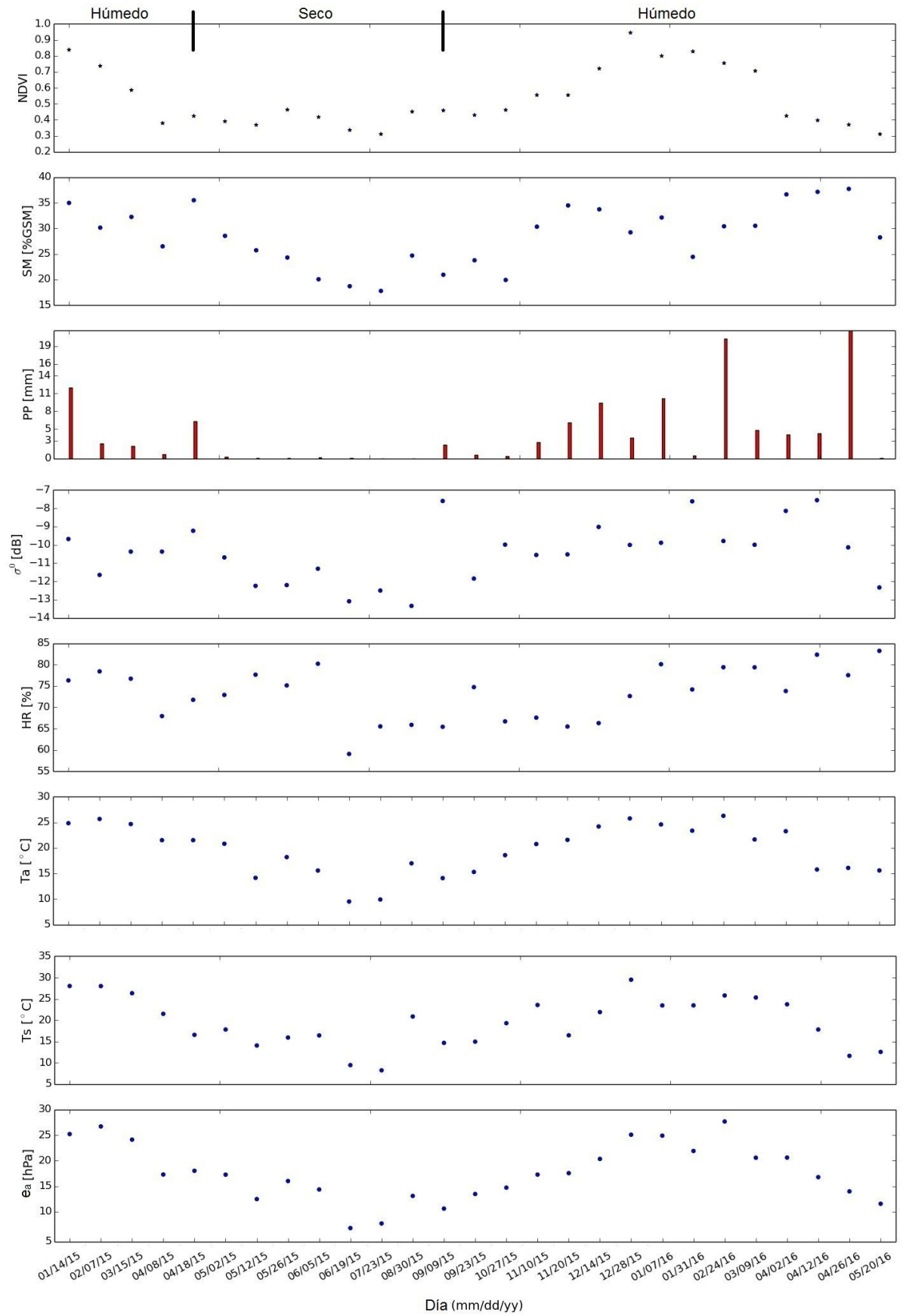


Figura 4.5: Variación temporal de las variables.

La matriz depurada se filtra mediante el índice NDVI para seleccionar los datos registrados en suelos desnudos y con escasa vegetación. La matriz de entrada depurada final solo contiene las muestras con valores de NDVI entre 0 y 0.5. Como resultado, la cantidad de muestras seleccionadas para la calibración y validación es 111. Los estadísticos de cada variable obtenidos con la matriz filtrada se presentan en la Tabla 4.2.

Tabla 4.2: Estadísticos de las variables de entrada: mínimo (Mín), máximo (Máx), media y S.

Variable	Unidades	Min	Max	Media	S
HS	%Vol.	6.25	43.36	26.26	8.00
σ^0	dB	-17.82	-4.53	-10.78	3.00
T_a	$^{\circ}C$	9.25	24.7	17.00	4.15
T_s	$^{\circ}C$	7.00	28.00	16.28	4.75
e_a	hPa	6.76	24.2	14.26	4.00
PP	mm	0.00	22.00	1.96	4.23
HR	%	58.00	83.63	71.74	6.64
NDVI		0.24	0.50	0.37	0.07

Cualquiera de las metodologías propuestas requiere variables estadísticamente independientes. Por lo tanto, la dependencia mutua de las variables de entrada se estableció utilizando el coeficiente de correlación de Pearson (r), como se muestra en la Figura 4.6. Esta figura sugiere que existe una relación lineal (también llamada co-linealidad) entre las variables T_s y T_a con un r mayor que 0.8. Este hallazgo es respaldado por Lin et al. (2016), quien mostró que T_s aumenta al aumentar T_a . Por otro lado, e_a está relacionado con T_a mediante una función exponencial (Buck, 1981), y se observa una relación similar en el gráfico de dispersión entre ambas variables, con r mayor que 0.9. Por lo tanto, e_a también está relacionado con T_s , el coeficiente r obtenido es igual a 0.83. La variable σ^0 no mostró ninguna relación lineal clara con otras variables. El diagrama de dispersión entre σ^0 y HS no sugiere una tendencia lineal entre ambas variables ($r = 0.3$), aunque HS aumenta al aumentar σ^0 , lo que indica que hay una relación no lineal. Cabe señalar que σ^0 presentó valores muy bajos de r , mientras que T_s tiene un valor menor de r .

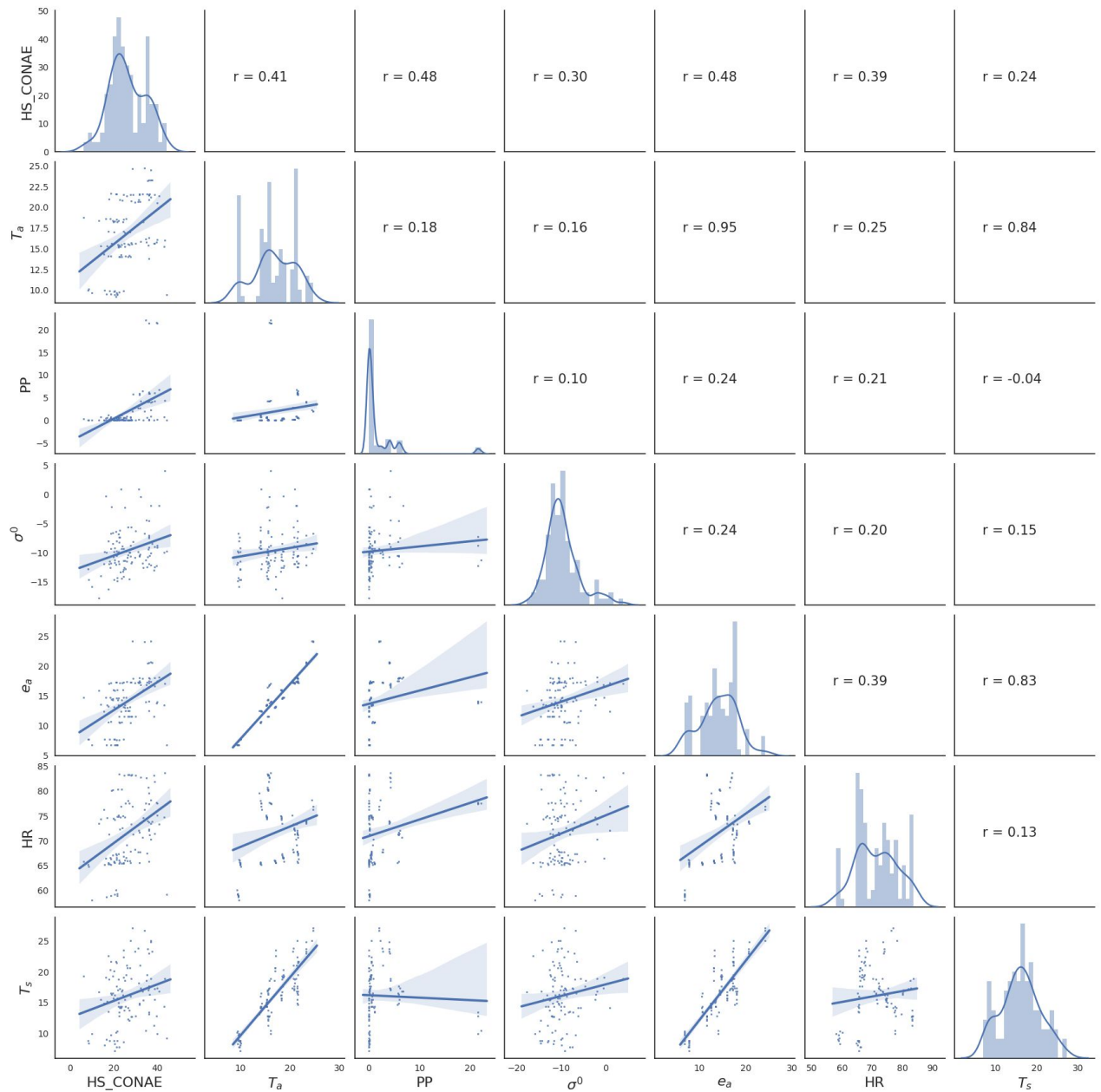


Figura 4.6: Matriz de correlación entre HS CONAE, T_a , PP, σ^0 , e_a , HR y T_s .

Como se mencionó anteriormente, es necesario conocer y cuantificar la cubierta vegetal sobre los puntos de medición para poder eliminar las áreas densamente vegetadas de la estimación de HS. Con esta tarea se busca reducir el efecto significativo de la vegetación sobre la señal de retrodispersión. Es por esto que la fenología de la vegetación limita la cantidad de datos disponibles para este tipo de modelos. La Figura 4.7 muestra una clasificación de la cobertura vegetal según los valores de NDVI, para cuatro imágenes representativas de las diferentes estaciones del año. En cada imagen, el NDVI se discriminó en cuatro clases: agua, suelo desnudo, vegetación escasa y

vegetación densa.

Cabe señalar que durante la época de invierno (Figura 4.7 a), la vegetación densa cubría solo 10.6 % del área; mientras que los píxeles con NDVI mayor que 0.5 aumentaron de 28.3 % en primavera a 99 % en verano (Figura 4.7 b y 4.7 c). Al final del verano, inicio del otoño (Figura 4.7 d) la vegetación densa bajó a 77 %. En base a la variación de la cobertura vegetal presentada en la Figura 4.7 la metodología propuesta sería aplicable durante la temporada de siembra y en los meses no lluviosos (de abril a septiembre).

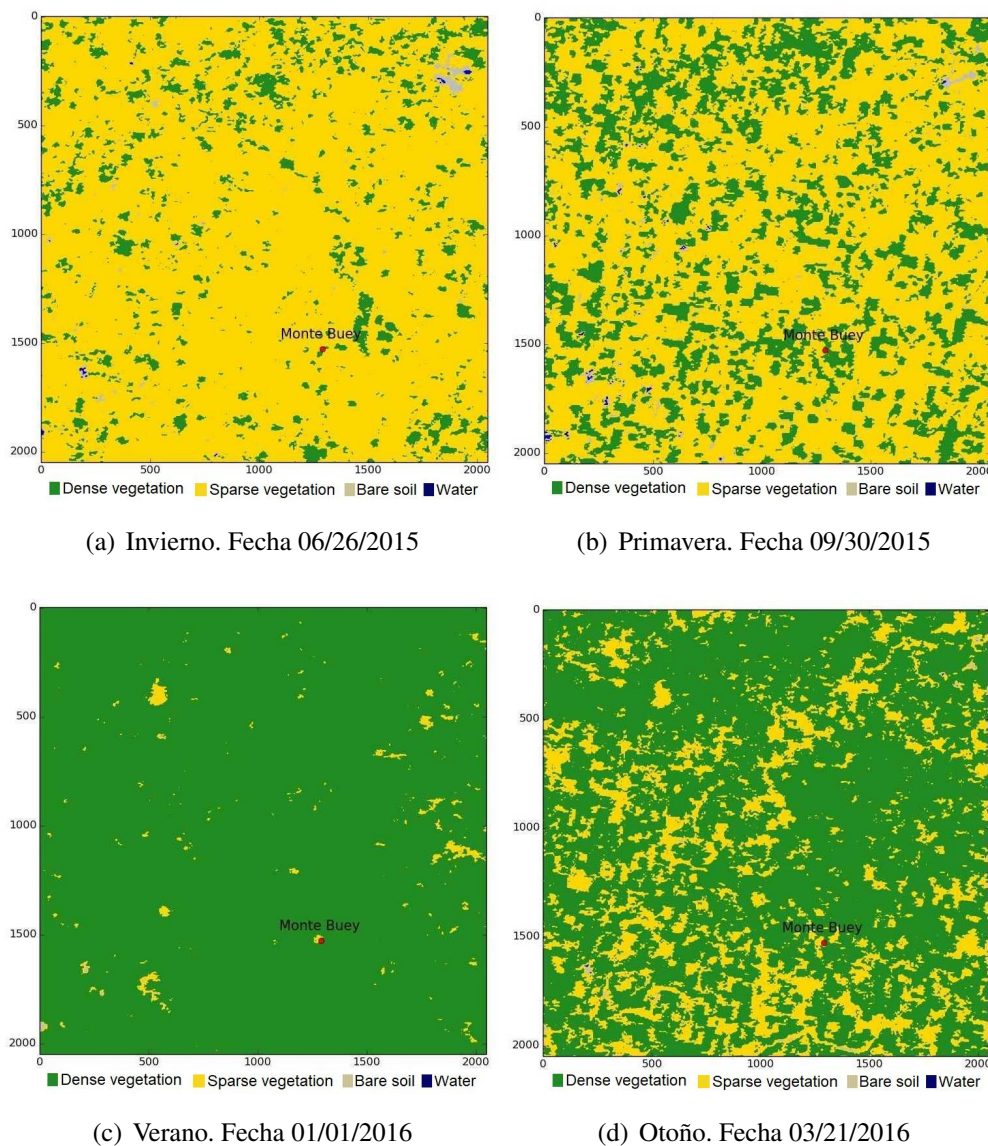


Figura 4.7: Clasificación de la cobertura con el producto NDVI MOD13Q1 a 30m: agua (Azul); suelo desnudo (Gris); vegetación escasa (Amarillo); vegetación densa (Verde).

4.2.1. Calibración de los métodos MLR y MLP

La ecuación de balance hídrico propuesta en la metodología (Ecuación 2.1) expresa la variabilidad de HS como la superposición de funciones no lineales que dependen de variables meteorológicas y propiedades del suelo. Para resolver esta ecuación con el método MLR las relaciones entre variables hidrometeorológicas y los procesos deben ser linealizadas, como se propuso en la Ecuación 2.3. Por lo que, la HS queda expresada de la siguiente manera:

$$\log(\widehat{HS}) \approx \log(T_a, T_s, HR, e_a) + \log(\text{propiedades del suelo}) + PP \quad (4.2)$$

donde las variables hidrometeorológicas y propiedades del suelo representan los procesos ET y Q , respectivamente.

Para aplicar el método MLR, primero las variables en la matriz de entrada fueron transformadas mapeando el rango de variabilidad de cada una al rango $[0,1]$. Luego, se realiza la selección de las variables que mejor explican la variabilidad de HS. Para ello se implementa el procedimiento denominado eliminación hacia atrás. Las variables en la matriz filtrada se transforman aplicando logaritmo a HS, T_a, HR, T_s, e_a . En esta tesis el coeficiente de retrodispersión se utiliza para representar las propiedades del suelo. Además, σ^0 está vinculado a HS por una función logarítmica como se propone en Zribi et al. (2005) y Baghdadi et al. (2012), por lo tanto los valores de σ^0 no son modificados en la matriz.

En la literatura revisada durante el desarrollo de esta tesis, no se encontró una relación general entre PP y HS. Por esto, en este trabajo se propone una función ajustada empíricamente a los datos observados, de la forma: $HS = 0.092 e^{0.093PP}$. En la Figura 4.8, se presentan los datos observados junto a la ecuación ajustada. Por lo tanto, PP se incluye en la matriz de entrada sin transformaciones adicionales.

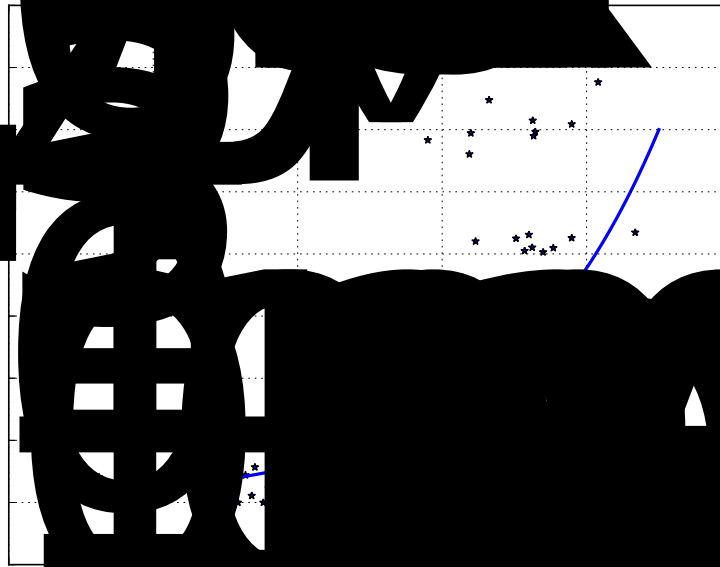


Figura 4.8: Relación entre PP y HS.

El primer modelo MLR creado con la matriz de datos arrojó un R^2 de aproximadamente 0.43. El estadístico p-valor fue significativo (valor $p < 0.05$) para σ^0 , HR y PP . El p-valor para e_a , T_s y T_a fue mayor que 0.05, siendo e_a el más crítico. Por otra parte, el análisis del estadístico FIV indicó a e_a como la variable más colineal seguida de T_a . Al aplicar los criterios de colinealidad y significancia combinados, se eliminó e_a del conjunto de datos y se volvió a aplicar el MLR. Los nuevos resultados mostraron que no había colinealidad entre las variables, sin embargo el p-valor de T_s fue de 0.538 por lo que la variable se eliminó del conjunto de datos, y se calcularon los coeficientes del método nuevamente. El modelo resultante luego de eliminar T_s arrojó valores de p-valor y FIV aceptables para todas las variables involucradas.

La matriz final de trabajo quedó compuesta por PP , $\log(T_a)$, $\log(HR)$ y σ^0 , que luego fue dividida en dos matrices: una para calibración y otra para validación. Tal como se explicó en la Sección 4.1, el número de muestras óptimo para calibrar el modelo se obtiene por medio del análisis de la curva de aprendizaje usando el RMSE. El número de datos óptimo se encuentra para un tamaño de conjunto de calibración del 75 % de los datos.

El conjunto de calibración, compuesto por 83 muestras, se utilizó para calcular los coeficientes del MLR. El modelo obtenido tiene un R^2 igual a 0.4, un RMSE de 6.36

%Vol. y un sesgo igual a 0.85 *%Vol.* Para comprobar cuan robusto es el modelo calibrado, se utilizó validación cruzada con k particiones (Haykin, 2009). El conjunto de calibración fue dividido aleatoriamente en $k = 10$ subconjuntos de datos de igual tamaño, con el 90 % de los datos para entrenar y el 10 % para probar. La precisión global del modelo se calcula promediando las medidas de precisión, R^2 y RMSE, obtenidas para los diez conjuntos de datos de prueba. Para este caso, los valores promedios de R^2 y RMSE obtenidos para el entrenamiento con validación cruzada con 10 particiones fueron 0.4 y 6.99 *%Vol.*, respectivamente.

La Tabla 4.3 resume los coeficientes del modelo, y las estadísticas p-valor y FIV. Se puede observar que las variables seleccionadas siguen siendo significativas (p-valor inferior a 0.05) y no son colineales (FIV inferior a 5).

Tabla 4.3: Estadísticas de las variables de entrada seleccionadas.

Variable	Coefficiente	p-valor	FIV
β_0	1.52	0.0	-
σ^0	0.17	0.009	1.2
PP	0.11	0.008	1.1
$\log(T_a)$	0.15	0.03	1.3
$\log(RH)$	0.82	0.009	1.1

Finalmente, el modelo MLR resultante está representado por la siguiente ecuación:

$$\widehat{HS} = 1.52 + 0.17 \sigma^0 + 0.11 PP + 0.15 \log(T_a) + 0.82 \log(HR) \quad (4.3)$$

Posteriormente, al mismo conjunto de variables se le aplicó el método MLP. La implementación del MLP como regresor se realizó utilizando la herramienta *scikit-learn* para aprendizaje automático en Python (Pedregosa et al., 2011).

En este caso, las variables de entrada no fueron linealizadas como en el caso anterior, debido a que las redes neuronales MLP pueden aproximar cualquier función no lineal arbitraria, continua y multidimensional (Hornik et al., 1989). En este caso, las variables en la matriz de entrada fueron transformadas mapeando el rango de variabilidad de cada una al rango $[-1, 1]$ para lograr un correcto aprendizaje de la red neuronal como se describió en la Sección 2.3. Luego, el conjunto de datos estandarizados se dividió aleatoriamente en subconjuntos de calibración (75 % de los datos) y validación (25 % restante).

La arquitectura del MLP apropiada para este trabajo, se obtuvo con búsqueda exhaustiva sobre el espacio de parámetros, probando métodos de optimización de pesos como quasi-Newton, gradiente descendiente (GD) estocástico y GD optimizado (Kingma and Ba, 2014). Cada uno de estos métodos tiene asociados diferentes parámetros que son presentados en la Tabla 4.4.

Tabla 4.4: Parámetros de los métodos de optimización de pesos del MLP.

Parámetro	Método		
	Quasi-Newton	Gradiente descendiente (GD)	GD optimizado
Capas ocultas	x	x	x
Función de activación	x	x	x
Penalidad	x	x	x
Tasa de aprendizaje	-	x	x
Tipo de tasa de aprendizaje	-	x	-
Momento	-	x	-

Las funciones de activación probadas fueron: sigmoidea, Unidades Lineales Rectificadas (ReLU, por sus siglas en inglés) y tangente hiperbólica. Dentro del método GD, se probaron diferentes tipos de tasas de aprendizaje como ser: *constante*, *invscaling* y *adaptativo*. El tipo *invscaling* disminuye gradualmente la tasa de aprendizaje en cada paso de tiempo; el tipo *adaptativo* mantiene constante la tasa de aprendizaje siempre que el error de entrenamiento siga disminuyendo, es decir, si en dos épocas consecutivas no se disminuye el error de entrenamiento por al menos una tolerancia, la tasa de aprendizaje se divide por 5.

También se probaron métodos automáticos de optimización de parámetros como búsqueda de grilla o barrido de parámetros y búsqueda aleatoria (Bergstra and Bengio, 2012), ambos con validación cruzada (k-fold). Hasta el momento no se han obtenido mejores resultados, aunque cabe destacar que no se realizó una aplicación exhaustiva de los mismos.

El número de neuronas de entrada fué cuatro, una para cada una de las variables propuestas. Durante la búsqueda exhaustiva se probaron diferentes números de capas ocultas y de neuronas en cada una, junto a los distintos métodos de optimización de pesos, para obtener la mejor configuración que garantice un rendimiento óptimo del

modelo (error mínimo y máxima convergencia), evitando cualquier posible sobreajuste. La red fue calibrada y probada para cada prueba buscando el RMSE más bajo.

La arquitectura MLP resultante tiene cuatro capas, es decir, una capa de entrada con cuatro neuronas; dos capas ocultas con tres y cuatro neuronas, respectivamente, y una capa de salida con una neurona. El método de optimización de pesos que arrojó mejores resultados fue GD junto con la función de transferencia ReLU para las capas ocultas y el tipo de tasa de aprendizaje adaptativo. El R^2 y el RMSE para esta configuración fueron 0.58 y 4.2 %Vol., respectivamente. Por lo tanto, la arquitectura del modelo MLP adoptada tiene una configuración de 4-3-4-1 neuronas.

Una vez derivada la arquitectura, el MLP fue calibrado con un R^2 igual a 0.54, un RMSE de 5.73 %Vol. y un sesgo igual a 0.01 %Vol. En la Tabla 4.5 se presentan los pesos sinápticos de la arquitectura MLP calibrada.

Tabla 4.5: Pesos sinápticos (W_{ij}) para el método MLP con arquitectura 4-3-4-1.

Variable de entrada	W_{i1}	W_{i2}	W_{i3}	W_{i4}
W_{ij} entre las entradas y la 1er capa oculta				
T_a	0.63	-0.30	1.15	
PP	2.50	-1.97	2.06	
σ^0	0.56	-0.33	-0.05	
HR	-0.18	-0.55	1.16	
W_{ij} entre la 1er y la 2da capa oculta				
W_{1j}	0.05	-2.43	-0.51	-0.74
W_{2j}	0.47	-1.55	-0.85	0.39
W_{3j}	-0.08	-0.69	0.79	0.46
W_k entre la capa oculta y la capa de salida				
W_1	0.89			
W_2	4.14			
W_3	3.47			
W_4	0.75			

En la Figura 4.9 se puede observar que el modelo MLP obtiene una correlación levemente mejor con respecto a los valores de HS observados, en comparación con el modelo MLR.

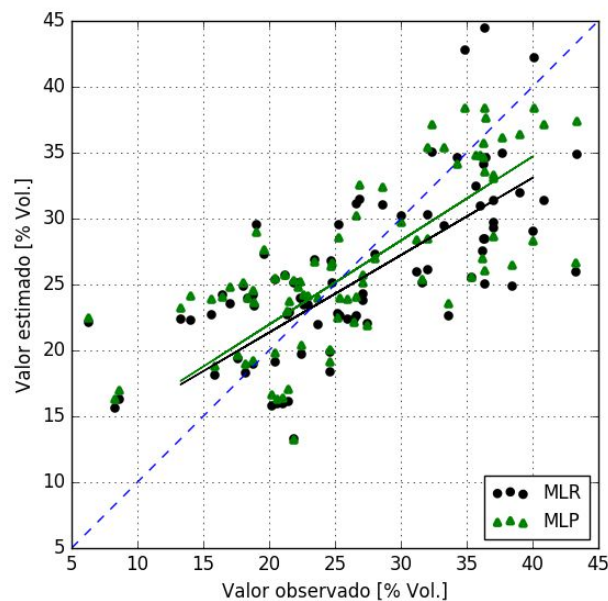


Figura 4.9: Comparación entre los valores de HS observados y estimados que resultan en la fase de calibración para los métodos MLR (Círculos) y MLP (Triángulos).

4.2.2. Validación de los métodos

Posteriormente, se validan los modelos resultantes para evaluar la bondad del ajuste de los mismos. Esta operación implica aplicar la metodología a un conjunto de datos que no se utilizó en la definición de los coeficientes del MLR y de los pesos del MLP. Luego, los valores de HS estimados se comparan con los valores de HS observados. Para cuantificar la precisión de la metodología, se calculan los índices propuestos en la Sección [2.4.1](#). El conjunto de datos de validación está compuesto por 28 muestras.

Para el método MLR, la comparación entre los valores estimados y observados de HS arrojó un R^2 igual a 0.58, un RMSE de 4.25 %Vol. y un sesgo de 0.25 %Vol. (Ver Figura [4.10](#) - círculos). Para el método MLP, la comparación entre los valores estimados y observados de HS arrojó un R^2 igual a 0.64, un RMSE de 4.17 %Vol. y un sesgo = -1.25 %Vol. (Ver Figura [4.10](#) - triángulos). El error obtenido por ambos modelos representa un 10 % de los valores máximos observados de HS. Se puede observar en la Figura [4.10](#) que los valores de HS estimados en torno a 25 %Vol. están adecuadamente representados por ambos modelos, sin embargo éstos sobreestiman las observaciones de HS menores a 25 %Vol. y subestiman las mayores a 25 %Vol.

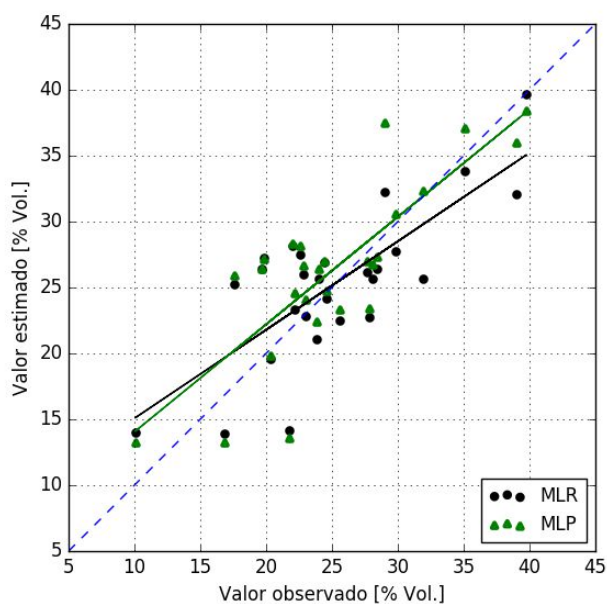


Figura 4.10: Comparación entre los valores de HS observados y estimados que resultan en la fase de validación para los métodos MLR (Círculos) y MLP (Triángulos).

4.2.3. Discusión de los resultados

La metodología propuesta en esta tesis tiene una estructura simple y es adaptable a los tipos de datos fácilmente disponibles. Los modelos implementados arrojaron resultados comparables con los publicados por otros autores utilizando enfoques diferentes. La Tabla 4.6 resume los resultados de los métodos disponibles en la literatura. En general, los errores varían de 3 a 20 %Vol., siendo máquina de vectores de soporte (SVR, por sus siglas en inglés) el método que mejor estima HS en terrenos con pendientes. Los modelos de inversión presentan una amplia gama de errores que básicamente dependen de la cobertura del suelo. El modelo MLR propuesto produce errores de aproximadamente 4 %Vol. comparable con los errores publicados por Merzouki et al. (2011), y Lievens and Verhoest (2012) con banda C en áreas de llanura con vegetación. En particular, los resultados presentados en este documento coinciden con los publicados por Paloscia et al. (2008), quienes compararon varios algoritmos para recuperar HS en una zona agrícola plana en el norte de Italia. Los autores encontraron errores entre 3 y 9 %Vol. Además, los resultados obtenidos por Lievens and Verhoest (2012) con un enfoque físico, utilizando imágenes RADARSAT-2 en topografía plana, resultaron similares a los obtenidos en este trabajo con MLR.

Tabla 4.6: Resumen de las metodologías para la estimación de HS a partir de datos SAR.

Metodología	Banda	Autores	Topografía	Cobertura	RMSE (%Vol.)
ANNs	C	Paloscia et al. (2008)	Llanura	Vegetado	9
		Paloscia et al. (2010)	Montaña		10
	X	Kseneman et al. (2012)	Llanura	Suelo desnudo	14
SVR	C	Pasolli et al. (2011)	Montaña	Vegetado	3
		Bertoldi et al. (2014)			5
Inversión	C	Paloscia et al. (2008)	Llanura	Vegetado	12
		Merzouki et al. (2011)			6
		Gherboudj et al. (2011)			20
		Lievens and Verhoest (2012)			4
	X	Baghdadi et al. (2012)		Suelo desnudo	3
Bayes	C	Paloscia et al. (2008)	Llanura	Vegetado	11
		Notarnicola et al. (2008)		Suelo desnudo	<10
		Pierdicca et al. (2010)		Vegetado	4

En resumen, la ventaja de la metodología propuesta se deriva del tipo de variables que podrían utilizarse dentro de la ecuación de balance para representar los procesos ET y Q , es decir, variables meteorológicas, propiedades del suelo, variables hidrológicas y datos satelitales. De hecho, las propiedades del suelo, las que son escasas y no siempre se encuentran en formato digital en la Argentina, y en muchas otras áreas remotas del mundo, no se requieren explícitamente para ensamblar un modelo MLR o MLP como lo es en los modelos de inversión.

Como desventaja, puede mencionarse que la banda C es muy sensible a la cobertura vegetal, por lo que en la literatura no se recomienda para la estimación HS (Das et al., 2011; Paloscia et al., 2008). Además, estos autores sugieren que la polarización VV no es adecuada para la recuperación de HS. Sin embargo, los resultados presentados mostraron que las imágenes de Sentinel-1 en polarización VV producen errores comparables con los obtenidos por otros autores, los que deberían mejorar si se utiliza la polarización HH de imágenes en banda C o L. Aún así, los datos de la banda C de Sentinel-1 son de acceso gratuito.

4.2.4. Análisis de incertidumbre del modelo

Los modelos presentados se basan en el ajuste de las relaciones funcionales entre HS y variables hidroambientales, por esto es necesario analizar la incertidumbre introducida por los errores típicos de las variables involucradas. Para realizar esta tarea se utilizó un análisis de primer orden.

El análisis de primer orden es apropiado para cuantificar las incertidumbres introducidas en una variable dependiente por variables y parámetros independientes debido a la incertidumbre inherente a las mediciones y estimaciones (Chow et al., 1988). Este método asume que una variable Y se expresa como una función de x , tal que $Y = f(x)$. Si el modelo f es correcto, las incertidumbres de Y están asociadas a errores de medición de x . Sabiendo que \bar{x} es el promedio de x , el promedio de Y , \bar{Y} se calcula como $\bar{Y} = f(\bar{x})$. Si el valor verdadero de x difiere de \bar{x} en cierta cantidad, el efecto de esta diferencia en Y puede examinarse expandiendo $f(x)$ como una serie de Taylor alrededor de $x = \bar{x}$. Si se descartan los términos de orden superior de la serie Taylor, la expresión del error se aproxima como se muestra en la Ecuación 4.4.

$$Y - \bar{Y} \approx \left. \frac{df}{dx} \right|_{x=\bar{x}} (x - \bar{x}) \quad (4.4)$$

La varianza del error de Y se expresa como:

$$S_Y^2 \approx \left(\left. \frac{df}{dx} \right|_{x=\bar{x}} \right)^2 S_x^2 \quad (4.5)$$

donde S_Y^2 y S_x^2 son la varianza de Y y x , respectivamente. Esta metodología puede extenderse a varias variables independientes (Chow et al., 1988).

La aplicación de este análisis requiere conocer la expresión analítica de los modelos, es decir la función f , para luego derivarla con respecto a la variable x y evaluarla en su valor medio \bar{x} . Por este motivo, el análisis de incertidumbre en esta tesis sólo se realizó para el modelo MLR, debido a la complejidad para expresar analíticamente la ecuación del modelo MLP.

Para calcular la incertidumbre del modelo MLR, la Ecuación 4.5 se aplicó a la Ecuación 4.3 de forma de derivar la varianza del error de HS:

$$S_{HS}^2 = \left(\frac{\overline{HS}\beta_{\sigma^0}}{\log(e)} \right)^2 S_{\sigma^0}^2 + \left(\frac{\overline{HS}\beta_{PP}}{\log(e)} \right)^2 S_{PP}^2 + \left(\frac{\overline{HS}\beta_{T_a}}{\overline{T_a}} \right)^2 S_{T_a}^2 + \left(\frac{\overline{HS}\beta_{HR}}{\overline{HR}} \right)^2 S_{HR}^2 \quad (4.6)$$

Reemplazando los valores medios y los coeficientes del modelo MLR de las varia-

bles se obtiene:

$$S_{HS}^2 = 105.8 S_{\sigma^0}^2 + 44.3 S_{PP}^2 + 57.4 S_{T_a}^2 + 689.6 S_{HR}^2 \quad (4.7)$$

Para examinar la ecuación 4.7, se agregó un ruido uniforme de 10 % y 20 % del valor de S a cada variable, como se muestra en la Tabla 4.7.

Tabla 4.7: Media y S de las variables de entrada para el análisis de primer orden.

Variable	Media	S	S 10 %	S 20 %
σ^0	0.52	0.22	0.24	0.26
PP	0.09	0.19	0.21	0.23
T_a	0.52	0.21	0.23	0.25
HR	0.82	0.1	0.11	0.12

Con un S de 10 % y 20 %, la incertidumbre introducida por σ^0 en la varianza de HS varía entre 6.3 y 7.4, respectivamente. HR es la variable con el mayor valor nominal que agrega un error de aproximadamente 6.9 y 9.6, con un S de 10 y 20 %, respectivamente. Sumando las varianzas de todas las variables, la incertidumbre de HS oscila entre 18 y 23 % del valor medio de HS. Este análisis mostró que los errores en el procesamiento de imágenes SAR afectarían menos en la estimación de HS, que los errores en la variables HR.

4.2.5. Mapas de humedad del suelo

Después de la calibración y validación de los modelos a escala puntual, la metodología se extiende sobre el área de estudio para generar mapas de HS. Para producir estos mapas, la imagen SAR es filtrada para eliminar áreas en las que el modelo no se implementa, tales como áreas con vegetación densa, masas de agua y ciudades. Para esta tarea, se utiliza el índice NDVI para definir una imagen con cuatro clases de cobertura, que se utiliza para enmascarar las áreas donde el modelo no es aplicable. Es bien sabido que en áreas con vegetación escasa el índice NDVI toma valores moderados, aproximadamente entre 0.2 y 0.5, en áreas de suelo desnudo NDVI tiene valores cercanos a cero (0 a 0.2), y NDVI tiene valores negativos para el agua (Holben, 1986; Lillesand and Kiefer, 1994). Por lo tanto, estos valores típicos se usan en este estudio para diferenciar los tipos de cobertura. Los modelos MLR y MLP se aplican a los píxeles que representan suelo desnudo y áreas con vegetación escasa.

La fecha de adquisición de las imágenes Sentinel-1 y Landsat 8 utilizadas en esta sección se presentan en la Tabla 4.8. Cabe señalar que el producto MODIS MOD13Q1 no es adecuado para la aplicación del modelo debido a su resolución espacial gruesa. En su lugar, se utiliza Landsat 8 que se caracteriza por tener una resolución espacial fina, adecuada para estos mapas.

Tabla 4.8: Inventario de las imágenes para generar mapas de HS.

Fechas (mm/dd/yyyy)	
Sentinel-1	Landsat 8
06/29/2015	06/18/2015
10/03/2015	10/08/2015
12/28/2015	12/27/2015
03/19/2016	03/16/2016

El índice NDVI se usó para obtener las clases arriba mencionadas, y se aplicó una técnica de clasificación supervisada para delimitar ciudades. Las imágenes con las clases resultantes se muestran en la Figura 4.11, los píxeles no utilizables se presentan en negro y los píxeles en los que se aplicaron los modelos se muestran en blanco. En estas imágenes se puede observar en detalle la evolución de la cobertura vegetal (píxeles en color negro) y cómo las parcelas van cambiando según el estadio del cultivo. En las imágenes para las fechas comprendidas en el período invierno-primavera (Figura 4.11- a y b) los píxeles con NDVI mayores a 0.5 (color negro) podrían estar asociados a pasturas o trigo. En el área de estudio, los principales cultivos se desarrollan durante el periodo estival (Figura 4.11- c y d) siendo los de mayor importancia soja y maíz.

Las imágenes de Sentinel-1 fueron re-escaladas desde un píxel de 14 m² hasta uno de 30 m², y los datos meteorológicos puntuales se interpolaron para obtener mapas con una resolución espacial de 30 m. Así, los modelos se aplicaron a los datos de σ^0 , PP , T_a y HR de los píxeles de suelo desnudo y vegetación rala (píxeles blancos en la Figura 4.11).

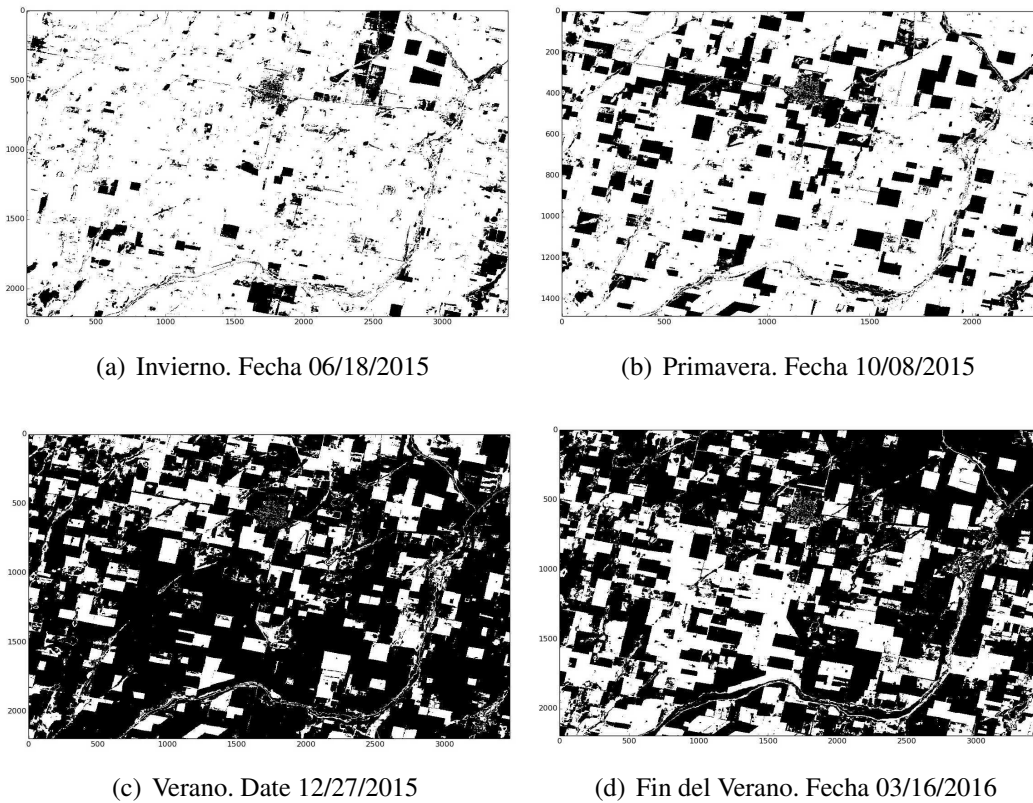


Figura 4.11: Imágenes de clases para cuatro fechas de Landsat 8. Negro: píxeles inutilizables; Blanco: píxeles donde se aplicó el modelo.

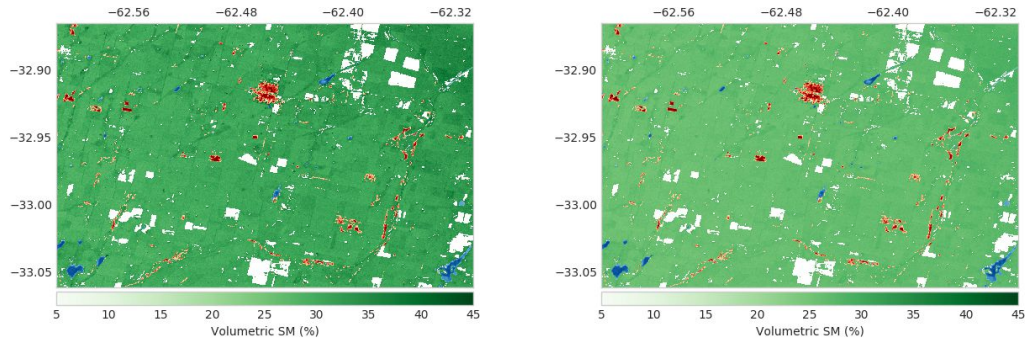
Los mapas de HS resultantes se muestran en la Figura [4.12](#), la primera columna presenta los mapas de HS estimada con el método MLR y la segunda, los mapas obtenidos con MLP, todos con resolución espacial de 30 m. En estos mapas la escala verde representa la HS, el azul se utiliza para los píxeles de agua, el rojo para las ciudades y el blanco para las zonas densamente vegetadas.

La media, Mín, Máx y S de los mapas de HS se presentan en la Tabla [4.9](#). En general, la HS estimada expresada en $\%Vol.$, varía entre 23.5 y 58.5 para el método MLR y, entre 21.3 y 39.2 para el método MLP. Existe una amplia variabilidad en el valor medio regional entre los métodos. La mayor diferencia entre los métodos se observa para las fechas de verano (es decir, 28/12/2015 y 16/03/2016) cuando el método MLR alcanzó un máximo de 58.5 $\%Vol.$ mientras que MLP alcanzó un máximo de 34 $\%Vol.$ Estas diferencias para las fechas de verano se pueden relacionar con el peso relativo que cada método asigna a PP, ya que es en esta temporada cuando se producen las mayores precipitaciones. El rango de HS estimada depende de los valores de las variables de entrada y la función de estandarización aplicada para resolver el problema de la no linealidad de la ecuación de balance. Para analizar la calidad de la estimación

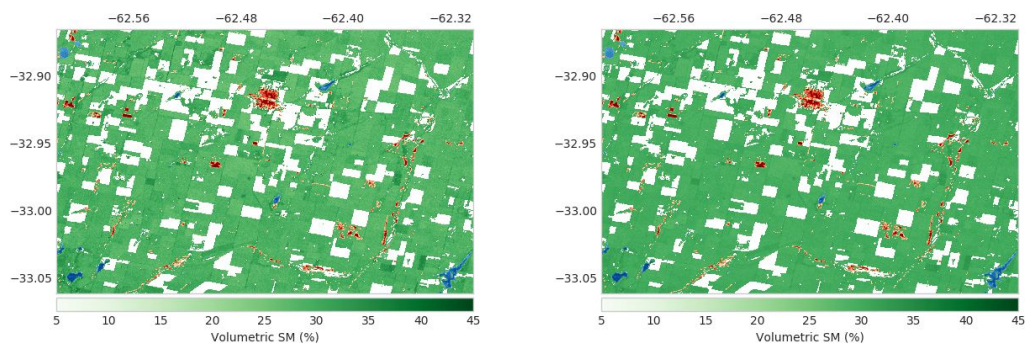
de los métodos, se calculó el RMSE para cada mapa contrastándolos con los datos *in situ* de CONAE. El mayor RMSE se obtuvo con el método MLR para todas las fechas, en particular para la fecha 16/03/2016, arrojó un error máximo de 17 %Vol. aproximadamente. Estos RMSE se redujeron utilizando el método MLP.

Aunque los valores medios de HS obtenidos con los métodos son diferentes, una comparación visual indicó que los mapas logrados con MLR y MLP proporcionaron una variabilidad espacial similar.

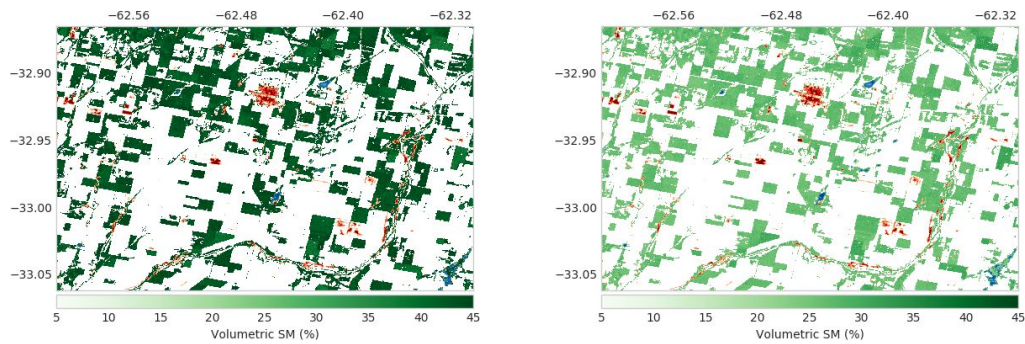
Invierno. Fecha 06/29/2015



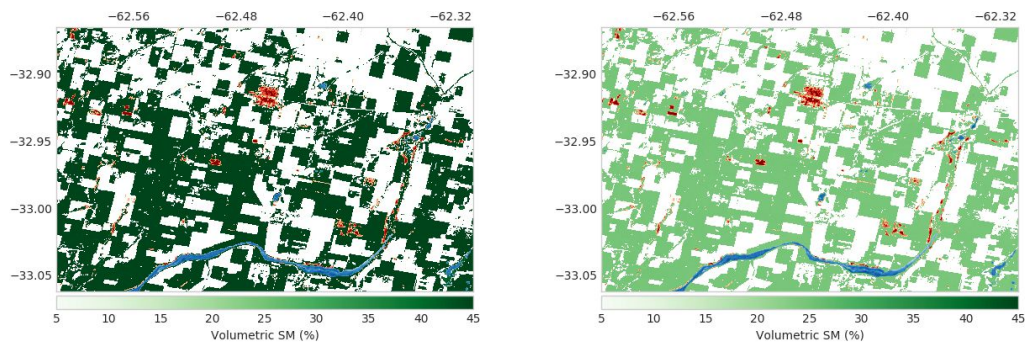
Primavera. Fecha 10/03/2015



Verano. Fecha 12/28/2015



Fin del Verano. Fecha 03/16/2016



(a)

(b)

Figura 4.12: Mapas de HS sobre el área de estudio: (a) HS obtenida con MLR, (b) HS utilizando MLP; HS (escala de Verdes); Cuerpos de agua (Azul); Ciudades (Rojo); Píxeles enmascarados donde no se aplican los modelos (Blanco).

Tabla 4.9: Análisis estadístico de los mapas de HS

Fecha	Método	Máx %Vol.	Mín %Vol.	Media %Vol.	S %Vol.	RMSE %Vol.
29/06/2015	MLR	41.9	25.6	31.3	2	14.3
	MLP	31.3	23.7	26.6	1	7.6
03/10/2015	MLR	37.1	23.5	28.7	1.7	9.5
	MLP	33.6	26.6	29.6	0.9	8.6
28/12/2015	MLR	56.5	35.9	43.5	1.7	14.9
	MLP	34	23.2	26.4	1.7	4.2
16/03/2016	MLR	58.5	37.6	48.5	2.2	17.4
	MLP	25.9	23.2	24.3	0.3	11.8

Capítulo 5

Desagregación de mapas de HS de SMAP

Los modelos encontrados hasta el momento para la estimación de HS requieren de observaciones *in situ* para su obtención. En este Capítulo se explora la posibilidad de independizar los modelos de los datos *in situ* de HS. Para lograr este objetivo se utilizan mapas de HS de SMAP en lugar de las observaciones de CONAE, y los datos de PP de las estaciones de INTA se reemplazan por datos de la misión GPM.

El desafío de esta configuración es la combinación de los mapas SMAP con datos Sentinel-1 y datos meteorológicos para obtener estimaciones de HS a escala espacial fina. La desagregación de los mapas de SMAP se realiza a partir de la resolución de la ecuación de balance de agua, aplicando las metodologías MLR y MLP.

Previo a la utilización de los datos HS, PP y ET obtenidos remotamente, se realizó un análisis comparativo de datos *in situ* con los productos provistos por las misiones SMAP y GPM sobre el área de estudio.

La comparación de los productos de SMAP y las observaciones provistas por CONAE se presentan en el Anexo [D](#). La validación del producto de HS arrojó un R^2 igual a 0.67 con un RMSE de 6.17 %Vol. (que representa un 14.4 % del valor máximo) y un sesgo de -2 %Vol. El error encontrado en el área de estudio se acercó a la precisión esperada para la misión (4 %Vol.) (Entekhabi et al., 2010). El sesgo demostró una leve subestimación en la HS. Por otro lado, en el análisis temporal de HS también se demuestra la buena relación entre los datos *in situ* y las observaciones de SMAP a lo largo del tiempo. La validación de HS presentó resultados similares o mejores a los hallados por Zeng et al. (2016). Por otro lado, la validación del producto de T_s arrojó un R^2 igual a 0.91 con un RMSE de 2.27 °K, que representa un 0.75 % del valor máximo,

y un sesgo de $-1.5 \text{ }^\circ K$.

Un análisis similar se aplicó a los datos de GPM. Se contrastaron los datos de PP obtenidos de la red de estaciones de INTA y los de la misión GPM obtenidos con el algoritmo IMERG. Primero se realizó un análisis puntual, es decir píxel-punto, con los datos de las estaciones meteorológicas. Se analizaron 105 muestras distribuidas en el período abril 2015 - mayo 2016. Las estadísticas de los datos se presentan en la Tabla 5.1. La diferencia entre los valores medios, máximos y S son 2.88 mm, 3.7 mm y 3.95 mm, respectivamente, en favor de GPM.

Tabla 5.1: Estadísticas de PP-INTA y PP-GPM en milímetros.

Dato PP	Media	Mín	Máx	S
INTA	20.17	0	136.1	26.35
GPM	23.05	0	139.8	30.03

En la Figura 5.1 se presenta el diagrama de dispersión para PP acumulada en siete días para INTA y GPM. La validación del producto de PP arrojó un R^2 igual a 0.4 con un RMSE de 25.63 mm (que representa un 19 % del valor máximo) y un sesgo de 4.2 mm. El sesgo indica que GPM sobrestima la PP medida puntualmente por las estaciones de INTA.

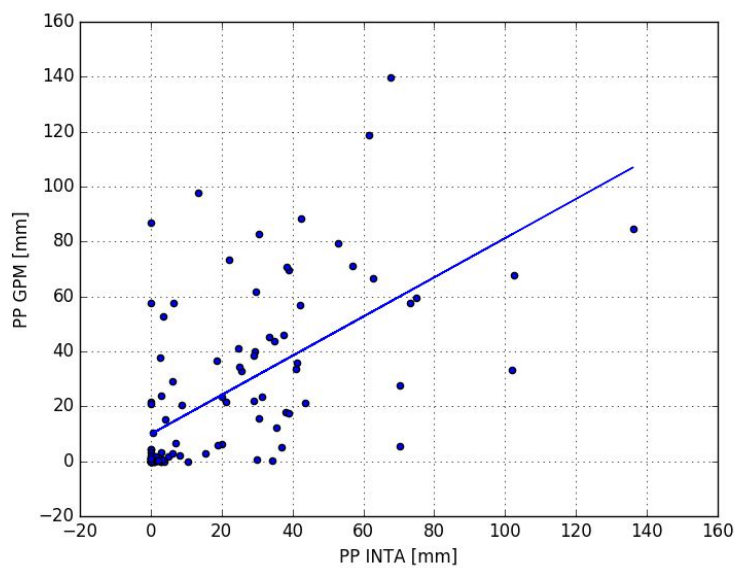


Figura 5.1: Comparación entre PP-INTA y PP-GPM en milímetros acumulados en siete días.

Para complementar la validación del producto PP se graficó la evolución temporal de la PP observada *in situ* y la obtenida por sensores satelitales. Para ello se utilizaron los valores puntuales de las estaciones y los obtenidos de los mapas de PP-GPM, tal como se observa en la Figura 5.2. El análisis temporal de HS demuestra una relación adecuada entre las observaciones *in situ* y los datos satelitales de PP a lo largo del tiempo. Se observa además, que a partir de octubre la frecuencia y el monto de las precipitaciones aumenta considerablemente. Durante el periodo 10/2015 - 05/2016, se observan las mayores diferencias entre las mediciones puntuales de INTA y los mapas GPM, lo que podría estar relacionado a eventos puntuales los que no son registrados en los mapas de GPM cuya resolución espacial es 10 km.

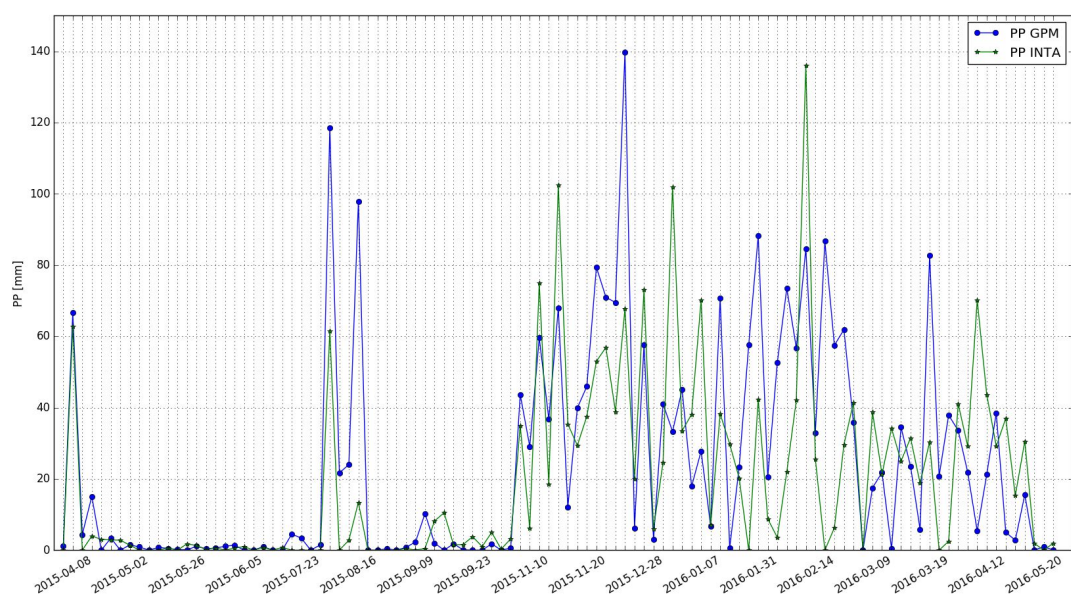


Figura 5.2: Variabilidad temporal de PP-INTA y PP-GPM en milímetros acumulados en siete días. PP-INTA (Verde); PP-GPM (Azul).

Por otro lado, el producto de ET de MODIS ha sido validado con estaciones en campo en diversos trabajos (Courault et al., 2005; Kalma et al., 2008). Estos autores han encontrado que los errores medios absolutos se encuentran entre 24.1 % y 24.6 %, para mediciones de ET realizadas con torres y calculada por el modelo Global Modelling and Assimilation Office, respectivamente.

Luego de analizar los errores de los productos satelitales, se ensambló el conjunto de datos de entrada. Este es una colección de productos satelitales (T_s , ET , PP , σ^0 y $NDVI$) y datos registrados por estaciones meteorológicas (HR). Para crear la matriz de entrada, se aplicaron técnicas de re-escalado para obtener datos con resoluciones espa-

cial compatibles. En este estudio, todos los datos se re-muestraron a una cuadrícula de 5 km^2 . Esta resolución espacial conservaría las propiedades estadísticas de la imagen original, mientras que resoluciones espaciales más refinadas requerirían de un análisis por cobertura vegetal (Van Der Velde et al., 2008).

Se probaron diferentes técnicas de remuestreo, como la interpolación del vecino más cercano, bilineal y cúbica. La revisión bibliográfica reveló que el método de interpolación bilineal es el método más usado para re-muestrear las variables geofísicas (Inamdar and French, 2009; Pasolli et al., 2011; Bai et al., 2017). Por lo tanto, como se muestra en la Figura 5.3, las imágenes de resolución espacial gruesa (HS-SMAP, T_s -SMAP y PP-GPM) y las imágenes de fina resolución espacial (σ^0 , ET y NDVI) se remuestraron a 5 km utilizando el método de interpolación bilineal.

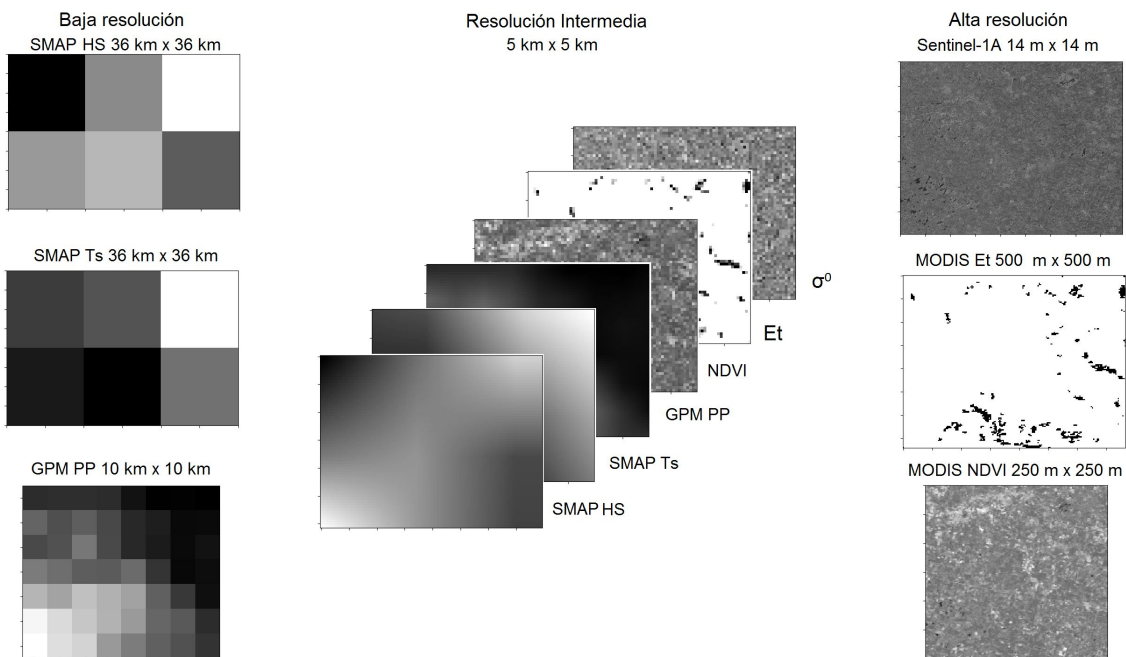


Figura 5.3: Integración de las resoluciones espaciales de las imágenes HS-SMAP y T_s -SMAP, PP-GPM, σ^0 -Sentinel y ET -Modis y NDVI-Modis.

Luego, las imágenes remuestreadas se organizaron en una matriz de entrada donde cada fila contiene la información de un píxel y cada columna una variable diferente. La matriz resultante posee 5184 filas y seis columnas. Posteriormente, durante el proceso de depuración se identificaron valores atípicos y las filas involucradas se eliminaron del conjunto de datos. Al igual que las aplicaciones anteriores, el criterio utilizado se basó en el percentil 95% que permite seleccionar el bloque de datos más representativo de todas las variables (Helsel and Hirsch, 2002).

Un píxel de SMAP representa el contenido de HS promedio para la heterogeneidad de la superficie sensada. Es decir, un píxel cuantifica el contenido de humedad para una parcela de 36 km² con distintos tipos de coberturas, desde suelo desnudo hasta bosques. En este estudio se eliminaron los píxeles que representan superficies con vegetación densa o bosques y los píxeles que representan cuerpos de agua. Por lo tanto se eliminaron de la matriz de datos los píxeles con NDVI menores a 0, que representan agua, y los píxeles con NDVI mayores a 0.8, que representan vegetación densa. El número final de muestras se redujo a 2356.

Las variables meteorológicas utilizadas en este estudio comprenden T_s , ET , HR y PP . Para detectar el tipo de relación que existe entre las variables de entrada, se utilizó la matriz de correlación, tal como se muestra en la Figura 5.4. Esta matriz sugiere que no hay una relación lineal (o co-linealidad) entre las variables de entrada. La variable σ^0 no presentó ninguna relación clara con las otras variables, excepto con HS-SMAP, cuya gráfica de dispersión sugiere una relación lineal entre ambas variables. Las gráficas de dispersión entre ET y las restantes variables no muestran relaciones, incluso con HS-SMAP.

Analizando los valores de r se encontró que, por un lado el coeficiente de correlación entre PP y HS-SMAP es mayor a 0.6, sin embargo, el gráfico de dispersión de ambas variables mostró un patrón de comportamiento no lineal. Mayores valores de PP se corresponden con mayores valores de HS, pero valores altos de HS no necesariamente están relacionados con valores altos de PP . Sorprendentemente, valores de PP elevados se asociaron a valores elevados de T_s , mientras que se espera una relación inversa. En un análisis espacial, la humedad bajaría la temperatura del suelo, mientras que los píxeles secos deberían tener mayor temperatura, pero si el análisis no es temporal no hay una relación predefinida (depende de la climatología de la zona y de los parámetros de circulación global). Sin embargo, se observó que los mayores valores de PP se registraron durante la primavera-verano del 2016. Cabe señalar que dicho año correspondió a un evento El Niño extremadamente húmedo. En contraste, los meses de otoño-invierno del 2015 fueron secos, con bajos valores de PP .

Por otro lado, el gráfico de dispersión entre HR y HS-SMAP posee un r igual a 0.475. A pesar de que estas variables no se relacionan linealmente, presentan un comportamiento escalonado, donde rangos de valores de HS-SMAP se relacionan con un único valor de HR . Un patrón de comportamiento similar ocurre entre las variables T_s y HS-SMAP, donde rangos de valores de HS toman un único valor de temperatura; y el mismo rango de valores de HS puede estar representado en distintos escalones de

T_s . Estos patrones de comportamientos encontrados entre las variables PP , HR y T_s , y HS-SMAP son significativos, por lo que al utilizarlas para desagregar la HS de SMAP se espera que aparezcan sus efectos en la HS estimada.

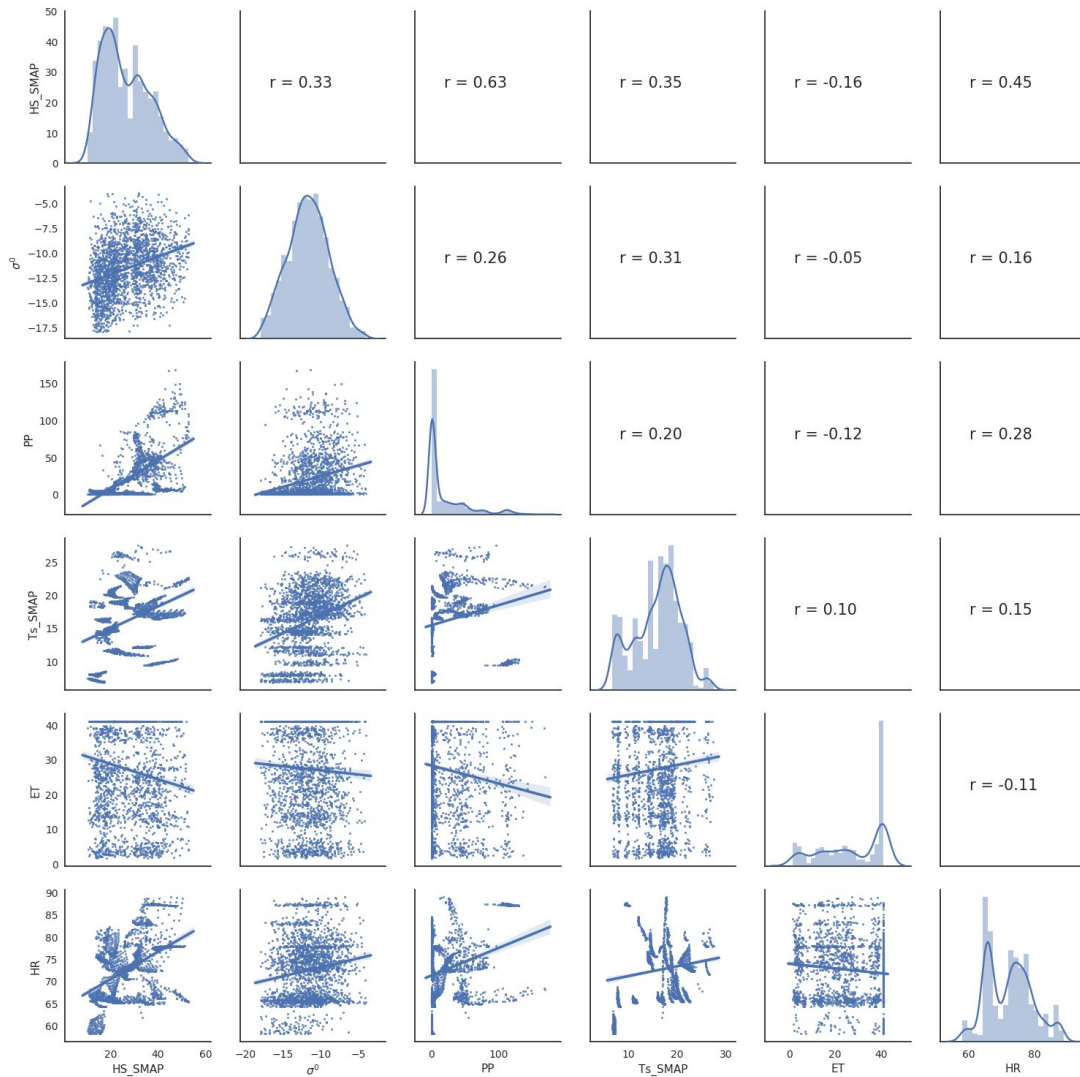


Figura 5.4: Matriz de correlación entre HS-SMAP, T_s -SMAP, ET , HR , PP -GPM y σ^0 .

En las siguientes secciones se presentan los modelos obtenidos con dos conjuntos de variables. El primer conjunto combina datos de estaciones meteorológicas terrestres (HR) y productos satelitales (PP , T_s y σ^0); mientras que el segundo conjunto es una compilación de datos obtenidos mediante sensoramiento remoto (PP , T_s , σ^0 y ET). Estos dos conjuntos de variables de entrada se usaron con el producto de HS de SMAP distribuido a una resolución espacial de 36 km.

5.1. Desagregación de HS-SMAP con PP , T_s , σ^0 y HR

En este apartado se presenta la calibración y la validación de los modelos MLR y MLP para el primer conjunto de datos propuesto.

5.1.1. Calibración método MLR

Al igual que en la Sección 4.2.1, las variables en la matriz de entrada se transformaron aplicando logaritmo a HS , HR y T_s antes de calibrar el modelo MLR. Las variables PP y σ^0 se incluyeron en la matriz sin ser transformadas.

La selección de las variables que mejor explican variabilidad de HS se realizó con el procedimiento de eliminación hacia atrás. El primer modelo de MLR creado con la matriz de trabajo compuesta por PP , $\log(T_s)$, $\log(HR)$ y σ^0 , produjo un R^2 de aproximadamente 0.57. El estadístico p-valor fue significativo para las variables y el FIV no mostró colinealidad entre las variables.

La matriz de trabajo final estaba compuesta por PP , $\log(T_s)$, σ^0 y $\log(HR)$, que se dividió en dos matrices: una para calibración y la otra para validación. La división de los datos se realizó sobre el número total de fechas (27 fechas), presentadas en la Tabla 3.2. El número de muestras óptimo para calibrar el modelo se obtiene por medio del análisis de la curva de aprendizaje usando el RMSE. El tamaño de datos óptimo se encontró para un tamaño de conjunto de calibración del 70 % de los datos. Para obtener la mayor representatividad en los dos conjuntos de datos, estos porcentajes se aplicaron al número total de fechas. Es por esto que se tomó aleatoriamente el 70% de las fechas (18 fechas) para la calibración y el 30 % restante (9 fechas) para la validación.

El conjunto de calibración, compuesto por 1501 muestras, se usó para calcular los coeficientes del MLR. El modelo obtenido en la fase de calibración tiene un R^2 igual a 0.64 y un RMSE de 7.07 %Vol. (Figura 5.5- a Círculos). Luego, se aplicó validación cruzada con 10 particiones para comprobar la robustez del modelo calibrado, obteniendo valores promedios de R^2 y RMSE iguales a 0.64 y 7.14 %Vol., respectivamente. La Tabla 5.2 resume los coeficientes del modelo, y los estadísticos p-valor y FIV. Se debe tener en cuenta que las variables seleccionadas siguen siendo significativas (p-valor inferiores a 0.05) y no son colineales (FIV inferior a 5).

Tabla 5.2: Estadísticas de las variables de entrada seleccionadas del primer conjunto de datos.

Variable	Coefficiente	p-valor	FIV
β_0	-2.45	0.0	
σ^0	0.0025	0.018	1.19
PP	0.0033	0.0	1.36
$\log(T_s)$	0.3	0.0	1.62
$\log(HR)$	1.86	0.0	1.31

Finalmente, el modelo MLR resultante está representado por la siguiente ecuación:

$$\widehat{HS} = -2.45 + 0.0025 \sigma^0 + 0.0033 PP + 0.3 \log(T_s) + 1.86 \log(HR) \quad (5.1)$$

El conjunto de datos de validación, compuesto por 513 datos, se usó para verificar la capacidad de estimación del modelo. La comparación entre los valores estimados y HS de SMAP arrojó un R^2 igual a 0.35 y un RMSE de 7.87 %Vol. (Figura 5.5-b Círculos).

En los diagramas de dispersión para ambas fases, Figura 5.5-Círculos, se puede observar que la HS estimada con MLR posee valores muy dispersos para HS superior a 30 %Vol., en particular en la fase de calibración (Figura 5.5-a), los valores de HS-SMAP próximos a 30 %Vol. poseen valores estimados superiores al 50 %Vol. Estos valores de HS anómalos pueden estar relacionados con valores erróneos en las variables de entrada. Además, las relaciones no son biunívocas, el rango de valores de HS-SMAP se relaciona con un único valor de HS estimada, por ejemplo en la Figura 5.5-a el valor estimado de HS igual a 18 %Vol. se relaciona con valores HS-SMAP en el rango 10 y 20 %Vol. Este efecto de escalonamiento de los valores estimados es evidente en la calibración y validación, y está relacionado con los patrones de comportamiento de las variables de entrada (HR , PP y T_s), mencionados cuando se realizó el análisis de la matriz de correlación (Figura 5.4). Por otra parte, para valores de HS-SMAP mayores a 30 %Vol. se observa una bifurcación de los valores estimados en ambas fases, este patrón de comportamiento también se observó en el diagrama de dispersión entre HS-SMAP y PP. Dicha bifurcación es más evidente en la Figura 5.5-a.

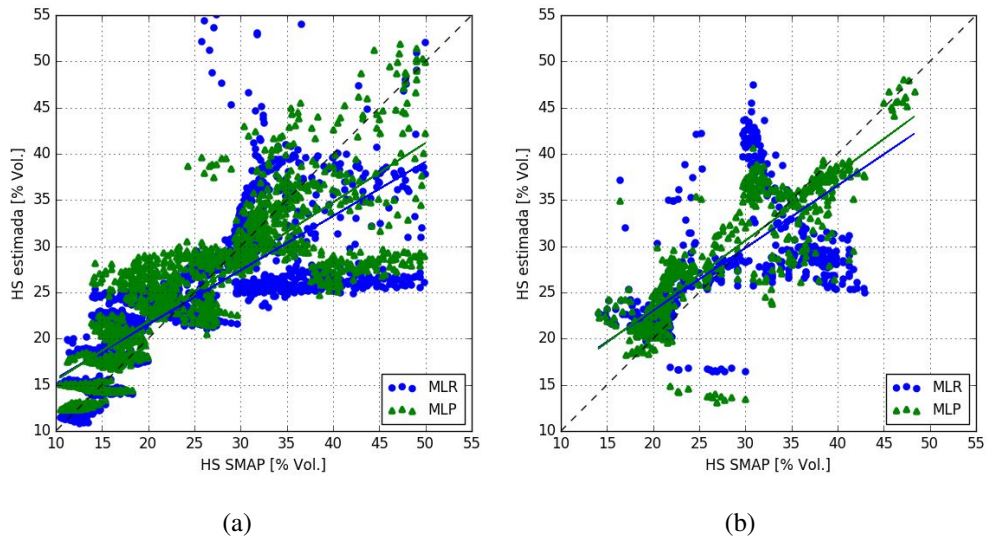


Figura 5.5: Gráficos de dispersión entre HS de SMAP y HS estimada utilizando HR con los métodos MLR (Círculos) y MLP (Triángulos) durante las fases de: (a) calibración y (b) validación.

5.1.2. Calibración del método MLP

Para este método, al igual que el método MLR, el conjunto de datos se dividió aleatoriamente en subconjuntos de calibración (70 % de las fechas) y validación (30 % de las fechas). Las entradas disponibles se escalan entre $[-1,1]$ para lograr un correcto entrenamiento de la red MLP, al igual que se realizó en la Sección [4.2.1](#).

Como se explicó anteriormente, la estructura del MLP se obtiene con búsqueda exhaustiva sobre el espacio de parámetros. La arquitectura MLP resultante tiene cuatro capas, es decir, una capa de entrada con cuatro neuronas (PP , T_s , HR y σ^0); dos capas ocultas con cinco y cuatro neuronas, respectivamente, y una capa de salida con una neurona. El método de optimización de pesos que arrojó mejores resultados fue GD junto con la función de transferencia ReLU para las capas ocultas y el tipo de tasa de aprendizaje adaptativo. Los R^2 y RMSE para esta configuración fueron 0.66 y 4.68 %Vol., respectivamente. Por lo tanto, la arquitectura del modelo MLP adoptada tiene una configuración de 4-5-4-1 neuronas.

El modelo MLP derivado fue calibrado con un R^2 igual a 0.66 y un RMSE de 5.82 %Vol. (Figura [5.5](#) -a Triángulos). En la Tabla [5.3](#) se presentan los pesos sinápticos de la arquitectura MLP calibrada. En la etapa de validación el modelo alcanzó un R^2 igual a 0.68 y un RMSE de 4.68 %Vol. (Figura [5.5](#) - b Triángulos), respectivamente.

Tabla 5.3: Pesos sinápticos (W_{ij}) para el método MLP con arquitectura 4-5-4-1.

Variable de entrada	W_{i1}	W_{i2}	W_{i3}	W_{i4}	W_{i5}
W_{ij} entre las entradas y la 1er capa oculta					
HR	-1.1	0.49	-0.53	3.34	0.76
PP	2.64	0.46	0.60	-0.76	-0.83
σ^0	-0.05	0.28	0.08	0.17	0.028
T_s	-1.92	-0.076	1.19	0.35	0.95
W_{ij} entre la 1er y la 2da capa oculta					
W_{1j}	3.85	2.85	5.33	3.77	
W_{2j}	-0.0008	-0.2	-1.6e-14	0.1	
W_{2j}	2.15	0.7	1.8	-1.7	
W_{3j}	1.2	1.3	1.9	-0.6	
W_{4j}	1.13	1.23	1.2	-1.95	
W_k entre la capa oculta y la de salida					
W_1	1.33				
W_2	1.51				
W_3	1.3				
W_4	-3.4				

En contraste con el método MLR, MLP reduce el RMSE en 1.25 %Vol. y 3.19 %Vol. en las fases de calibración y validación, respectivamente. Esto se puede observar en los gráficos de dispersión, donde MLP no presenta valores anómalos en la calibración para valores de HS-SMAP próximos a 30 %Vol. (Figura 5.5-a). Por otro lado, en la validación los valores de HS estimada con MLP poseen menor dispersión en comparación con el método MLR (Figura 5.5-b). Sin embargo, también están presentes los patrones de comportamiento mencionados para el método MLR, el escalonamiento de los valores de HS estimada y la bifurcación de los valores estimados para HS-SMAP mayor a 30 %Vol.

5.1.3. Contraste HS estimada versus HS de otras fuentes

Una vez que se construyeron los modelos, se obtuvieron mapas de HS con alta resolución espacial. Para esto, se aplicaron los modelos a las imágenes de PP , T_s , σ^0 y HR . Los mapas creados son 9 y se corresponden con las fechas del bloque de va-

lidación. Las fechas utilizadas son: 04/08/2015, 05/26/2015, 08/17/2015, 09/09/2015, 11/10/2015, 12/28/2015, 01/27/2016, 03/09/2016 y 03/19/2016.

Una vez obtenidos los mapas, se realizó una comparación espacial entre las estimaciones de HS y las mediciones *in situ* para analizar las correlaciones espaciales de los mapas desagregados. La comparación de las estimaciones de HS con observaciones puntuales implica la suposición de que las mediciones son válidas dentro de un radio de 5 km, lo que no es necesariamente cierto, sin embargo, este tipo de validación se ha realizado en teledetección desde hace tres décadas (Venturini et al., 2008; Venturini et al., 2011; Sánchez et al., 2014).

Para realizar este análisis se extrajeron de los mapas los puntos pertenecientes a las estaciones de CONAE. La comparación inicial entre la HS de SMAP original y HS observada arrojó un RMSE igual a 10.4 %Vol. y un sesgo = 3.4 %Vol. En la Figura 5.6 se presenta un diagrama de la evolución temporal de PP-GPM y mediciones de HS, junto a la diferencia puntual entre los valores de HS obtenidos re-escalando las imágenes de HS-SMAP a 5 km mediante los métodos MLR y MLP y los valores observados. El método MLR produce un RMSE igual a 9.26 %Vol. y un sesgo = 4.04 %Vol. (Figura 5.6 - Línea de trazo y punto Azul), mientras que con el método MLP el RMSE es igual a 11 %Vol. y sesgo = 2.25 %Vol. (Figura 5.6 - Línea de trazo Verde).

La Figura 5.6 evidencia que las diferencias entre la HS observada y la estimada por ambos métodos tienen un comportamiento similar, excepto para los valores de error entorno a la fecha 09/09/2015, donde MLR toma valores positivos con un pico de 25 %Vol., es decir que el método sobreestima la HS, y MLP posee valores negativos con un pico en 7 %Vol. aproximadamente, es decir, subestima la HS. Por otro lado, entorno a la fecha 03/09/2016 se observa una mayor diferencia entre los errores de la HS estimada, si bien el comportamiento es similar, el error para el método MLP es mayor.

En general, las fechas de mayor error en los mapas a escala reducida se corresponden con valores máximos de HS observada. También se aprecia que entorno a la fecha 09/09/2015, HS-CONAE presenta dos picos de HS (45 %Vol. aprox.), pero PP-GPM no registra valores elevados de lluvia acumulada, estos eventos locales no capturados en la resolución espacial de 10 km de GPM podrían ser una de las causas de la elevada diferencia (20 %Vol. aprox.) entre HS observada y estimada.

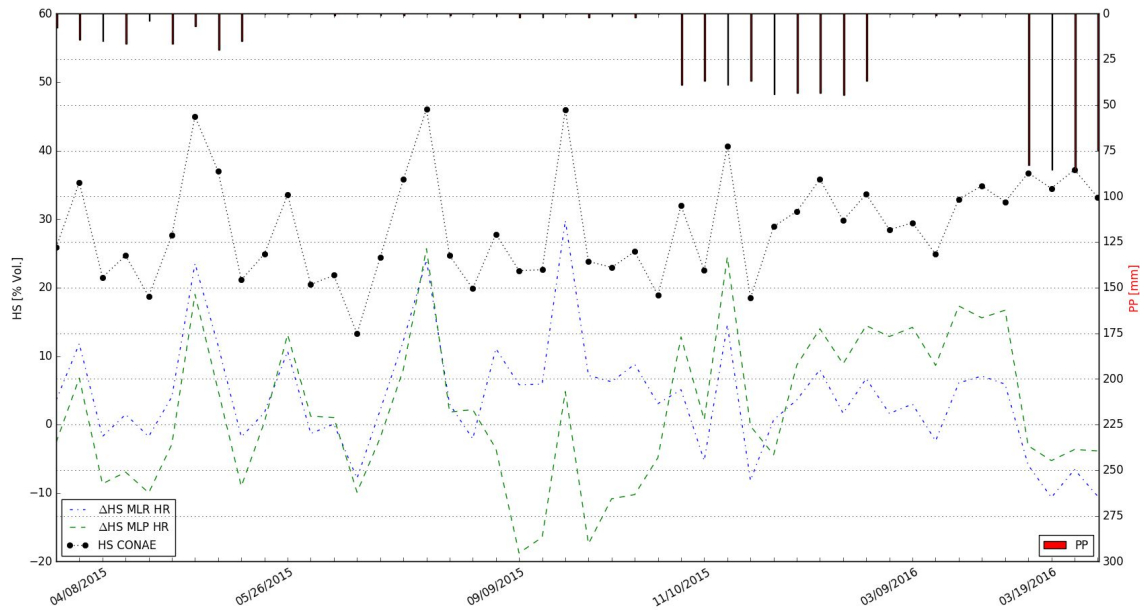


Figura 5.6: Evolución temporal de PP-GPM (Barras rojas) y HS-CONAE (Líneas de punto Negro). Las líneas de trazo de colores corresponden a la diferencia entre la HS observada y la estimada utilizando *HR* por los métodos MLR (Línea de trazo y punto Azul) y MLP (Línea de trazo Verde).

5.2. Desagregación de HS-SMAP con PP , T_s , σ^0 y ET

Los resultados de la sección anterior plantearon el desafío de obtener un modelo completamente satelital. Siguiendo la misma justificación que se describió en la Sección 2.1, las observaciones puntuales *HR* fueron reemplazadas por el producto *ET* que cuantifica la pérdida vertical de agua. Por lo tanto, se utilizó un segundo conjunto de datos que es una suma de productos satelitales. Las variables utilizadas fueron PP , T_s , σ^0 y ET . Con este conjunto de datos de entrada los modelos fueron calibrados y validados.

5.2.1. Calibración del método MLR

El modelo MLR con toda la matriz de entrada produce un R^2 de aproximadamente 0.5. Los estadísticos p-valor y FIV, sugirieron que todas las variables de entrada eran significativas y que no existe colinealidad entre las mismas. El modelo obtenido en la fase de calibración tiene R^2 igual a 0.5 y un RMSE de 7.88 %Vol. (Figura 5.7 - a Círculos). Nuevamente, se aplicó validación cruzada con 10 particiones para comprobar la robustez del modelo calibrado, obteniendo valores promedios de R^2 y RMSE

iguales a 0.5 y 7.93 %Vol., respectivamente.

La Tabla 5.4 resume los coeficientes del modelo y los estadísticos p-valor y FIV, donde se mantiene que todas las variables de entrada son estadísticamente significativas y no son colineales.

Tabla 5.4: Estadísticas de las variables de entrada seleccionadas del segundo conjunto de datos.

Variable	Coefficiente	p-valor	FIV
β_0	0.8	0.0	
σ^0	0.006	0.0	1.17
PP	0.0025	0.0	1.27
$\log(T_s)$	0.54	0.0	1.32
$\log(ET)$	-0.019	0.032	1

El modelo MLR resultante está representado por la siguiente ecuación:

$$\widehat{HS} = 0.8 + 0.006 \sigma^0 + 0.0025 PP + 0.54 \log(T_s) - 0.019 \log(ET) \quad (5.2)$$

La comparación entre los valores estimados y HS de SMAP arrojó un R^2 igual a 0.3 y RMSE de 6.92 %Vol. Las gráficas de dispersión para la validación del método MLR se representa en la Figura 5.7-b (círculos).

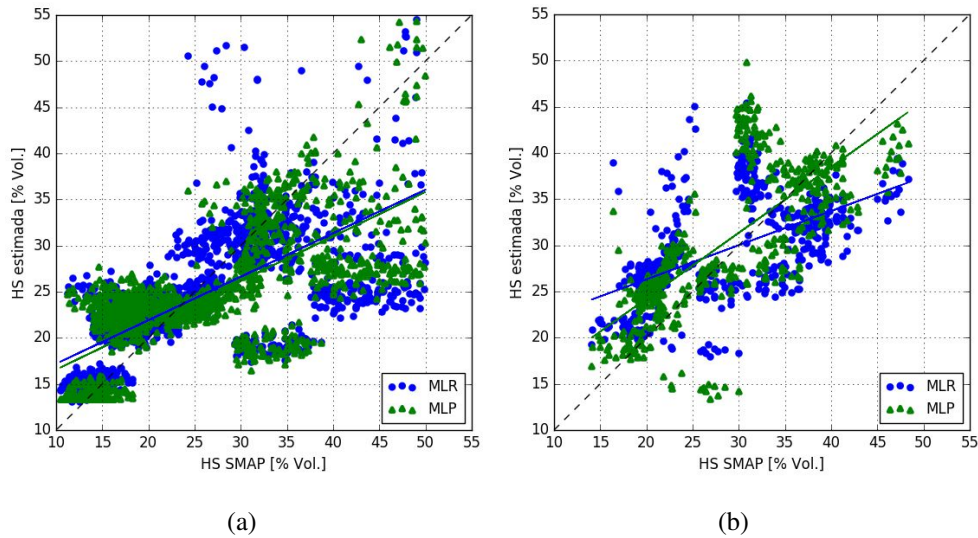


Figura 5.7: Gráficos de dispersión entre HS de SMAP y HS estimada utilizando ET con los métodos MLR (Círculos) y MLP (Triángulos) durante las fases de: (a) calibración y (b) validación.

En comparación con el caso presentado en la Sección [5.1](#), utilizando la variable ET los valores anómalos de HS estimada para HS-SMAP proximos a 30 %Vol. siguen presentes. A diferencia, el efecto bifurcación no es tan notorio como en el caso mencionado, esto puede deberse a que la variable *ET* compensa el comportamiento de *PP*. En el caso anterior, esto no sucedía ya que *HR* es una variable subrogante de *ET*.

5.2.2. Calibración método MLP

La arquitectura MLP obtenida tiene una red de cuatro capas que consiste en una capa de entrada (cuatro neuronas: σ^0 , T_s , *PP* y *ET*), dos capas ocultas (3 y 6 neuronas) y una capa de salida (HS). El R^2 y RMSE para esta configuración fueron 0.42 y 6.11 %Vol. Por lo tanto, la arquitectura del modelo MLP desarrollado tiene una configuración de 4-3-6-1 neuronas.

En la fase de calibración, el modelo alcanzó un R^2 igual a 0.51 y RMSE de 7.21 %Vol. (Figura [5.7](#) -a Triángulos). Los pesos sinápticos de la arquitectura MLP calibrada se presentan en la Tabla [5.5](#). La fase de validación se llevó a cabo con el 30 % restante de los datos y el modelo MLP tuvo un R^2 igual a 0.53 y el valor correspondiente de RMSE fue 6.11 %Vol. (Figura [5.7](#) - b Triángulos).

El método MLP reduce el RMSE en 0.67 %Vol. y 0.81 %Vol. en contraste con MLR para las fases de calibración y validación, respectivamente. En los gráficos de dispersión, Figura [5.7](#), se puede observar que el efecto de bifurcación para valores de

Tabla 5.5: Pesos sinápticos (W_{ij}) para el método MLP con arquitectura 4-3-6-1.

Variable de entrada	W_{i1}	W_{i2}	W_{i3}	W_{i4}	W_{i5}	W_{i6}
W_{ij} entre las entradas y la 1er capa oculta						
ET	0.007	0.02	-0.79			
PP	-0.56	1.45	0.21			
σ^0	0.21	0.23	0.94			
T_s	3.35	1.3	1.53			
W_{ij} entre la 1er y la 2da capa oculta						
W_{1j}	-0.51	-1.56	-0.64	-1.47	-1.35	0.015
W_{2j}	1.89	0.76	1.39	-0.19	1.81	-0.8
W_{3j}	-0.3	-0.28	1.05	-1.37	0.33	0.24
W_k entre la capa oculta y la de salida						
W_1	1.57					
W_2	2.24					
W_3	1.38					
W_4	0.48					
W_5	1.83					
W_6	-0.22					

HS-SMAP mayor a 30 %Vol. también se ve reducido en contraste con el caso anterior, utilizando HR.

5.2.3. Contraste HS estimada versus HS de otras fuentes

Los resultados de la diferencia entre la HS de SMAP re-escalada contra las observaciones de HS puntuales se muestran en la Figura 5.8. El modelo MLR arrojó un RMSE = 8.5 %Vol. (Figura 5.8- Línea de trazo y punto Azul) y sesgo = 1.1 %Vol. En contraste, el RMSE para MLP fue 11.75 %Vol. (Figura 5.8- Línea de trazo Verde) y el sesgo de -1.65 %Vol. El comportamiento de los errores, al igual que el caso anterior, es muy similar para ambos métodos, excepto para la fecha 09/09/2015, donde nuevamente MLR sobreestima y MLP subestima la HS. En este caso, también se observa un mayor error para MLP entorno a la fecha 03/09/2016.

Finalmente, la HS desagregada con MLR tiene menor error con las observaciones puntuales que la HS desagregada con el método MLP. Estos resultados estadísticos sugieren que el modelo con variables detectadas remotamente funciona tan bien como con variables de diferentes fuentes. Más aún, para el método MLR el RMSE es menor con ET que con HR.

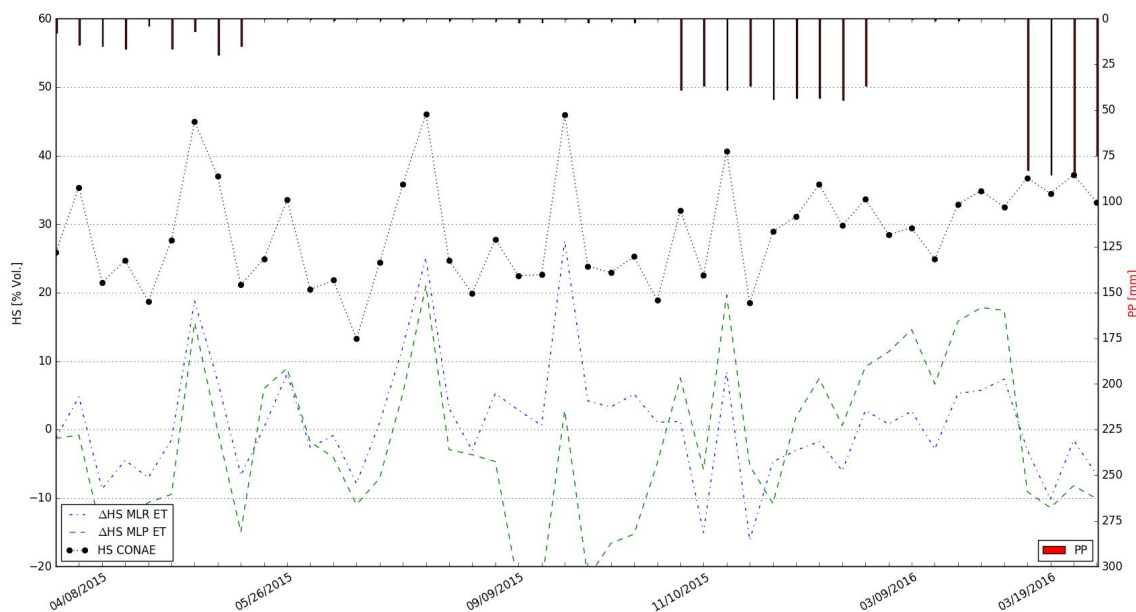


Figura 5.8: Evolución temporal de PP-GPM (Barras rojas) y HS-CONAE (Líneas de punto Negro). Las líneas de colores corresponden a la diferencia entre la HS observada y la estimada utilizando *ET* por los métodos MLR (Línea de trazo y punto Azul) y MLP (Línea de trazo Verde).

Para resumir los resultados y mostrar la capacidad de los modelos MLR y MLP en el re-escalado de los mapas de HS de SMAP se utilizó el diagrama de Taylor (Figura 5.9). Este diagrama resume los valores de S (eje x y eje y) y r (eje radial) obtenidos de la comparación entre los resultados de los modelos y las observaciones *in situ*. El símbolo con forma de estrella en el eje x , llamado punto de referencia, representa los estadísticos de los datos *in situ*. Además, cuanto más cerca esté un resultado al punto de referencia, mejor será el rendimiento del modelo.

El r y S entre la HS de SMAP y la HS *in situ*, fueron 0.22 y 8.15 %Vol., respectivamente (cuadrado verde en la figura 5.9). La desagregación con *HR* alcanzó $r = 0.35$ y $S = 7$ %Vol. para el método MLR (círculo azul en la Figura 5.9) y $r = 0.06$ y $S = 8.26$ %Vol. para el método MLP (triángulo rojo hacia arriba en la Figura 5.9). El modelo que usa *ET* da como resultado r y S igual a 0.3 y 6.5 %Vol., respectivamente para MLR (triángulo gris hacia abajo en la figura 5.9), y $r = 0.07$ y $S = 9.46$ %Vol. para MLP (rombo rosa en la Figura 5.9). Este diagrama resaltó el desempeño del método MLR, ya sea con *HR* o *ET*, como el más apropiado para la desagregación de los mapas de HS de SMAP.

La Tabla 5.6 resume los resultados obtenidos hasta el momento en términos del

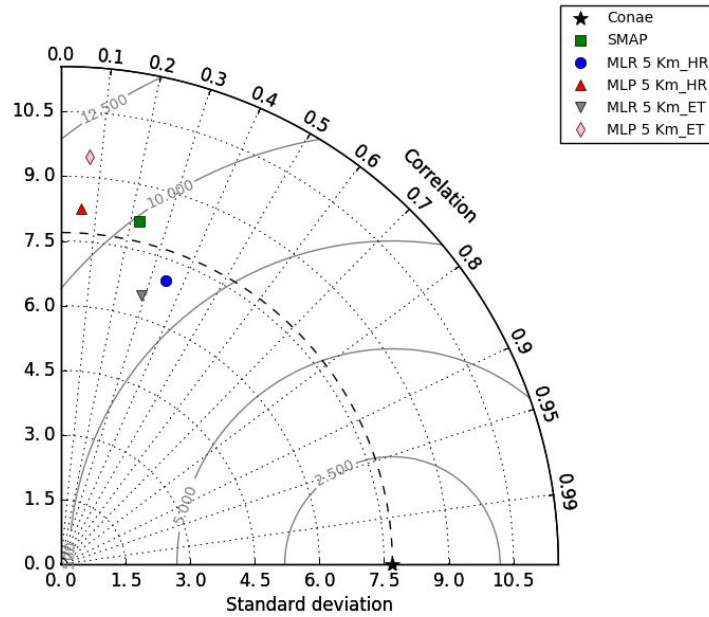


Figura 5.9: Diagrama de Taylor para el desempeño de las metodologías. SMAP (cuadrado); CONAE (estrella); MLR con HR (círculo); MLP con HR (triángulo hacia arriba); MLP con ET (triángulo hacia abajo); MLP con ET (diamante).

RMSE. La tabla permite visualizar el desempeño de los modelos y las alternativas propuestas. MLR presenta un error de aproximadamente 7-8 %Vol. en todas las desagregaciones realizadas, y de aproximadamente 9 %Vol. en el contraste con datos observados. Si bien estos errores representan el 16 % y 20 % del error máximo, respectivamente, disminuyen el error intrínseco de SMAP observado en la zona de estudio. MLP presenta errores que varían entre 5 y 7 %Vol. para la desagregación realizada con los distintos datos, y arrojó un error significativamente más alto en el contraste con los datos observados. Este comportamiento de MLP pone en duda la robustez de la metodología al aplicarla a datos no utilizados en la calibración. Es decir, MLP logra una buena desagregación de SMAP para el conjunto de datos para el que se lo calibró pero no mantiene el desempeño con datos nuevos. Por el contrario, MLR si bien tiene errores más altos que MLP, su desempeño es consistente en todas las desagregaciones realizadas. Además, el modelo MLR resultante es simple y fácil de analizar en términos de errores como se realizó en el Capítulo 4, inciso 4.2.4. Esto permite suponer que mejores fuentes de datos redundarán en mejores resultados para ambos métodos, pero no se puede asegurar que MLP se convierta en un modelo más robusto como MLR.

Tabla 5.6: RMSE de los modelos MLR y MLP con *HR* y *ET* durante las fases de calibración, validación y prueba.

RMSE	Calibración		Validación		Prueba	
	HR	ET	HR	ET	HR	ET
MLR	7	7.88	7.87	6.92	9.26	8.5
MLP	5.8	7.21	4.7	6.11	11.07	11.75

Como se mencionó anteriormente, el RMSE entre los datos de HS-SMAP y observados es igual a 10.4 %Vol., con un sesgo igual a 3.4 %Vol., por lo que el método MLR reduce el error entorno al 11 % y 19 %, según se utilice *HR* o *ET*, respectivamente. Estos resultados son comparables con los encontrados en la literatura (Piles et al., 2011, 2014). Piles et al. (2014) proponen una técnica de desagregación que combina datos de SMOS y MODIS para obtener mapas de HS a resolución espacial fina. Estos autores encontraron que su modelo mejora el RMSE de HS de SMOS entre 13 % y 30 %, para el producto MODIS nodo ascendente o descendente, respectivamente.

Si bien el modelo propuesto por Piles et al. (2014) obtiene mejores resultados, sólo puede ser mejorado cambiando el producto T_s de MODIS por uno con mayor precisión. Por otra parte, los modelos propuestos en esta tesis pueden ser mejorados introduciendo mejores productos para T_s y/o *PP* y/o *ET*, por lo que ofrece una gama más amplia de alternativas para reducir el RMSE.

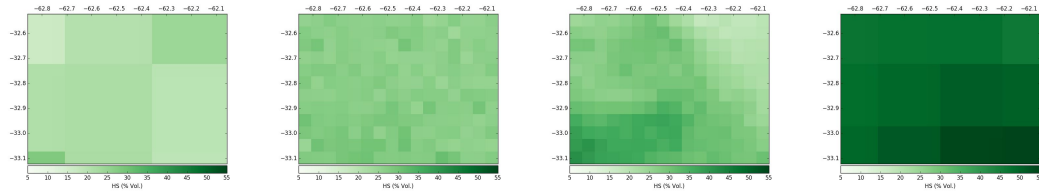
5.2.4. Análisis de la variabilidad regional

En esta sección se analizan los mapas de HS desagregados resultantes de la aplicación de los métodos MLR y MLP con la variable *ET*. La Figura 5.10 compila las imágenes de HS así obtenidas para las fechas 04/08/2015, 26/05/2015, 16/08/2015, 09/09/2015 y 10/11/2015. La primera columna muestra la HS original de SMAP a una resolución espacial de 36 km, la segunda y la tercera columnas muestran HS desagregada con los métodos MLR y MLP, respectivamente. La cuarta columna muestra HS estimada con el modelo VIC a 25 km de resolución. En general, la HS estimada, expresada en %Vol., varía entre 12 y 35 para SMAP, entre 13 y 30 para el método MLR y entre 17 y 55 para el método MLP.

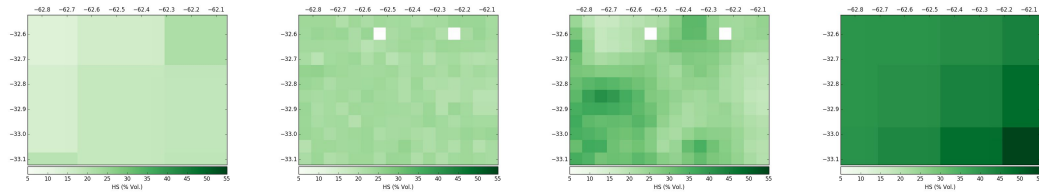
Una comparación visual indica que la distribución espacial general de HS de SMAP a 36 km y HS desagregada con MLR guardan similitud, y ambos difieren de los mapas obtenidos con MLP. Aunque el rango de valores de los mapas de HS de VIC

difieren de los de HS de SMAP, la distribución espacial general de ambas fuentes de HS coincide. Los mapas obtenidos con el método MLR proporcionaron una variabilidad espacial más suave que los del método MLP. Tanto MLR como MLP reproducen el rango de variabilidad de HS de SMAP. Se debe notar que el gradiente espacial de HS de MLP es diferente al de los otros mapas para las fechas 16/08/2015, 09/09/2015 y 10/11/2015. Este patrón es causado por el hecho de que T_s es la variable de mayor peso en MLR, mientras que PP parece ser la característica más importante en el modelo MLP. La inspección visual de las imágenes de T_s y PP para estas fechas, confirmó que MLP copió la distribución espacial de PP. El hecho de que el modelo de regresión represente mejor la variabilidad espacial de HS podría estar relacionado con las funciones logarítmicas introducidas en las variables de entrada al método, que parece ser más apropiado que la función polinómica no lineal encontrada por el método MLP.

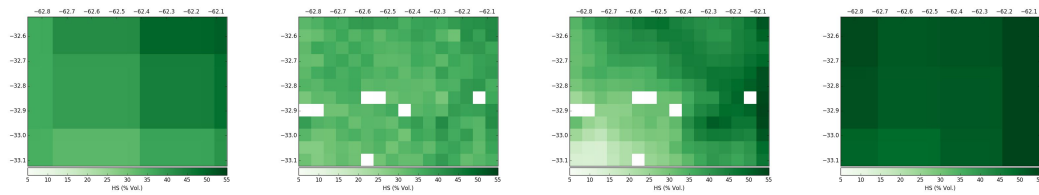
04/08/2015



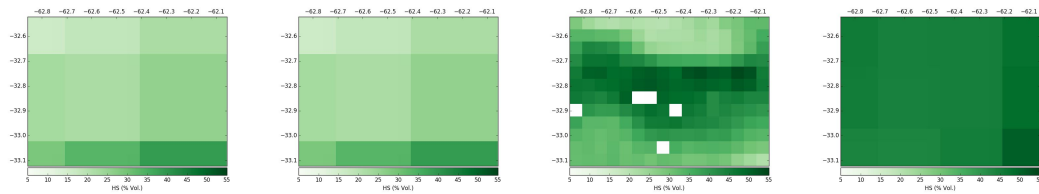
05/26/2015



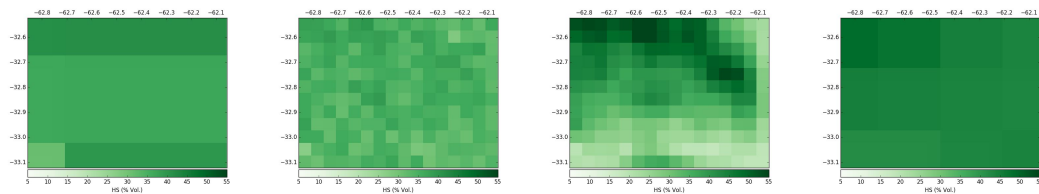
08/16/2015



09/09/2015



11/10/2015



(a)

(b)

(c)

(d)

Figura 5.10: HS para la localidad de Monte Buey (Córdoba), comparación entre (a) SMAP, desagregada utilizando *ET* (b) con MLR, (c) con MLP, (d) resultado del modelo VIC.

En la Tabla 5.7 se presentan el RMSE y el SSIM entre los mapas obtenidos con los métodos MLR y MLP y los productos de la misión SMAP y del modelo VIC. Es decir, con el RMSE se cuantifican las diferencias entre un método y otro, no es un error en el sentido estricto de la definición de RMSE. Como era de esperar los menores RMSE se obtienen entre los mapas derivados con MLR y MLP y los productos de SMAP, particularmente, con el método MLR se logran los menores RMSE. Las diferencias encontradas en el contraste con VIC son muy superiores a los de SMAP, dentro de estos, los RMSE más bajos se alcanzan para la fecha 11/10/2015 y son 9.94 y 12.31 %Vol. para MLR y MLP, respectivamente. Estos resultados pueden atribuirse al hecho que la HS modelada con VIC varía entre 40 y 59 %Vol., siendo el límite inferior mayor a los obtenidos con los métodos MLR (13 %Vol.) y MLP (17 %Vol.), y al del producto SMAP (12 %Vol.).

El índice SSIM toma valores cercanos a cero en casi todas las comparaciones para todas las fechas. Excepto para algunas de las comparaciones en las fechas 04/08/2015, 08/16/2015 y 11/10/2015, donde toma valores próximos a 0.2. En general, no se encuentra similitud estructural en los resultados analizados.

Tabla 5.7: RMSE y SSIM entre los mapas obtenidos con MLR y MLP, y los mapas de SMAP y VIC.

Fecha	Mapa contraste	Método	RMSE (%Vol.)	SSIM
04/08/2015	SMAP	MLR	6.42	0.05
		MLP	9.46	0.04
	VIC	MLR	22.8	0.13
		MLP	21.38	0.24
05/26/2015	SMAP	MLR	6.64	-0.01
		MLP	11.36	-0.07
	VIC	MLR	21.5	0.018
		MLP	18.92	-0.15
08/16/2015	SMAP	MLR	10.14	0.12
		MLP	10.09	0.22
	VIC	MLR	20.62	0.004
		MLP	19.65	0.066
09/09/2015	SMAP	MLR	8.58	0.013
		MLP	17.05	-0.01
	VIC	MLR	26.67	0.035
		MLP	14.15	-0.006
11/10/2015	SMAP	MLR	4.8	-0.004
		MLP	10.06	0.12
	VIC	MLR	9.94	0.18
		MLP	12.31	0.17

Capítulo 6

Conclusiones y trabajos futuros

6.1. Conclusiones

En esta tesis se abordó el problema de la estimación de la HS bajo cobertura vegetal moderada en zonas de llanura mediante la fusión de datos de microondas pasivas y activas. En la literatura es común encontrar que la retrodispersión de la onda SAR está relacionada principalmente a propiedades del suelo, incluida la HS. El razonamiento detrás del enfoque propuesto en esta tesis es que la HS no sólo es el resultado de las propiedades del suelo, sino que también responde a procesos más complejos como la infiltración, y la *ET*, entre otros. La metodología propuesta para estimar HS se basó en la ecuación de balance hídrico que considera la HS como el resultado de las entradas y salidas de agua al sistema suelo.

Las variables que se usaron para representar los movimientos del agua se obtuvieron de diferentes fuentes. Los datos meteorológicos disponibles de las estaciones terrestres operadas por INTA fueron T_a , e_a , *PP* y *HR*. Todos ellos representan un proceso en el balance hídrico, pero sólo *PP*, *HR* y T_a fueron estadísticamente significativas para la HS. Por otra parte, los productos satelitales también brindan información sobre variables como ser *PP*, HS, *ET* y T_s , entre otras. Las variables remotamente obtenidas se validaron con observaciones *in situ* para analizar los errores introducidos a los métodos seleccionados. La variable *PP* de GPM introduce un error igual al 19% del valor máximo de *PP*-GPM, el error introducido por la variable T_s de SMAP es igual a 0.75% del valor máximo de T_s -SMAP y la variable *ET* de MODIS introduce un error aproximado del 25%. Por lo tanto, es claro que la calidad de los datos de *PP*, T_s y *ET* puede afectar directamente la precisión de las estimaciones de HS resultantes.

La ecuación de balance hídrico se simplificó a partir de los métodos estadísticos

MLR y MLP. Cada uno de estos métodos presenta ventajas y desventajas. La principal ventaja del MLR es que permite aproximar la ecuación de balance hídrico con funciones físicamente significativas y con una cantidad razonable de datos. En oposición, MLP puede resolver cualquier tipo de no linealidad, aunque se requiere una gran cantidad de muestras. En cuanto a la cantidad de parámetros a calibrar, MLR posee un parámetro por cada variable de entrada, en cambio en el método MLP se deben calibrar múltiples parámetros como ser, cantidad de capas ocultas, número de neuronas por capa oculta, pesos, función de activación, y tasa de aprendizaje, entre otros. Ambos métodos son adecuados para simplificar la ecuación de balance no lineal propuesta en la metodología, sin embargo, las ecuaciones resultantes con MLP son más difíciles de analizar en términos de incertidumbre en comparación a las obtenidas con MLR. Es decir, la forma de la ecuación encontrada con MLR permite analizar de manera más sencilla el error aportado por cada variable de entrada al error total del modelo. Cabe aclarar que los errores encontrados en esta tesis tienen origen no sólo en la simplificación de la ecuación de balance, sino también en los errores propios de las fuentes de datos.

Se propusieron dos configuraciones para estimar HS. Bajo la primera configuración, la HS es estimada mediante las variables *in situ* PP , HR y T_a , y el dato satelital σ^0 . Ambos modelos arrojaron resultados similares, con errores de aproximadamente 10 % del valor medio observado de HS. El error obtenido por ambos métodos es comparable con los publicados por otros autores con diferentes metodologías, tales como modelos de inversión, modelos bayesianos y SVR. Los modelos propuestos se diferencian de los mencionados en la forma que se utiliza la retrodispersión de las microondas, ya que en esta tesis las mismas representan las propiedades del suelo. Si bien la retrodispersión está fuertemente asociada a las propiedades del suelo puede ocurrir que no sea suficiente para representar el medio suelo. Por lo que debería investigarse la incorporación de mapas de propiedades del suelo, como por ejemplo la textura, para representar mejor el medio físico en el que se aplica la ecuación de balance.

La segunda configuración permite desagregar los mapas de HS de la misión SMAP con resolución espacial de 36 km en mapas de HS con resolución de 5 km a partir de la ecuación de balance y los métodos MLR y MLP. En esta instancia, se compararon dos conjuntos de datos, en uno la salida del agua del sistema estaba representada por observaciones *in situ* de HR y en otro por el producto satelital ET provisto por MODIS. Los resultados mostraron que los modelos aplicados con HR arrojaron menores diferencias que cuando se los utilizó con ET . El análisis de los errores de MLR sugiere que los

errores en *HR* y *ET* tienen gran peso en el error total de HS. Las observaciones de *HR* podrían ser ligeramente mejoradas si se dispone de estaciones meteorológicas con sensores de mayor precisión que los que dispone INTA. Sin embargo, los productos de *ET* pueden ser reemplazados por otras fuentes de mejor calidad que la utilizadas aquí, lo que influiría positivamente en ambos modelos. En efecto, hoy en día se dispone de metodologías para estimar esta variable con errores de 10-15 % del valor medio. En cuanto a los modelos utilizados, el método MLR, ya sea con *HR* o *ET*, es el resultó el más apropiado para la desagregación de los mapas de HS de SMAP. El método MLP logró resultados correctos en la validación con SMAP, pero arrojó errores significativamente más altos que MLR en el contraste con datos observados *in situ*.

Entre las desventajas de los modelos propuestos se puede mencionar la cantidad de datos necesarios para calibrar el modelo MLP, debido a que su robustez depende del tamaño de la base de datos. Si bien MLR requirió un volumen de aproximadamente 1000 muestras para obtener una estructura de modelo robusta, MLP demandó mayor cantidad de muestras para el entrenamiento de todos los parámetros (pesos, velocidad de aprendizaje, momento y penalidad). Cabe señalar que la serie temporal de imágenes Sentinel-1 tiene tan sólo dos años de datos. Por lo tanto, se podría esperar que a medida que aumente el número de datos, la calibración y validación del modelo mejoren.

Se destaca que la banda C en polarización VV no es recomendada para la estimación de HS debido a su sensibilidad a la cobertura vegetal. Aun así, los datos de Sentinel-1 proporcionados gratuitamente por ESA representan una ventaja importante en comparación con otros productos costosos, con una polarización y/o frecuencia más adecuada para la estimación de HS. En efecto, esta tesis muestra que se puede aplicar una metodología simple con resultados aceptables con un costo prácticamente nulo.

En resumen, una de las ventajas de la metodología propuesta radica en el tipo de variables que podrían utilizarse dentro de la misma para estimar HS, es decir, variables meteorológicas, propiedades del suelo, variables hidrológicas y datos satelitales gratuitos. Otra ventaja destacable es que no requieren usuarios expertos en sensoramiento remoto y en informática; por ejemplo un modelo MLR sólo requiere conocimientos básicos de estadística. Estas particularidades serían relevantes para la generación automática de mapas, y en el caso de aplicaciones para organizaciones agrícolas del interior del país donde no se encuentran fácilmente especialistas en sensoramiento remoto o disciplinas como la física y otras.

Las metodologías propuestas en la literatura introducen los detalles de la variabilidad de HS a partir de información de temperatura, vegetación y parámetros de la

superficie. En esta tesis se propone refinar los productos de SMAP introduciendo la variabilidad espacial de la precipitación, la temperatura y una variable atmosférica por excelencia como *HR* o *ET*. Los resultados presentados muestran que el enfoque propuesto es adecuado y flexible en relación a la fuente de la información necesaria. De hecho, las propiedades del suelo no son estrictamente necesarias para ensamblar los modelos propuestos en esta tesis como si lo son en los modelos de inversión.

6.2. Trabajos Futuros

Teniendo en cuenta que a lo largo del desarrollo de la tesis surgieron determinadas inquietudes acerca de las metodologías utilizadas que, por los plazos de tiempo no pudieron ser abordadas, se plantean propuestas para continuar su abordaje en el futuro.

Por un lado, se plantea extender la aplicabilidad de la metodología desarrollada a regiones con pendiente pronunciada. Esta aplicación requeriría la incorporación de la topografía del área de estudio (es decir, modelos digitales de terreno), que no se puede incorporar como variable de entrada al modelo. Esto plantea un reto que hay que resolver. A priori una de las formas de representar la topografía es por medio de mapas de densidad de precipitación. De esta manera, se pretende extender el procedimiento a tipos de superficies más generales.

Por otro lado, en este trabajo se sugirió una combinación de datos de entrada disponibles hoy en día. Se debería investigar en forma más exhaustiva sobre nuevas fuentes de datos para lograr mapas de HS más precisos. Una alternativa podría ser la utilización de información de otras misiones dentro de Constelación Copernicus y del Programa SAOCOM. La utilización de imágenes SAR en banda L permitiría obtener estimaciones con mayor precisión, debido a la penetración más profunda del suelo, y a la mejor percepción sobre las áreas de vegetación. Además, debido a que el contenido de agua de la vegetación atenúa la señal del suelo, se plantea la necesidad de estudiar la incorporación de índices de vegetación radar, que caracterizan la retrodispersión del volumen de vegetación.

En la última etapa de la tesis se comenzó a trabajar con la técnica de regresión no paramétrica denominada Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991). MARS es una extensión del método de regresión lineal que resuelve las no linealidades y la dependencia entre conjuntos de datos. Los primeros resultados obtenidos, Anexo E, demuestran su capacidad para capturar la no linealidad de la ecuación de balance de agua. Es por esto que se considera necesario profundizar la investigación

en la aplicabilidad de MARS para la estimación de HS.

Por último, en esta tesis se trabajó en la incorporación de detalles de la variabilidad espacial de HS a los mapas de SMAP reduciendo su resolución espacial a aproximadamente 5 km. Los resultados obtenidos hasta el momento proveen mapas refinados de HS con una frecuencia temporal provista por el tiempo de revisita de los satélites SAR. Cabe mencionar que la resolución temporal de los radiómetros es superior a los SAR. Por lo tanto, una propuesta de trabajo a futuro sería la integración de la variabilidad espacio-temporal en un mismo producto de HS. Este producto sería de gran valor para la asimilación de la variable en diferentes aplicaciones hidro-ambientales, como ser modelos de la superficie terrestre en tiempo real de alta resolución para la llanura pampeana, vulnerable a los impactos de las sequías e inundaciones.

Bibliografía

- Ahmad, S., A. Kalra, and H. Stephen. (2010). Estimating soil moisture using remote sensing data: A machine learning approach. *Advances in Water Resources*, 33(1):69–80.
- Al-Shrafany, D., M. A. Rico-Ramirez, D. Han, and M. Bray. (2014). Comparative assessment of soil moisture estimation from land surface model and satellite remote sensing based on catchment water balance. *Meteorological Applications*, 21(3):521–534.
- Al-Yaari, A., J. P. Wigneron, Y. Kerr, R. de Jeu, N. Rodriguez-Fernandez, R. van der Schalie, A. Al Bitar, A. Mialon, P. Richaume, A. Dolman, and A. Ducharne. (2016). Testing regression equations to derive long-term global soil moisture datasets from passive microwave observations. *Remote Sensing of Environment*, 180:453–464.
- Al-Yaari, A., J. P. Wigneron, Y. Kerr, N. Rodriguez-Fernandez, P. E. O'Neill, T. J. Jackson, G. J. M. De Lannoy, A. Al Bitar, A. Mialon, P. Richaume, J. P. Walker, A. Mahmoodi, and S. Yueh. (2017). Evaluating soil moisture retrievals from ESA's SMOS and NASA's SMAP brightness temperature datasets. *Remote Sensing of Environment*, 193:257–273.
- Albergaria, J. T., F. G. Martins, and C. Alvim-Ferraz, M. C. M. and Delerue-Matos. (2014). Multiple linear regression and artificial neural networks to predict time and efficiency of soil vapor extraction. *Water, Air, and Soil Pollution*, 225(8).
- Alexakis, D. D., F. D. K. Mexis, A. E. K. Vozinaki, I. N. Daliakopoulos, and I. K. Tsanis. (2017). Soil moisture content estimation based on Sentinel-1 and auxiliary earth observation products. A hydrological approach. *Sensors (Switzerland)*, 17(6):1–16.
- Ali, I., F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola. (2015). Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sensing*, 7(12):16398–16421.

- Alley, W. M.. (1984). On the treatment of evapotranspiration, soil moisture accounting, and aquifer recharge in monthly water balance models. *Water Resources Research*, 20(8):1137–1149.
- Baghdadi, N., M. Aubert, and M. Zribi. (2012). Use of TerraSAR-X data to retrieve soil moisture over bare soil agricultural fields. *IEEE Geoscience and Remote Sensing Letters*, 9(3):512–516.
- Baghdadi, N., N. Holah, and M. Zribi. (2006). Soil moisture estimation using multi-incidence and multi-polarization ASAR data. *International Journal of Remote Sensing*, 27(10):1907–1920.
- Bai, X., B. He, X. Li, J. Zeng, X. Wang, Z. Wang, Y. Zeng, and Z. Su. (2017). First assessment of Sentinel-1A data for surface soil moisture estimations using a coupled water cloud model and advanced integral equation model over the Tibetan Plateau. *Remote Sensing*, 9(7):1–20.
- Barrett, B. W., E. Dwyer, and P. Whelan. (2009). Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sensing*, 1(3):210–242.
- Bedano, J. C., A. Domínguez, R. Arolfo, and L. G. Wall. (2016). Effect of good agricultural practices under no-till on litter and soil invertebrates in areas with different soil types. *Soil and Tillage Research*, 158:100–109.
- Behari, J.. (2005). *Microwave Dielectric Behavior of Wet Soils*, volume 8 of *Remote Sensing and Digital Image Processing*. Dordrecht: Springer Netherlands.
- Belloni, M. C., M. DÍndio, R. O. Rodríguez, N. R. Fernández, A. F. Moltoni, and A. D. Blasón. (2011). Desarrollo de un sistema de observación y análisis climático y ambiental. Diseño de estaciones agrometeorológicas automáticas NIMBUS THP. *Rumbos Tecnológicos*, 3:95–105.
- Bergstra, J. and Y. Bengio. (2012). Random Search for Hyper-Parameter Optimization James. *Journal of Machine Learning Research*, 13:281–305.
- Bertoldi, G., S. Della Chiesa, C. Notarnicola, L. Pasolli, G. Niedrist, and U. Tappeiner. (2014). Estimation of soil moisture patterns in mountain grasslands by means of SAR RADARSAT2 images and hydrological modeling. *Journal of Hydrology*, 516:245–257.

- Beyene, T., D. Lettenmaier, and P. Kabat. (2010). Hydrologic impacts of climate change on the Nile River Basin: Implications of the 2007 IPCC scenarios. *Climatic Change*, 100(3):433–461.
- Bisquert, M., J. M. Sánchez, and V. Caselles. (2016). Evaluation of Disaggregation Methods for Downscaling MODIS Land Surface Temperature to Landsat Spatial Resolution in Barrax Test Site. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(4):1430–1438.
- Brocca, L., S. Camici, F. Melone, T. Moramarco, J. Martínez-Fernández, J. F. Didon-Lescot, and R. Morbidelli. (2014). Improving the representation of soil moisture by using a semi-analytical infiltration model. *Hydrological Processes*, 28(4):2103–2115.
- Brogioni, M., L. Rodriguez, S. Graciani, G. Fontanelli, M. C., and S. Paloscia. (2014). Aplicación de red neuronal artificial sobre imágenes sar para determinar la humedad del suelo. *2º Congreso Internacional de Hidrología de Llanuras, Santa Fe, Argentina, 23-26 Sept.*
- Brutsaert, W.. (2005). *Hydrology: An Introduction*, first edition. Cambridge University Press.
- Buck, A. L.. (1981). New equations for computing vapor pressure and enhancement factor. *Journal of Applied Meteorology*, 20(12):1527–1532.
- Chen, N., Y. He, and X. Zhang. (2017). NIR-red spectra-based disaggregation of SMAP soil moisture to 250 m resolution based on SMAPE_{x-4/5} in Southeastern Australia. *Remote Sensing*, 9(1).
- Chow, V., D. Maidment, and L. Mays. (1988). *Applied Hydrology*. McGraw-Hill Publishing Company.
- Cohen, J., P. Cohen, S. G. West, and L. Aiken. (2003). *Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences*, volume Third Edit.
- Courault, D., B. Seguin, and A. Olioso. (2005). Review on estimation of evapotranspiration from remote sensing data: From empirical to numerical modeling approaches. *Irrigation and Drainage Systems*, 19(3):223–249.

- Cybenko, G.. (1989). Approximation by superpositions of a sigmoidal function. *Approximation Theory and its Applications*, 9(3):17–28.
- Das, N. N., D. Entekhabi, S. Kim, S. Yueh, and P. O'Neill. (2016). Combining SMAP and Sentinel data for high-resolution Soil Moisture product. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016-November:129–131.
- Das, N. N., D. Entekhabi, and E. G. Njoku. (2011). An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 49(5):1504–1512.
- Das, N. N., D. Entekhabi, E. G. Njoku, J. J. C. Shi, J. T. Johnson, and A. Colliander. (2014). Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(4):2018–2028.
- Del Valle, H., P. Blanco, and W. Sione. (2011). Aplicaciones sustentables de las mediciones de humedad del suelo usando teledetección radar. In *VIII Jornadas de Educación en Percepción Remota en el Ámbito del Mercosur. SELPER, UNLu, CONAE, P. R. y. S. C. y. C. CD-ROM, Grupo 1: Enseñanza de la Currícula*, ed., Pp. 35–41.
- Dubois, P. C. and T. Engman. (1995). Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing*, 33(4):915–926.
- Duval, M. E., J. A. Galantini, J. M. Martínez, F. M. López, and L. G. Wall. (2016). Sensitivity of different soil quality indicators to assess sustainable land management: Influence of site features and seasonality. *Soil and Tillage Research*, 159:9–22.
- Engman, E. T.. (1991). Applications of microwave remote sensing of soil moisture for water resources and agriculture. *Remote Sensing of Environment*, 35(2-3):213–226.
- Engman, E. T. and N. Chauhan. (1995). Status of microwave soil moisture measurements with remote sensing. *Remote Sensing of Environment*, 51(1):189–198.
- Entekhabi, D., N. Das, S. Kim, T. Jagdhuber, M. Piles, S. H. Yueh, A. Colliander, E. Lopez-baeza, and J. Martínez-Fernández. (2016). High-resolution enhanced product based on SMAP active-passive approach and Sentinel 1A radar data. *AGU Fall Meeting Abstracts*.

- Entekhabi, D., E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. D. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. Van Zyl. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, 98(5):704–716.
- Fox, J. and G. Monette. (2002). *R and S-Plus companion to applied regression*. Sage Publications, Inc. Thousand Oaks, CA, USA ©2002.
- Friedman, J. H.. (1991). Multivariate adaptive regression splines. *The Annals of Statistics. Institute of Mathematical Statistics*, 19(1):1–67.
- Fung, A.. (1994). *Microwave Scattering and Emission Models and Their Applications*. Artech House Inc., Norwood, MA.
- Gherboudj, I., R. Magagi, A. A. Berg, and B. Toth. (2011). Soil moisture retrieval over agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data. *Remote Sensing of Environment*, 115(1):33–43.
- Grayson, R. B., A. W. Western, F. H. S. Chiew, and G. Blöschl. (1997). Preferred states in spatial soil moisture patterns: Local and nonlocal controls. *Water Resources Research*, 33(12):2897.
- Hamlet, A. F. and D. P. Lettenmaier. (1999). Effects of climate change on hydrology and water resources in the columbia river basin. *Journal of the American water resources association. American water resources association*, 35(6):1597–1623.
- Harrell, F.. (2001). *With Applications to Linear Models, Logistic Regression, and Survival Analysis*, volume First Edit.
- Haykin, S.. (2009). *Neural Networks and Learning Machines*, Neural networks and learning machines. Prentice Hall.
- Helsel, D. R. and R. M. Hirsch. (2002). Statistical Methods in Water Resources. *Technometrics*, 36(3):323.
- Holah, N., N. Baghdadi, M. Zribi, A. Bruand, and C. King. (2005). Potential of ASAR/ENVISAT for the characterization of soil surface parameters over bare agricultural fields. *Remote Sensing of Environment*, 96(1):78–86.

- Holben, B. N.. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11):1417–1434.
- Hornik, K.. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257.
- Hornik, K., M. Stinchcombe, and H. White. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- Inamdar, A. K. and A. French. (2009). Disaggregation of GOES land surface temperatures using surface emissivity. *Geophysical Research Letters*, 36(2):1–5.
- Jackson, T. J., D. M. Le Vine, A. Y. Hsu, A. Oldak, P. J. Starks, C. T. Swift, J. D. Isham, and M. Haken. (1999). Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains Hydrology Experiment. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5):2136–2151.
- Jackson, T. J. and T. J. Schmugge. (1991). Vegetation effects on the microwave emission of soils. *Remote Sensing of Environment*, 36(3):203–212.
- Joseph, A. T., R. V. D. Velde, P. E. O. Neill, S. Member, R. H. Lang, and T. Gish. (2008). Soil moisture retrieval during a corn growth cycle. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2365–2374.
- Kalma, J. D., T. R. McVicar, and M. F. McCabe. (2008). Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surveys in Geophysics*, 29(4-5):421–469.
- Kerr, Y. H. and E. G. Njoku. (1990). A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Transactions on Geoscience and Remote Sensing*, 28(3):384–393.
- Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M.-J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg. (2010). The SMOS Mission: New Tool for Monitoring Key Elements of the Global Water Cycle. *Proceedings of the IEEE*, 98(5):666–687.

- Kerr, Y. H., P. Waldteufel, J. P. Wigneron, J. M. Martinuzzi, J. Font, and M. Berger. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8):1729–1735.
- Kim, Y. and J. J. van Zyl. (2009). A time-series approach to estimate soil moisture using polarimetric radar data. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(8):2519–2527.
- Kingma, D. P. and J. Ba. (2014). Adam: A Method for Stochastic Optimization. In *the 3rd International Conference for Learning Representations, San Diego, 2015*, Pp. 1–15.
- Kornelsen, K. C. and P. Coulibaly. (2013). Advances in soil moisture retrieval from synthetic aperture radar and hydrological applications. *Journal of Hydrology*, 476:460–489.
- Krepper, C. M. and V. Venturini. (2009). Assessing interannual water balance of La Plata river basin. *Atmosfera*, 22(4):387–398.
- Kseneman, M., D. Gleich, and B. Potočnik. (2012). Soil-moisture estimation from TerraSAR-X data using neural networks. *Machine Vision and Applications*, 23(5):937–952.
- Le Hegarat-Masclé, S., M. Zribi, F. Alem, A. Weisse, and C. Loumagne. (2002). Soil moisture estimation from ERS/SAR data: toward an operational methodology. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2647–2658.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, 99(D7):14415–14428.
- Lievens, H., R. H. Reichle, Q. Liu, G. J. M. De Lannoy, R. S. Dunbar, S. B. Kim, N. N. Das, M. Cosh, J. P. Walker, and W. Wagner. (2017). Joint Sentinel-1 and SMAP data assimilation to improve soil moisture estimates. *Geophysical Research Letters*, 44(12):6145–6153.
- Lievens, H. and N. E. C. Verhoest. (2012). Spatial and temporal soil moisture estimation from RADARSAT-2 imagery over Flevoland, The Netherlands. *Journal of Hydrology*, 456-457:44–56.

- Lillesand, T. and R. Kiefer. (1994). *Remote Sensing and Image Interpretation*. John Wiley Sons, New York; Third edition.
- Lin, X., W. Zhang, Y. Huang, W. Sun, P. Han, L. Yu, and F. Sun. (2016). Empirical estimation of near-surface air temperature in China from MODIS LST data by considering physiographic features. *Remote Sensing*, 8(8):1–15.
- Liu, D., A. K. Mishra, and Z. Yu. (2016). Evaluating uncertainties in multi-layer soil moisture estimation with support vector machines and ensemble Kalman filtering. *Journal of Hydrology*, 538:243–255.
- Loew, A. and W. Mauser. (2006). A semiempirical surface backscattering model for bare soil surfaces based on a generalized power law spectrum approach. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4):1022–1035.
- Mallat, G.. (1989). A Theory for Multiresolution Signal Decomposition : The Wavelet Representation. II(7).
- Mallick, K., B. K. Bhattacharya, and N. K. Patel. (2009). Estimating volumetric surface moisture content for cropped soils using a soil wetness index based on surface temperature and NDVI. *Agricultural and Forest Meteorology*, 149(8):1327–1342.
- Marquardt, D. W.. (1970). Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12(3):591.
- Mattia, F., G. Satalino, L. Dente, and G. Pasquariello. (2006). Using a priori information to improve soil moisture retrieval from ENVISAT ASAR AP data in semiarid regions. *IEEE Transactions on Geoscience and Remote Sensing*, 44(4):900–911.
- McCulloch, W. S. and W. Pitts. (1943). A logical calculus of the ideas immanent in neurons activity. *Bull. Math. Biophys*, Pp. 115–133.
- Merlin, O., C. Rüdiger, A. Al Bitar, P. Richaume, J. P. Walker, and Y. H. Kerr. (2012). Disaggregation of SMOS soil moisture in Southeastern Australia. *IEEE Transactions on Geoscience and Remote Sensing*, 50(5 PART 1):1556–1571.
- Merlin, O., V. G. Stefan, A. Amazirh, A. Chanzy, E. Ceschia, T. Tallec, J. Beringer, P. Gentine, S. Er-Raki, S. Bircher, and S. Khabba. (2016). Modeling soil evaporation efficiency in a range of soil and atmospheric conditions: A downward approach based on multi-site data. *Submitted to Water Resources Research*, Pp. 3663–3684.

- Merlin, O., J. P. Walker, A. Chehbouni, and Y. Kerr. (2008). Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment*, 112(10):3935–3946.
- Merzouki, A., H. McNairn, and A. Pacheco. (2011). Evaluation of the Dubois, Oh, and IEM radar backscatter models over agricultural fields using C-band RADARSAT-2 SAR image data. 36(October 2009).
- Minsky, M. and S. Papert. (1969). *Perceptrons*. Oxford, England: M.I.T. Press.
- Molero, B., O. Merlin, Y. Malbêteau, A. Al Bitar, F. Cabot, V. Stefan, Y. Kerr, S. Bacon, M. H. Cosh, R. Bindlish, and T. J. Jackson. (2016). SMOS disaggregated soil moisture product at 1 km resolution: Processor overview and first validation results. *Remote Sensing of Environment*, 180:361–376.
- Montgomery, D. C. and E. A. Peck. (1992). *Introduction to Linear Regression Analysis*. Wiley-Interscience; Second edition.
- Montzka, C., T. Jagdhuber, R. Horn, H. R. Bogena, I. Hajnsek, A. Reigber, and H. Vereecken. (2016). Investigation of SMAP fusion algorithms with airborne active and passive L-Band microwave remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7):3878–3889.
- Moran, M. S., D. C. Hymer, J. Qi, and E. E. Sano. (2000). Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. *Agricultural and Forest Meteorology*, 105:69–80.
- Mu, Q., S. W. Running, H. A. Cleugh, and R. Leuning. (2007). Regional evaporation estimates from flux tower and MODIS satellite data. *Remote Sensing of Environment*, 106(3):285–304.
- Mu, Q., M. Zhao, and S. W. Running. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, 115(8):1781–1800.
- Narayan, U., V. Lakshmi, and T. J. Jackson. (2006). High-resolution change estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1545–1554.

- Narayan, U., V. Lakshmi, and E. G. Njoku. (2004). Retrieval of soil moisture from passive and active L/S band sensor (PALS) observations during the Soil Moisture Experiment in 2002 (SMEX02). *Remote Sensing of Environment*, 92(4):483–496.
- Njoku, E. G. and D. Entekhabi. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology*, 184(1-2):101–129.
- Njoku, E. G., T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem. (2003). Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2 PART 1):215–228.
- Njoku, E. G., W. J. Wilson, S. H. Yueh, S. J. Dinardo, F. K. Li, T. J. Jackson, V. Lakshmi, and J. Bolten. (2002). Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2659–2673.
- Notarnicola, C., M. Angiulli, and F. Posa. (2008). Sensed data : neural network approach versus bayesian method. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2):547–557.
- Oh, Y., K. Sarabandi, and F. T. Ulaby. (1992). An empirical model and an inversion technique for radar scattering from bare surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2):370–381.
- Olivera-Guerra, L., C. Mattar, O. Merlin, C. Durán-Alarcón, A. Santamaría-Artigas, and R. Fuster. (2017). An operational method for the disaggregation of land surface temperature to estimate actual evapotranspiration in the arid region of Chile. *ISPRS Journal of Photogrammetry and Remote Sensing*, 128:170–181.
- Orth, R., R. D. Koster, and S. I. Seneviratne. (2013). Inferring soil moisture memory from streamflow observations using a simple water balance model. *Journal of Hydrometeorology*, 14(6):1773–1790.
- Paloscia, S., G. Macelloni, and E. Santi. (2006). Soil moisture estimates from AMSR-E brightness temperatures by using a dual-frequency algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11):3135–3144.
- Paloscia, S., P. Pampaloni, S. Pettinato, and E. Santi. (2008). A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 46(10):3274–3284.

- Paloscia, S., P. Pampaloni, S. Pettinato, and E. Santi. (2010). Generation of soil moisture maps from ENVISAT/ASAR images in mountainous areas: a case study. *International Journal of Remote Sensing*, 31(9):2265–2276.
- Pandey, V. and P. K. Pandey. (2010). Spatial and temporal variability of soil moisture. *International Journal of Geosciences*, 1(August):87–98.
- Parton, W. J.. (1984). Predicting soil temperatures in a shortgrass steppe. *Soil Science*, 138(2):93–101.
- Pasolli, L., C. Notarnicola, L. Bruzzone, G. Bertoldi, S. Della Chiesa, V. Hell, G. Niedrist, U. Tappeiner, M. Zebisch, F. Del Frate, and G. Vaglio Laurin. (2011). Estimation of soil moisture in an Alpine catchment with RADARSAT2 images. *Applied and Environmental Soil Science*, 2011:1–12.
- Pathe, C., W. Wagner, S. Member, D. Sabel, M. Doubkova, and J. B. Basara. (2009). Using ENVISAT ASAR global mode data for surface soil moisture retrieval over. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):468–480.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Peng, J., A. Loew, O. Merlin, and N. E. Verhoest. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture. *Reviews of Geophysics*, Pp. 1–26.
- Penman, H. L.. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 193(1032):120–145.
- Pierdicca, N., L. Pulvirenti, and C. Bignami. (2010). Soil moisture estimation over vegetated terrains using multitemporal remote sensing data. *Remote Sensing of Environment*, 114(2):440–448.
- Piles, M., A. Camps, M. Vall-Llossera, I. Corbella, R. Panciera, C. Rudiger, Y. H. Kerr, and J. Walker. (2011). Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9):3156–3166.

- Piles, M., D. Entekhabi, and A. Camps. (2009). A change detection algorithm for retrieving high-resolution soil moisture from SMAP radar and radiometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, 47(12):4125–4131.
- Piles, M., N. Sánchez, M. Vall-Llossera, A. Camps, J. Martínez-Fernández, J. Martínez, and V. Gonzalez-Gambau. (2014). A downscaling approach for SMOS land observations: Evaluation of high-resolution soil moisture maps over the Iberian peninsula. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(9):3845–3857.
- Priestley, C. H. B. and R. J. Taylor. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, 100(2):81–92.
- Puri, S., H. Stephen, and S. Ahmad. (2011). Relating TRMM precipitation radar land surface backscatter response to soil moisture in the Southern United States. *Journal of Hydrology*, 402(1-2):115–125.
- Rubiolo, M.. (2014). *Desarrollo de nuevos modelos y algoritmos basados en redes neuronales para tareas de minería de datos*. PhD thesis, Facultad Regional Santa Fe. Universidad Nacional Tecnológica Nacional.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. (1988). Learning internal representations by error propagation. In *Readings in Cognitive Science*, Pp. 399–421. Elsevier.
- Said, S., U. C. Kothiyari, and M. K. Arora. (2008). ANN-based soil moisture retrieval over bare and vegetated areas using ERS-2 SAR data. *Journal of Hydrologic Engineering*, Pp. 461–475.
- Sánchez, N., M. Piles, J. Martínez-Fernández, M. Vall-llossera, L. Pipia, A. Camps, A. Aguasca, F. Pérez-Aragüés, and C. M. Herrero-Jiménez. (2014). Hyperspectral optical, thermal, and microwave L-Band observations for soil moisture retrieval at very high spatial resolution. *Photogrammetric Engineering & Remote Sensing*, 80(8):745–755.
- Santi, E., S. Paloscia, S. Pettinato, D. Entekhabi, S. H. Alemohammad, and A. G. Konings. (2016). Integration of passive and active microwave data from SMAP, AMSR2 and Sentinel-1 for soil moisture monitoring. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016-November:5252–5255.

- Schmugge, T. and T. J. Jackson. (1994). Mapping surface soil moisture with microwave radiometers. *Meteorology and Atmospheric Physics*, 54, Issue 1–4:213–223.
- Shi, J., J. Wang, A. Y. Hsu, P. E. O'Neill, and E. T. Engman. (1997). Estimation of bare surface soil moisture and surface roughness parameter using L-band SAR image data. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1254–1266.
- Sobrino, J. A., B. Franch, C. Mattar, J. C. Jiménez-Muñoz, and C. Corbari. (2012). A method to estimate soil moisture from Airborne Hyperspectral Scanner (AHS) and ASTER data: Application to SEN2FLEX and SEN3EXP campaigns. *Remote Sensing of Environment*, 117:415–428.
- Srivastava, P. K., D. Han, M. A. Rico-Ramirez, D. Al-Shrafany, and T. Islam. (2013). Data fusion techniques for improving soil moisture deficit using SMOS satellite and WRF-NOAH land surface model. *Water Resources Management*, 27(15):5069–5087.
- Thibeault, M., J. Cáceres, D. Dadamia, A. Soldano, M. Uriburu Quirno, J. Guerrieri, R. Edrosa, M. Palomeque, L. Romaldi, J. Pucheta, J. Mogadouro, E. De Luca, S. Bustos, S. Aguero, I. Pascual, and M. Mariotti. (2015). Spatial and temporal analysis of the monte buey saocom and smap core site. In *IGARSS 2015*, Pp. 969–971.
- Thorntwaite, C. W.. (1948). An approach toward a rational classification of climate. *American Geographical Society*, 38:55–94.
- Ulaby, F., R. Moore, and A. Fung. (1982). *Microwave Remote Sensing, Volumes 1, 2, and 3*. Addison Wesley, Reading, MA.
- Ulaby, F., R. Moore, and A. K. Fung. (1986). *Microwave Remote Sensing: Active and Passive, Vol. III, From Theory to Applications*. Artech House, Norwood, MA.
- Ulaby, F. T., P. C. Dubois, and J. Van Zyl. (1996). Radar mapping of surface soil moisture. *Journal of Hydrology*, 184(1-2):57–84.
- Van Der Velde, R., Z. Su, and Y. Ma. (2008). Impact of soil moisture dynamics on ASAR σ_0 signatures and its spatial variability observed over the Tibetan Plateau. *Sensors*, 8(9):5479–5491.

- Venturini, V. (2007). *Estimation of evaporative fraction and evapotranspiration from remotely sensed data using a complementary relationship*. PhD thesis, Universidad Nacional del Litoral.
- Venturini, V., S. Islam, and L. Rodriguez. (2008). Estimation of evaporative fraction and evapotranspiration from MODIS products using a complementary based model. *Remote Sensing of Environment*, 112(1):132–141.
- Venturini, V., L. Rodriguez, and G. Bisht. (2011). A comparison among different modified priestley and taylor equations to calculate actual evapotranspiration with MODIS data. *International Journal of Remote Sensing*, 32(5):1319–1338.
- Verheye, W.. (2004). *Agro-Climate-Based Land Evaluation Systems, in Land Use and Land Cover*, willy h. verheye edition. Encyclopedia of Life Support Systems (EOLSS), Developed under the Auspices of the UNESCO, Eolss Publishers, Oxford, UK.
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- Watson, D. and G. M. Philip. (1985). A refinement of inverse distance weighted interpolation. *Geoprocessing*, (2):315–327.
- Weisberg, S.. (2005). *Applied Linear Regression*, third edition. Minneapolis, Minnesota: A John Wiley & Sons, Inc., Publication.
- Western, A. W.. (1999). Observed spatial organization of soil moisture. *Water Resources Research*, 35(3):797–810.
- Wickel, A. J., T. J. Jackson, and E. F. Wood. (2001). Multitemporal monitoring of soil moisture with RADARSAT SAR during the 1997 Southern Great Plains hydrology experiment. *International Journal of Remote Sensing*, 22(8):1571–1583.
- Zeng, J., K. S. Chen, H. Bi, and Q. Chen. (2016). A preliminary evaluation of the SMAP radiometer soil moisture product over United States and Europe using ground-based measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 54(8):4929–4940.

- Zhan, X., P. R. Houser, J. P. Walker, and W. T. Crow. (2006). A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1534–1544.
- Zribi, M., N. Baghdadi, N. Holah, and O. Fafin. (2005). New methodology for soil surface moisture estimation and its application to ENVISAT-ASAR multi-incidence data inversion. *Remote Sensing of Environment*, 96(3-4):485–496.

Anexos

Anexo A

Análisis de la consistencia de los datos

Antes de comenzar a utilizar los datos provenientes de las estaciones *in situ* ubicadas en la provincia de Córdoba, se realizó un análisis de consistencia de los mismos. Para determinar la consistencia relativa, se comparan las observaciones obtenidas a partir de una cierta estación con la media de las observaciones de varias estaciones cercanas.

En las siguientes figuras se presentan los diagramas boxplot de las variables hidrometeorológicas de las estaciones *in situ* de la red de INTA y CONAE. El diagrama boxplot describe la mediana de los conjuntos de datos como la línea horizontal en la caja. El percentil 25 y 75 de los datos se define por los límites inferior y superior de la caja, respectivamente. La línea en el extremo inferior de los datos es el percentil 5 mientras que la línea en el extremo superior es el percentil 95.

En la Figura [A.1](#) se presenta la variable *HR*, en la que se observan valores medios entre 65 y 95 %. Estos valores de humedad altos durante todo el año podrían asociarse al fenómeno de “El Niño” que afectó a la región en el periodo 2015/2016. Por otra parte, se insinúa un ciclo anual que va desde diciembre a mayo con mayor humedad que durante los meses de junio a noviembre.

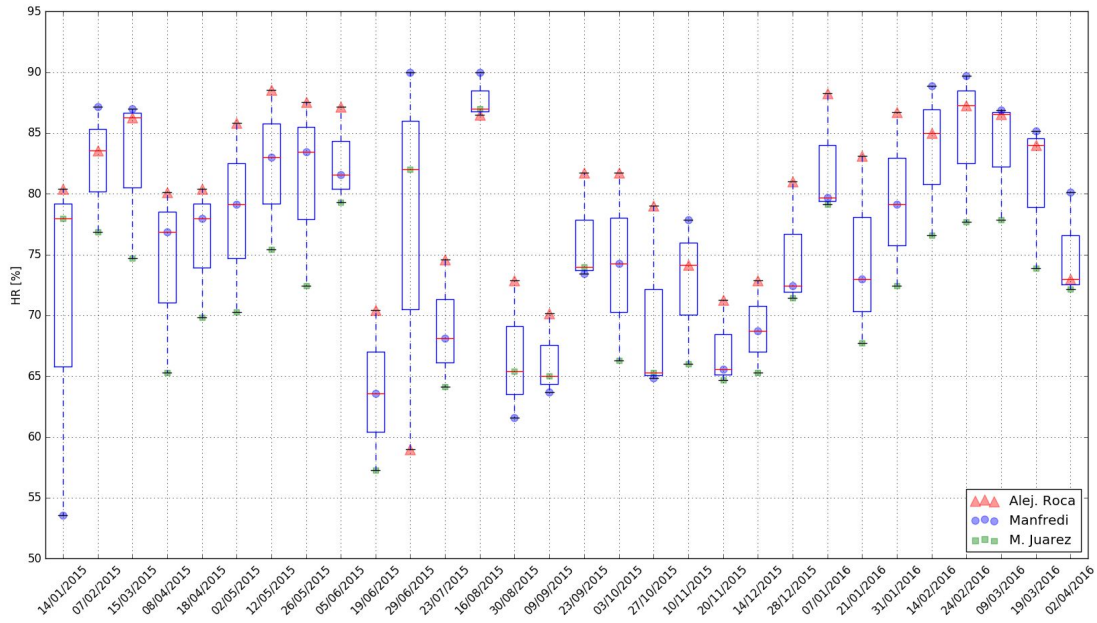


Figura A.1: Boxplot de la variación de HR de INTA respecto a la media regional

En las Figuras [A.2](#) y [A.3](#) se presentan las variables T_a y e_a , respectivamente. Estas variables presentan menor dispersión entre las estaciones INTA, observando un ciclo anual con valores máximos de T_a y e_a en los meses de verano, y valores mínimos en los meses de invierno.

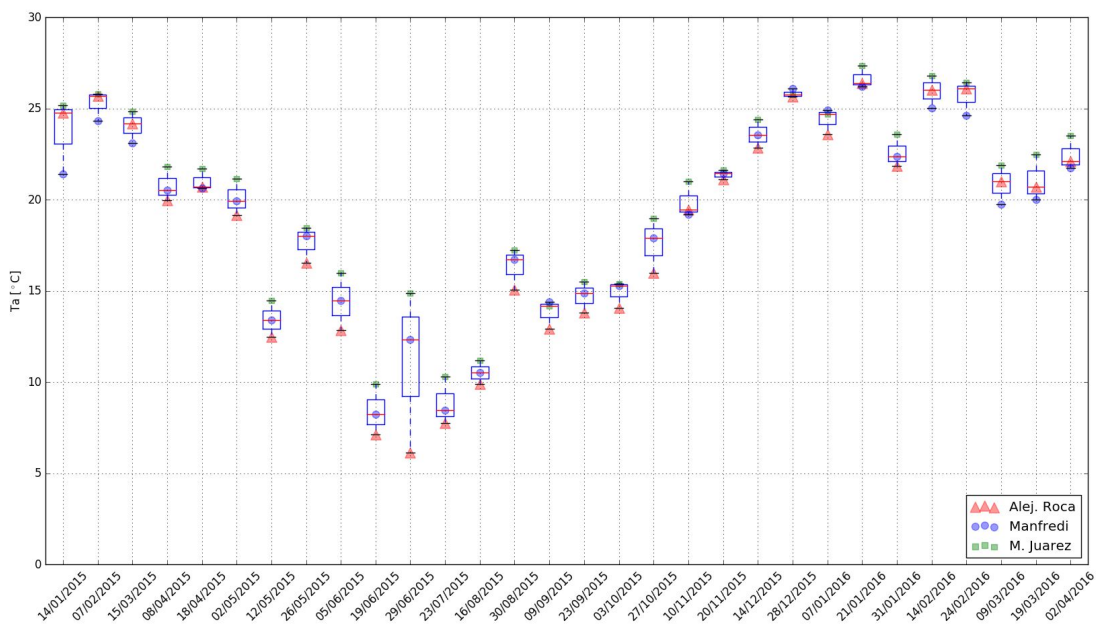


Figura A.2: Boxplot de la variación de T_a de INTA respecto a la media regional

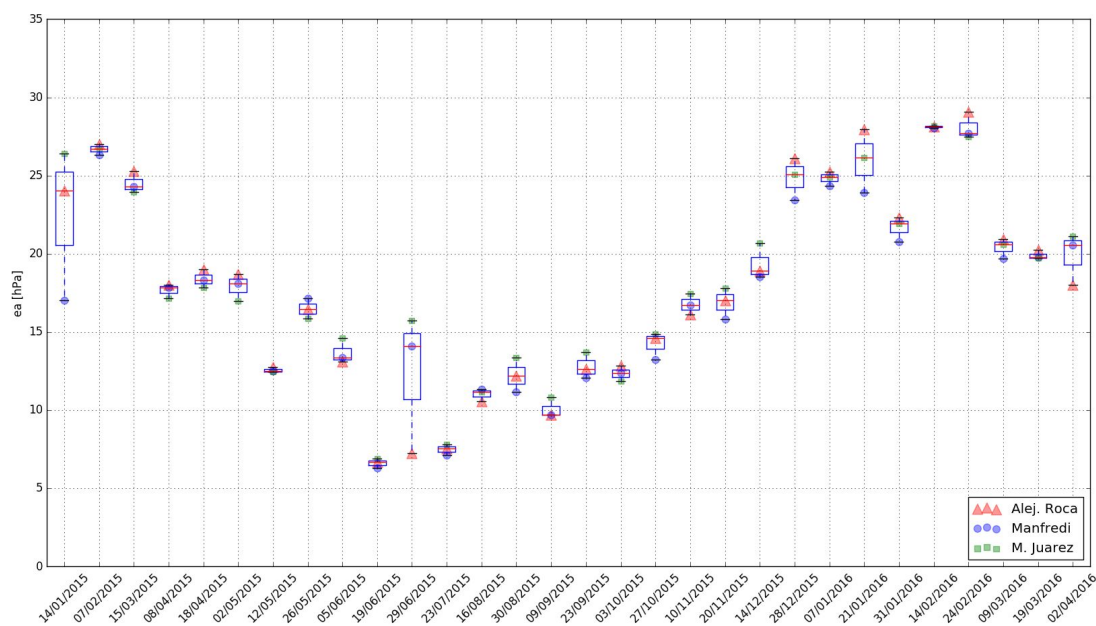


Figura A.3: Boxplot de la variación de e_a de INTA respecto a la media regional

En la Figura [A.4](#) se muestra la PP medida en las tres estaciones. Se observa que las mayores precipitaciones se producen en el período septiembre- abril y en los meses de mayo a agosto, las precipitaciones son prácticamente nulas excepto por eventos puntuales como es el caso para la fecha 16/08/2015. Las mayores diferencias entre las mediciones obtenidas por las estaciones se producen en los meses lluviosos, estas diferencias están asociadas a las variaciones espaciales propias del fenómeno.

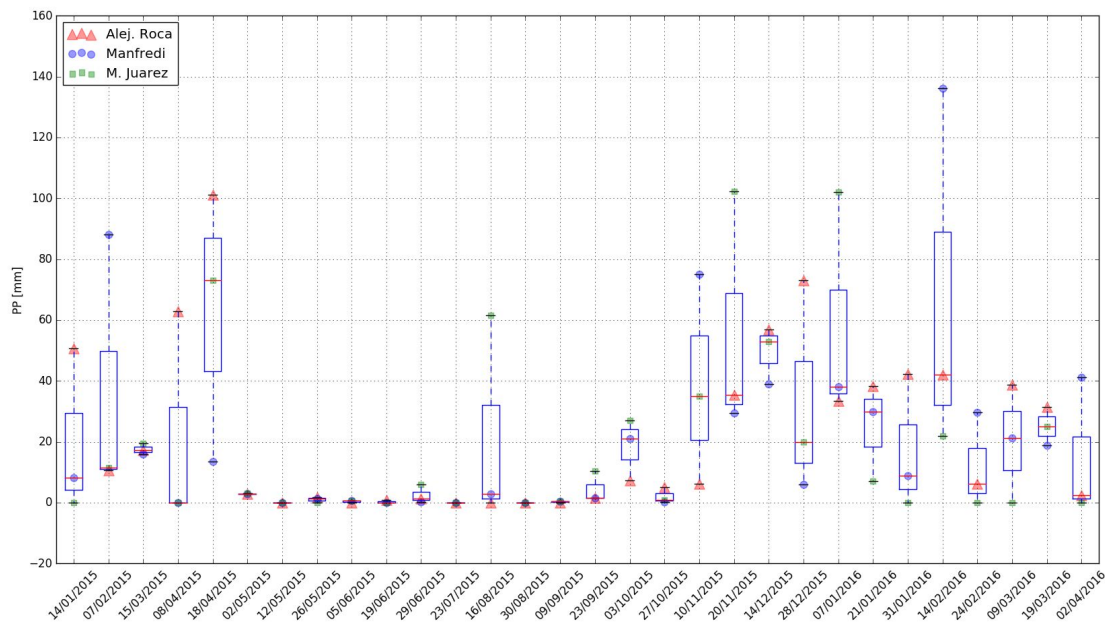


Figura A.4: Boxplot de la variación de PP de INTA respecto a la media regional

En las Figuras [A.5](#) y [A.6](#) se presentan las variaciones espaciales de las observaciones *in situ* obtenidas por la red de estaciones de CONAE. En la Figura [A.5](#) se observa la variación de la HS respecto de la media regional. Las mayores diferencias entre las muestras se producen durante la época lluviosa, probablemente causado por dispersión de las precipitaciones ya mencionadas. Los valores de HS más bajos se observan durante los meses de mayo a septiembre, y los más altos en los meses de febrero a abril.

La T_s (Figura [A.6](#)), presenta un comportamiento anual similar al de HS pero con menor dispersión. Además, los valores más altos de temperatura se dan de diciembre a abril con muy poca variación.

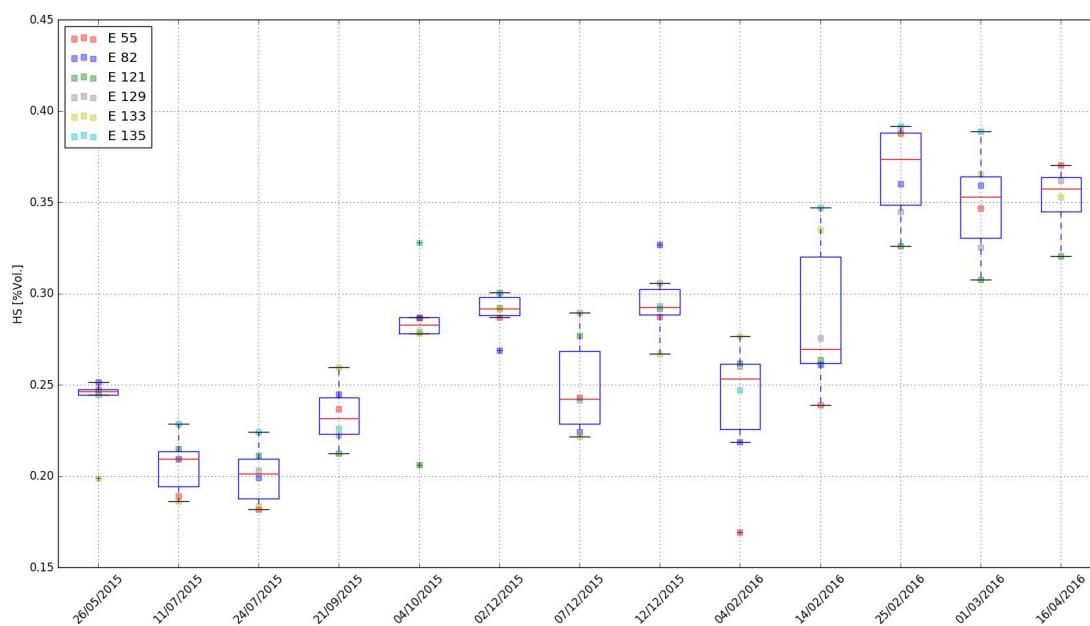


Figura A.5: Boxplot de la variación de HS de CONAE respecto a la media regional

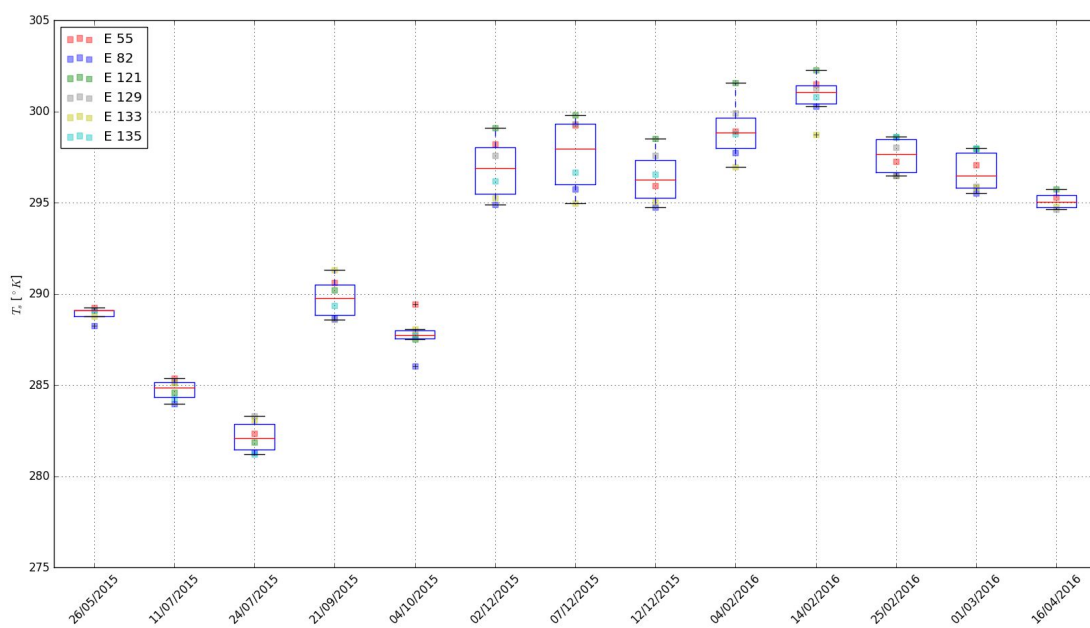


Figura A.6: Boxplot de la variación de T_s de CONAE respecto a la media regional

Anexo B

Implementación en Python

Parte del trabajo presentado en esta tesis se realizó con software desarrollado bajo el lenguaje de programación Python (<http://www.python.org/>). Python es un lenguaje interpretado, multiparadigma y multiplataforma que presenta una sintaxis limpia favoreciendo la legibilidad del código. Otra ventaja de Python es la gran cantidad de librerías existentes, entre ellas cabe destacar las de procesamiento de imágenes satelitales.

El software desarrollado permitió el procesamiento de las imágenes satelitales, el manejo de la información, la implementación de los métodos estadísticos, la obtención de mapas de humedad, entre otros. Las librerías utilizadas más importantes fueron las siguientes:

- **Scikit-learn**: esta librería se utilizó para implementar el modelo estadístico MLP.
- **Statsmodels**: utilizada para implementar el modelo MLR.
- **Pandas**: permitió estructurar los datos, como así también su procesamiento y análisis.
- **NumPy**: permitió procesar matrices y realizar cálculos en forma matricial.
- **Matplotlib**: se utilizó para realizar los gráficos.
- **Gdal**: proporcionó funcionalidades para la manipulación y procesamiento de las imágenes satelitales.

A continuación se presentan fragmentos de código de los programas implementados más importantes. Las versiones completas se encuentran alojadas en el siguiente

repositorio:

<https://github.com/gabrielagustin/Desarrollo-Tesis-Doctoral>.

Estimación de HS con datos observados

```
1  #!/usr/bin/python
2  import pandas as pd
3  import statsmodels.formula.api as smf
4  import selection
5  import matplotlib.pyplot as plt
6  import numpy as np
7  import statistics
8  import sklearn
9  import lectura
10 import application
11 import MLPregression
12 import copy
13 import SMmaps
14
15 # se lee el archivo .csv utilizando la librería lectura
16 # en lectura se realiza la limpieza y el formateo de los datos
17 file = "xxx.csv"
18 data = lectura.lecturaCompleta_etapa1(file)
19
20 # se mezclan las observaciones de las tablas
21 # semilla para mezclar los datos en forma aleatoria
22 np.random.seed(0)
23 dataNew = selection.shuffle(data)
24 dataNew = dataNew.reset_index(drop=True)
25
26 # formula
27 formula = "SM_CONAE ~ 1+T_aire+HR+PP+Sigma0"
28 print "Modelo planteado: " + str(formula)
29 model2 = smf.ols(formula, dataNew).fit()
30 print "R^2 del modelo: " + str(model2.rsquared)
31
32 # porcentaje de datos para la calibración
33 porc = 0.75
34 print "Porcentaje de datos para la calibración: " + str(porc)
35
36 # division de los datos para calibración y validación
```

```

37 nRow = len(dataNew.index)
38 numTraining=int(round(nRow)*porc)
39 print "Cantidad de elementos para el cálculo de coeff: "
40 + str(numTraining)
41 numTest=int((nRow)-numTraining)
42 print "Cantidad de elementos para la validación: " +str(numTest)
43
44 dataTraining = dataNew.ix[:numTraining, :]
45 dataTraining = selection.shuffle(dataTraining)
46 dataTraining = dataTraining.reset_index(drop=True)
47
48 dataTest = dataNew.ix[numTraining + 1:, :]
49
50 # Calibración
51 MLRmodel = smf.ols(formula, dataTraining).fit()
52 print MLRmodel.summary()
53 print "R^2 del modelo: " + str(MLRmodel.rsquared)
54
55 # Error de calibración
56 xxx = copy.copy(dataTraining)
57 del xxx['SM_CONAE']
58 yTraining = dataTraining['SM_CONAE']
59 yCal = MLRmodel.predict(xxx)
60 # se aplica la operación inversa a la operación logaritmo
61 # aplicada en lectura
62 yTraining = 10**(yTraining)
63 yCal = 10**(yCal)
64
65 rmse = statistics.RMSE(np.array(yTraining),np.array(yCal))
66 print "RMSE:" + str(rmse)
67 bias = statistics.bias(yTraining,yCal)
68 print "Bias:" + str(bias)
69
70 # se guardan los coeficientes del modelo entrenado
71 print "Los coeficientes del modelo son: "
72 coeff = MLRmodel.params
73
74 # Validación del modelo
75 y = np.array(dataTest["SM_CONAE"])
76 pred = MLRmodel.predict(dataTest)
77 yAprox = np.array(pred)
78 # se aplica la operación inversa a la operación logaritmo
79 # aplicada en lectura

```

```

80
81 y = 10**(y)
82 yAprox = 10**(yAprox)
83
84 bias = statistics.bias(yAprox,y)
85 print "Bias Validacion:" + str(bias)
86
87 # Error de validación
88 rmse = 0
89 rmse = statistics.RMSE(y,yAprox)
90 print "RMSE:" + str(rmse)
91 RR = smf.ols('y ~ 1+ yAprox', dataTest).fit().rsquared
92 print "R^2: "+str(RR)
93
94 # Se calibra y valida MLP para esto se utiliza la librería
95 # MLPregression
96 MLPmodel, yCalMLP, yAproxMLP = MLPregression.mlp(porc,file, "etapa1")
97
98 # Se obtienen los mapas de HS con los modelos calibrados
99 SMmaps.calculateMaps(MLRmodel, MLPmodel)

```

```

1 # -*- coding: utf-8 -*-
2 import lectura
3 import numpy as np
4 import selection
5 from sklearn.neural_network import MLPRegressor
6 import matplotlib.pyplot as plt
7 import statistics
8 import sklearn
9 import pandas as pd
10 import statsmodels.formula.api as smf
11
12 def mlp(porc, file, type):
13     data = lectura.lecturaCompletaMLP_etapa1(file)
14     varCal = 'SM_CONAE'
15     # números de neuronas en las 2 capas ocultas
16     indexC1 = 1
17     indexC2 = 4
18

```

```

19     # parámetros del perceptron multicapa
20     sol = 'sgd'
21     iter = 5000
22     random_s = 9
23     lr = 'adaptive'
24     l_rate = 0.0001
25     alpa = 0.015
26     momen = 0.75
27
28     # se dividen los datos según porc en los bloques de
29     # calibración y validación
30     np.random.seed(0)
31     dataNew = selection.shuffle(data)
32     dataNew = dataNew.reset_index(drop=True)
33     nRow = len(dataNew.index)
34     numTraining=int(round(nRow)*porc)
35     print "Cantidad de elementos para el calculo de coeff: " + str(numTraining)
36     numTest=int((nRow)-numTraining)
37     print "Cantidad de elementos para prueba: " +str(numTest)
38     dataTraining = dataNew.ix[:numTraining, :]
39     dataTraining = selection.shuffle(dataTraining)
40     dataTraining = dataTraining.reset_index(drop=True)
41     dataTest = dataNew.ix[numTraining + 1:, :]
42
43     # se normalizan los datos entre -1 y 1
44     OldRange = (np.max(dataTraining.T_aire) - np.min(dataTraining.T_aire))
45     NewRange = (1 + 1)
46     dataTraining.T_aire = (((dataTraining.T_aire - np.min(dataTraining.T_aire))
47     * NewRange) / OldRange) -1
48
49     OldRange = (np.max(dataTraining.HR) - np.min(dataTraining.HR))
50     NewRange = (1 + 1)
51     dataTraining.HR = (((dataTraining.HR - np.min(dataTraining.HR))
52     * NewRange) / OldRange) -1
53
54     OldRange = (np.max(dataTraining.PP) - np.min(dataTraining.PP))
55     NewRange = (1 + 1)
56     dataTraining.PP = (((dataTraining.PP - np.min(dataTraining.PP))
57     * NewRange) / OldRange) -1
58
59     OldRange = (np.max(dataTraining.Sigma0) - np.min(dataTraining.Sigma0))
60     NewRange = (1 + 1)
61     dataTraining.Sigma0 = (((dataTraining.Sigma0 - np.min(dataTraining.Sigma0))

```

```

62     * NewRange) / OldRange) -1
63
64     OldRange = (np.max(dataTest.T_aire) - np.min(dataTest.T_aire))
65     NewRange = (1 + 1)
66     dataTest.T_aire = (((dataTest.T_aire - np.min(dataTest.T_aire))
67     * NewRange) / OldRange) -1
68
69     OldRange = (np.max(dataTest.HR) - np.min(dataTest.HR))
70     NewRange = (1 + 1)
71     dataTest.HR = (((dataTest.HR - np.min(dataTest.HR))
72     * NewRange) / OldRange) -1
73
74     OldRange = (np.max(dataTest.PP) - np.min(dataTest.PP))
75     NewRange = (1 + 1)
76     dataTest.PP = (((dataTest.PP - np.min(dataTest.PP))
77     * NewRange) / OldRange) -1
78
79     OldRange = (np.max(dataTest.Sigma0) - np.min(dataTest.Sigma0))
80     NewRange = (1 + 1)
81     dataTest.Sigma0 = (((dataTest.Sigma0 - np.min(dataTest.Sigma0))
82     * NewRange) / OldRange) -1
83
84     yTraining = dataTraining[varCal]
85     del dataTraining[varCal]
86     xTraining = dataTraining
87
88     yTest = dataTest[varCal]
89     del dataTest[varCal]
90     test_x = dataTest
91
92     # se define la red neuronal
93     print "indice capa 1: " + str(indexC1)
94     print "indice capa 2: " + str(indexC2)
95     if (indexC2 == 0):
96         reg = MLPRegressor(hidden_layer_sizes=(indexC1), activation=acti,
97         solver=sol, alpha=alpa, batch_size='auto', learning_rate=lr,
98         learning_rate_init=l_rate, power_t=0.5, max_iter=iter,
99         shuffle=True, random_state=random_s, tol=0.0001,
100        verbose=False, warm_start=False, momentum=momen,
101        nesterovs_momentum=True, early_stopping=False,
102        validation_fraction=0.3, beta_1=0.9, beta_2=0.999,
103        epsilon=1e-08)
104     else:

```

```

105     reg = MLPRegressor(hidden_layer_sizes=(indexC1,indexC2), activation= acti,
106     solver= sol, alpha=alpa,batch_size='auto', learning_rate=lr,
107     learning_rate_init=l_rate, power_t=0.5, max_iter=iter,
108     shuffle=True, random_state=random_s, tol=0.0001,
109     verbose=False, warm_start=False, momentum=momen,
110     nesterovs_momentum=True, early_stopping=False,
111     validation_fraction=0.3, beta_1=0.9, beta_2=0.999,
112     epsilon=1e-08)
113
114     # Calibración de la red
115     reg = reg.fit(xTraining, yTraining)
116     # se predice los valores de humedad
117     yCal = reg.predict(xTraining)
118
119     # Error de calibración
120     rmse = statistics.RMSE(np.array(yTraining),np.array(yCal))
121     print "RMSE:" + str(rmse)
122     data2 = pd.DataFrame({'yCal' :yCal,'yTraining' :yTraining})
123     RR = smf.ols('yCal ~ 1+ yTraining', data2).fit().rsquared
124     print "Coeficiente de Determinacion:" + str(RR)
125     bias = statistics.bias(yTraining, yCal)
126     print "Bias:" + str(bias)
127
128     # Validación de la red
129     yAprox = reg.predict(test_x)
130
131     # Error de validación
132     rmse = statistics.RMSE(np.array(yTest),np.array(yAprox))
133     print "RMSE:" + str(rmse)
134     data2 = pd.DataFrame({'yTest' :yTest,'yAprox' :yAprox})
135     RR = smf.ols('yTest ~ 1+ yAprox', data2).fit().rsquared
136     print "Coeficiente de Determinacion:" + str(RR)
137     bias = statistics.bias(yTest, yAprox)
138     print "Bias:" + str(bias)
139
140     print "Correlación de Pearson: "
141     print np.corrcoef(yAprox,yTest)[1,0]
142
143     return reg, yCal, yAprox

```

Obtención de mapas de HS

```

1  #!/usr/bin/python
2  from gdalconst import *
3  import matplotlib.pyplot as plt
4  import numpy.ma as np
5  from matplotlib import cm
6  from osgeo import gdal, ogr
7  import sys
8  import pandas as pd
9  import functions
10
11 def calculateMaps(MLRmodel, MLPmodel):
12     dir = "xxx"
13     path = "xxx"+dir+"xxx"
14
15     # vectores fechas de las imágenes de entrada
16     fechaSentinel= []
17     fechaNDVI= []
18     fechaLandsat8=[]
19     fechaSMAP=[]
20     fechaMYD=[]
21
22     # vectores de las variables
23     Ta = []
24     HR = []
25     PP = []
26     sigma0 = []
27     # se recorre las fechas indicadas en fechaSentinel
28     for i in range(0,len(fechaSentinel)):
29         # por cada fecha se cargan las imágenes de entrada
30         fileTa = "xxx"+fechaSentinel[i]+"/T_aire.asc"
31         src_ds_Ta, bandTa, GeoTTa, ProjectTa = functions.openFileHDF(fileTa,1)
32
33         filePP = "xxx"+"/PP.asc"
34         src_ds_PP, bandPP, GeoTPP, ProjectPP = functions.openFileHDF(filePP,1)
35
36         fileHR = "xxx"+fechaSentinel[i]+"/HR.asc"
37         src_ds_HR, bandHR, GeoTHR, ProjectHR = functions.openFileHDF(fileHR,1)
38
39         fileNDVI = "xxx"+fechaLandsat8[i]+"/NDVI_recortado"
40         src_ds_NDVI, bandNDVI, GeoTNDVI, ProjectNDVI =
41         functions.openFileHDF(fileNDVI,1)

```



```

42
43     fileSar = "xxx"+fechaSentinel[i]+
44     ".SAFE/subset.data/recorte_30mx30m.img"
45
46     nameFileMLR = "mapa_MLR_30m_"+str(fechaSentinel[i])
47     nameFileMLP = "mapa_MLP_30m_"+str(fechaSentinel[i])
48
49     src_ds_Sar, bandSar, GeoTSar, ProjectSar =
50     functions.openFileHDF(fileSar, 1)
51
52     fileMascara = "xxx"+ "mascaraciudad.img"
53     src_ds_Mas, bandMas, GeoTMas, ProjectMas =
54     functions.openFileHDF(fileMascara, 1)
55
56     # se cambian las resoluciones y proyecciones de todas las
57     # imágenes a la de las imágenes SAR
58
59     # tipo de interpolación
60     #type = "Nearest"
61     type = "Bilinear"
62     nRow, nCol = bandSar.shape
63
64     data_src = src_ds_Mas
65     data_match = src_ds_Sar
66     match = functions.matchData(data_src, data_match, type, nRow, nCol)
67     band_matchCity = match.ReadAsArray()
68
69     data_src = src_ds_Ta
70     data_match = src_ds_Sar
71     match = functions.matchData(data_src, data_match, type, nRow, nCol)
72     band_matchTa = match.ReadAsArray()
73
74     data_src = src_ds_PP
75     data_match = src_ds_Sar
76     match = functions.matchData(data_src, data_match, type, nRow, nCol)
77     band_matchPP = match.ReadAsArray()
78
79     data_src = src_ds_HR
80     data_match = src_ds_Sar
81     match = functions.matchData(data_src, data_match, type, nRow, nCol)
82     band_matchHR = match.ReadAsArray()
83
84     data_src = src_ds_NDVI
85     data_match = src_ds_Sar

```

```

86     match = functions.matchData(data_src, data_match, type, nRow, nCol)
87     band_matchNDVI = match.ReadAsArray()
88
89     # se filtra la imagen SAR
90     # con NDVI mayores a 0.45 y con NDVI menores a 0
91     sarEnmask, maskNDVI = applyNDVIFilter(bandSar, band_matchNDVI, etapa)
92     r, c = bandSar.shape
93
94     # Se APLICA MÉTODO MLR
95     # los datos para el modelo MLR se modifican con la función
96     # logaritmo
97
98     # se crea la matriz de datos para MLR
99     dataMap_MLR = pd.DataFrame({'Sigma0' :sarEnmask.flatten(),
100     'T_aire' : (np.log10(band_matchTa)).flatten(),
101     'HR' : (np.log10(band_matchHR)).flatten(), 'PP' :band_matchPP.flatten()})
102
103     # se obtiene mapa
104     dataMap_MLR = dataMap_MLR.fillna(0)
105     mapSM_MLR = MLRmodel.predict(dataMap_MLR)
106     # se revierte la funcion flatten()
107     mapSM_MLR = mapSM_MLR.reshape(r, c)
108     mapSM_MLR = 10**(mapSM_MLR)
109     mapSM_MLR[mapSM_MLR < 0] = 0
110     mapSM_MLR = mapSM_MLR*maskNDVI
111
112     # Se APLICA MÉTODO MLP
113     # los datos para el modelo MLP se normalizan
114     # entre -1 y 1
115     OldRange = (np.max(band_matchTa) - np.min(band_matchTa))
116     NewRange = (1 + 1)
117     Ta = (((band_matchTa - np.min(band_matchTa)) *NewRange)/OldRange) -1
118
119     OldRange = (np.max(band_matchHR) - np.min(band_matchHR))
120     NewRange = (1 + 1)
121     HR = (((band_matchHR - np.min(band_matchHR)) *NewRange)/OldRange) -1
122
123     OldRange = (np.max(band_matchPP) - np.min(band_matchPP))
124     NewRange=(1 + 1)
125     PP = (((band_matchPP - np.min(band_matchPP)) *NewRange)/OldRange) -1
126
127     OldRange = (np.max(sarEnmask2) - np.min(sarEnmask2))
128     NewRange = (1 + 1)

```

```
129 sar2 = (((sarEnmask2 - np.min(sarEnmask2)) * NewRange)/OldRange)-1
130
131 # se crea matriz de datos para MLP
132 dataMap_MLP = pd.DataFrame({'T_aire' :Ta.flatten(),
133 'Sigma0' :sar2.flatten(), 'HR' :HR.flatten(), 'PP' :PP.flatten()})
134 dataMap_MLP = dataMap_MLP[['T_aire', 'PP', 'Sigma0', 'HR']]
135
136 # se obtiene mapa
137 dataMap_MLP = dataMap_MLP.fillna(0)
138 mapSM_MLP = MLPmodel.predict(dataMap_MLP)
139 mapSM_MLP = mapSM_MLP.reshape(r,c)
140 mapSM_MLP[mapSM_MLP < 0] = 0
141 mapSM_MLP = mapSM_MLP*maskNDVI
142
143 # se guardan los mapas creados por cada método en la fecha i
144 functions.createHDFfile(path, nameFileMLR, 'ENVI', mapSM_MLP,
145 c, r, GeoTSar, ProjectSar)
146 functions.createHDFfile(path, nameFileMLP, 'ENVI', mapSM_MLP,
147 c, r, GeoTSar, ProjectSar)
```


Anexo C

Remuestreo mapas NDVI utilizando transformada ondita

En este apartado se propone analizar la eficiencia del algoritmo propuesto para realizar el rescalado de las imágenes NDVI del sensor MODIS. La técnica implementada está basada en la transformada ondita.

El análisis multiresolución, en el cual se basa la teoría ondita, consiste en expandir una señal en una base de funciones cuyas propiedades tiempo-frecuencia se adapten a la estructura parcial de la señal. La transformada ondita discreta unidimensional (DWT-1D, por sus siglas en inglés) de una señal se obtiene mediante el algoritmo denominado representación ondita, basado en una estructura de banco de filtros (Mallat, 1989). Es decir, la señal $f[n]$ se transfiere a través de un filtro paso alto ($h[n]$) y un filtro de paso bajo ($L[n]$), seguido de una operación de sub-muestreo que reduce a la mitad la señal, para calcular tanto los coeficientes de detalle, $D_{j-1}^2 f$, como de aproximación, $A_{j-1}^2 f$. Esta operación se repite de forma iterativa hasta lograr el nivel de descomposición deseado, aplicando el mismo proceso a los coeficientes de aproximación obtenidos en el nivel anterior. La transformada ondita discreta de dos dimensiones (DWT -2D) separa las filas de las columnas de una imagen y realiza el mismo proceso que en la DWT-1D, considerando a cada fila y a cada columna como una señal unidimensional (Figura C.1). En cada paso se descompone la señal $A_j^2 f$ en $A_{j-1}^2 f$, $D_{2j-1}^1 f$, $D_{2j-1}^2 f$ y $D_{2j-1}^3 f$, (Mallat, 1989).

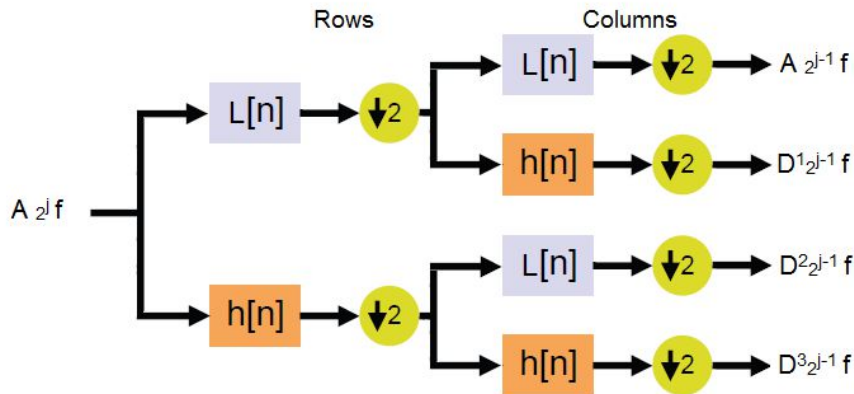


Figura C.1: Esquema transformada ondita discreta en dos dimensiones (DWT -2D)

En esta tesis se propone utilizar el proceso inverso de esta metodología, es decir, utilizar la reconstrucción de la señal partiendo de la imagen original y de esta manera aumentar su resolución espacial. La reconstrucción se realiza utilizando la transformada ondita discreta inversa (IDWT) en dos dimensiones. Para lo cual, se toma como el coeficiente de aproximación inicial (A_j^i) la imagen MODIS con resolución espacial 250 m, y se aplica la IDWT en tres ocasiones. Los coeficientes de detalle se asumen del mismo tamaño que la imagen de aproximación pero iguales a cero. El esquema propuesto se muestra en la Figura C.2.

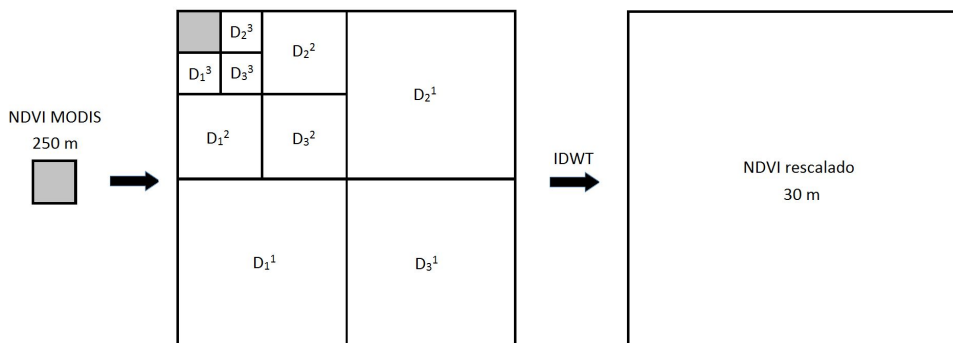


Figura C.2: Rescalado con transformada ondita inversa (IDWT)

Para realizar el cambio de resolución se han probado diferentes familias de onditas: *Haar*, *Daubechies*, *Coiflets* y *Symlets*, con diferentes órdenes, es decir, diferentes onditas con distintos números de coeficientes. La finalidad de este análisis es encontrar la familia ondita y el orden más adecuado para esta aplicación. Para comprobar el desempeño de cada configuración se contrastaron las imágenes NDVI MODIS re-escaladas a 30 m con imágenes Landsat 8, las fechas utilizadas se presentan en la Tabla C.1.

Tabla C.1: Inventario de las imágenes NDVI MODIS y Landsat 8

Fechas (mm/dd/yyyy)	
Terra-MOD13Q1	Landsat 8
06/26/2015	06/18/2015
07/28/2015	07/20/2015
10/16/2015	10/08/2015
12/19/2015	12/27/2015
03/21/2016	03/16/2016

El RMSE y R^2 obtenidos para las distintas familias de onditas y diferentes órdenes se presenta en la Tabla C.2. El mejor desempeño se obtuvo utilizando la familia *Coiflets* de orden 2, con la cual se alcanzó un RMSE igual a 0.19 adimensional y un R^2 igual a 0.46.

Tabla C.2: RMSE y R^2 para distintas familias onditas

Familia Ondita	Orden	RMSE	R^2
Daubechies	2	0.20	0.38
	4	0.20	0.41
Coiflets	2	0.19	0.46
	4	0.22	0.24
Symlets	2	0.20	0.39
	4	0.23	0.17
Haar	2	0.22	0.24

Anexo D

Análisis de los productos satelitales utilizados

En este apartado se presenta la validación de los datos satelitales utilizados en esta tesis. El contraste y análisis de los productos de SMAP y las observaciones *in situ* fue presentado en el XVII Simposio Internacional en Percepción Remota y Sistemas de Información Geográfica en Puerto Iguazú Misiones en 2016. El trabajo se denominó “Análisis de los productos de las misiones SMOS y SMAP en la región pampeana Argentina”.



ANÁLISIS DE LOS PRODUCTOS DE LAS MISIONES SMOS Y SMAP EN LA REGIÓN PAMPEANA ARGENTINA

Walker E.¹, G. García^{1,2} y V. Venturini¹

¹ Centro de estudios Hidro-Ambientales, Facultad de Ingeniería y Ciencias Hídricas.

Universidad Nacional del Litoral

ewalker@fich.unl.edu.ar

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

RESUMEN

El uso de las microondas ha potenciado el monitoreo de diferentes ambientes terrestres en condiciones climáticas adversas. Los datos remotamente obtenidos con sensores de microondas activos (Radar) y pasivos (Radiómetro) han probado ser los mejores instrumentos para cuantificar variables en los primeros centímetros del suelo. Los radiómetros poseen una óptima sensibilidad a la Humedad del Suelo (HS) inclusive bajo moderada cobertura vegetal. Estos sensores tienen la desventaja de proveer datos a baja resolución espacial (aproximadamente 40 Km). En contraste, los radares tienen la capacidad de proporcionar información a resolución espacial considerablemente mejor (menos de 10 Km) pero con menor sensibilidad a HS por la influencia de la rugosidad de la superficie. En este trabajo se propone verificar la capacidad de los radiómetros para captar la heterogeneidad natural de la HS y de la Temperatura de Superficie (TS) mediante un análisis comparativo de datos *in situ* con los productos provistos por las misiones satelitales Soil Moisture Active Passive (SMAP) y Soil Moisture Ocean Salinity (SMOS). Los datos de campo pertenecen a una red de sensores, que miden ambas variables, ubicados en la región Pampeana Argentina. En un primer análisis se comparan los promedios espaciales de las mediciones registradas dentro de un pixel con los valores provistos por ambas misiones. De este análisis se verifica que para HS hay una diferencia en los valores promedios de $0.0944 \text{ m}^3/\text{m}^3$ y $0.0205 \text{ m}^3/\text{m}^3$ para SMOS y SMAP respectivamente, presentando ambas series un R^2 de 0.30 para SMOS y 0.66 para SMAP. Los resultados muestran que la misión SMAP representa mejor que SMOS la HS promedio en el pixel. En el caso de la temperatura, solo se efectuó el análisis del producto provisto por SMAP. La diferencia de promedios es de $1.581 \text{ }^\circ\text{K}$ y la comparación arroja un R^2 de 0.91. Posteriormente se analizó estadísticamente la diferencia de las medias de ambas misiones. En el caso de la misión SMAP ambas variables (HS y TS) demostraron que no hay diferencia significativa ($p < 0.05$) entre los datos *in situ* y los estadísticos obtenidos por cada misión. Si bien la dinámica espacial de la HS es muy importante comparada con la de TS, el radiómetro de dicha misión representa correctamente la dinámica espacial de las variables. Para la misión satelital SMOS se presentan diferencias significativas para la HS al comparar los datos, lo que podría sugerir que los productos de este radiómetro deberían ser fusionados con los de otras fuentes de medición para lograr información con aplicabilidad agronómica e hidrológica. Por último, se analizó gráficamente la variabilidad para determinar la significancia de la heterogeneidad de las mediciones registradas en la superficie.

PALABRAS CLAVE: HUMEDAD DE SUELO, TEMPERATURA DE SUPERFICIE, SMOS, SMAP

ABSTRACT

The use of microwaves has enhanced monitoring of various terrestrial environments under adverse weather conditions. Remotely sensed data with active (radar) and passive (radiometer) microwave sensors have proven to be the best instruments to observe variables in the first centimeters of the soil. Radiometers have better sensitivity to the Soil Moisture (SM) even under moderate vegetation cover. These sensors have the disadvantage of providing data at lower spatial resolution (approximately 40 Km). In contrast, radars have the ability to provide information with better spatial resolution (less than 10 Km) but with less sensitivity to SM by the influence of the surface roughness. This paper purpose is to verify the radiometers capability to capture the natural heterogeneity of the SM and land surface temperature (ST), by comparing *in situ* data with the products provided by satellite missions Soil Moisture Active Passive (SMAP) and Soil Moisture Ocean Salinity (SMOS). Field data are obtained by a network of sensors that measure both variables, located in the Pampa region of Argentina. First of all, the spatially averaged observations *in situ* within a pixel and measurements provided by both missions are compared. This analysis verified that there is a difference for HS in the mean values of 0.0944 m³/m³ and 0.0205 m³/m³ for SMOS and SMAP respectively, both series present an R² of 0.30 and 0.66 for SMOS and SMAP. The results show that SMAP mission represents the pixels average SM better than SMOS. In the case of ST, only the product provided by SMAP was analyzed. The mean difference is 1.581 °K and the comparison yield an R² of 0.91. Then the differences of the means of both missions were statistically analyzed. In the case of the SMAP mission both variables (SM and ST) showed no significant difference (p <0.05) between the data and the SMAP products. While the spatial dynamics of the SM is very important compared to the ST, the radiometer of this mission represents correctly the spatial dynamics of both variables. For the SMOS mission significant differences for the SM was obtained when comparing the data, which could suggest that the products of this radiometer should be merged with other sources of measurement to achieve information for agronomic and hydrological applications. Finally, the variability is graphically analyzed to determine the significance of the heterogeneity of the surface measurements registered.

KEYWORDS: SOIL MOISTURE, SURFACE TEMPERATURE, SMOS, SMAP

INTRODUCCIÓN

La predicción espacial de características de la superficie terrestre, tales como humedad y temperatura, son necesarias para múltiples aplicaciones en agronomía, hidrología y meteorología. La humedad del suelo (HS) es una variable de gran importancia en el ciclo hidrológico y en los procesos que ocurren en la interfaz suelo-atmósfera. HS refleja la partición de la precipitación en infiltración y escorrentía superficial, y controla la partición de la energía disponible entre el flujo de calor sensible y calor latente (Houser, 2005). La temperatura de superficie (TS), es uno de los parámetros más importante en los procesos físicos de la superficie terrestre y en los balances hídricos en general (Anderson et al., 2008). El seguimiento de la HS y de la TS en grandes áreas a través de mediciones *in situ* es prácticamente imposible de realizar. La teledetección aparece como una herramienta adecuada para la obtención frecuente de dichas observaciones en escala regional



y global. Investigaciones previas han demostrado el potencial de los sensores de microondas pasivos y activos en la cuantificación de variables en los primeros centímetros del suelo (Njoku et al., 1999; Wigneron et al., 1995; Jackson et al., 1981; Ulaby et al., 1978; Narayan y Njoku, 2004). Los radiómetros poseen una óptima sensibilidad a la HS pero una baja resolución espacial (aproximadamente 40 Km). En contraste, los radares tienen la capacidad de proporcionar información a una mejor resolución (menos de 10 Km) pero con una disminuida sensibilidad a la humedad por la influencia de la rugosidad de la superficie y la vegetación (Sobrino et al., 2012). Las bajas frecuencias de microondas (banda L) son consideradas óptimas para la estimación de la humedad superficial, debido a que penetran en profundidad el suelo y vegetación y son menos afectadas por las condiciones atmosféricas (Schmugge, 1983; Kerr et al., 2001). Por esta razón diferentes misiones satelitales basadas en radiómetros de microondas han sido propuestas por las agencias espaciales. Ya desde 1970 información lograda por los radiómetros Scanning Multichannel Microwave Radiometer (SMMR) y Advanced Microwave Scanning Radiometer (AMSR-E) ha sido utilizada para la estimación de la HS. Sin embargo, el primer satélite dedicado específicamente al sensoramiento remoto de la humedad, fue lanzado en el 2009 constituyendo la misión Soil Moisture and Ocean Salinity (SMOS) de la Agencia Espacial Europea (Kerr et al., 2010a). Más tarde, en el año 2015 la Administración Nacional de la Aeronáutica y del Espacio (NASA) ha puesto en órbita la misión satelital Soil Moisture Active Passive (SMAP) combinando un radiómetro y un radar de microondas para la medición del estado hídrico de la superficie terrestre (Entekhabi et al., 2010).

La heterogeneidad de las variables de la superficie terrestre existe ampliamente dentro de un píxel de teledetección pasiva de microondas. La variación espacial de la HS se puede atribuir a factores de escala local como el tipo de suelo, la topografía y la vegetación, y factores de gran escala, como la precipitación (Entin et al., 2000). Actualmente, los algoritmos desarrollados a partir de los sensores remotos pasivos se utilizan para obtener el contenido de HS promedio del píxel (Jackson y Le Vine, 1996; Njoku et al., 2003; Owe et al., 2008). La TS se presenta también como una variable promediada en la grilla de 36 kilómetros utilizada por la misión SMAP.

Por dicho motivo, es de gran importancia validar la precisión de los productos de las misiones satelitales y verificar la capacidad de los mismos en captar la heterogeneidad presente en las mediciones. En este trabajo la estrategia de validación adoptada se basa en la comparación de los productos satelitales con información provista por una red de sensores *in situ*. A pesar de ser conocido que la exactitud de los sensores en tierra tienen sus propios errores, se asume que estos (promediados sobre la pasada del satélite) son significativamente menores que los errores de los productos satelitales (Gebremichael y Vivoni, 2008). En la actualidad se encuentran trabajos de validación como el aquí presentado en diferentes partes del mundo, por ejemplo España (Sanchez et al., 2012), China (Dente et al., 2012), Estados Unidos y Europa (Zeng et al., 2016), entre otros. El objetivo de este trabajo es validar la capacidad de los radiómetros SMOS y SMAP en captar los parámetros superficiales y su correspondiente heterogeneidad en un área de la región Pampeana Argentina.

MATERIALES Y METODOS

Área De Estudio

La ciudad de Monte Buey se halla situada en el sudeste de la provincia de Córdoba, Argentina. Se encuentra junto a otras ciudades de importancia, formando parte de la Región Pampeana, una extensa llanura fértil que abarca la cuarta parte del territorio nacional (Figura1).

La economía del área de estudio se establece fundamentalmente en los sectores primarios y secundarios. La principal actividad es la agricultura basada en la producción de soja, trigo y maíz como cultivos primordiales, sucedidos en la rotación por sorgos, alfalfas o pasturas. Otras actividades de importancia son la ganadería y la elaboración e industrialización de las materias primas agropecuarias.

Desde el punto de vista climático, la zona de estudio se caracteriza por una isoterma media anual de 17.4°C siendo enero el mes más cálido (24.3 °C) y julio el más frío (10.5 °C). La precipitación media anual es de 920 mm/año, los cuales se distribuyen principalmente entre los meses de Septiembre-Abril. Fisiográficamente es representada por llanuras, determinada por relieves planos o con suaves pendientes. La zona de estudio se caracteriza por presentar suelos Argiudoles típicos, los cuales por su buen drenaje y alta fertilidad se consideran adecuados para el desarrollo de actividades agrícolas-ganaderas.



Figura 1: Ubicación del área de estudio, centrada en la ciudad de Monte Buey, Provincia de Córdoba, Argentina.

Datos Satelitales

Para la realización de este trabajo se han utilizado las observaciones de las misiones satelitales SMOS y SMAP.

La misión Soil Moisture and Ocean Salinity (SMOS) perteneciente a la Agencia Espacial Europea (ESA) tiene a bordo el instrumento Microwave Imaging Radiometer by Aperture Synthesis (MIRAS) un radiómetro interferométrico que proporciona datos espectrales en banda L (1.4 GHz). El producto utilizado para este estudio es el L2 Soil Moisture User Data Product (MIR_SMUDP2) el cual es obtenido mediante un algoritmo basado en el modelo tau-omega (Kerr y Njoku, 1990). El mismo utiliza la temperatura de brillo en múltiples ángulos de incidencia

junto a información adicional para obtener la HS buscada. Este producto se encuentra geolocalizado y se proporciona sobre la grilla Icosahedral Snyder Equal Area Earth fixed (ISEA-4H9) referida como una grilla global discreta (DGG). SMOS cuenta con una resolución espacial de 30-50 km (43 km en promedio), un período de revisita menor a tres días y una precisión para la HS resultante de 0.04 m³/m³ (Kerr et al., 2010a).

La misión Soil Moisture Active Passive (SMAP) puesta en órbita por la Administración Nacional de la Aeronáutica y del Espacio (NASA) combina los instrumentos de microondas pasivas y activas en banda L (1.4 GHz) con el objetivo de mapear la HS en un período de revisita de 2-3 días. Para el análisis propuesto se seleccionó el producto L3 Soil Moisture Passive (SMAP_L3_SM_P) del cual se toman las variables HS y TS. El modelo utilizado para la obtención del producto es el mismo que para SMOS, pero su principal diferencia radica en que la temperatura de brillo posee un ángulo de incidencia fijo (40°). La HS así obtenida se presenta con una precisión de 0.04 m³/m³ (Entekhabi et al., 2010). El radiómetro de exploración de SMAP provee observaciones de huella elíptica, la cual luego es remuestreada en una grilla fija con una resolución espacial 36 km.

Las fechas de adquisición de las imágenes que se procesaron en este trabajo son para el año 2015 el 26/05; 11/07; 24/07; 21/09; 04/10; 02/12; 07/12; 12/12 y para el año 2016 el 04/02; 14/02; 25/02; 01/03; 16/04.

Datos In Situ

Las observaciones de campo utilizadas fueron proporcionadas por la Comisión Nacional de Actividades Espaciales (CONAE). La red se encuentra en los campos cercanos a la ciudad de Monte Buey, donde cada sitio está equipado con sensores HydraProbe II que registran humedad, temperatura y salinidad a una profundidad de 5 cm. Los mismos han sido adecuadamente distribuidos en el área de la zona de estudio con el fin de cubrir la variabilidad espacial de los parámetros superficiales de la zona (Figura 2).

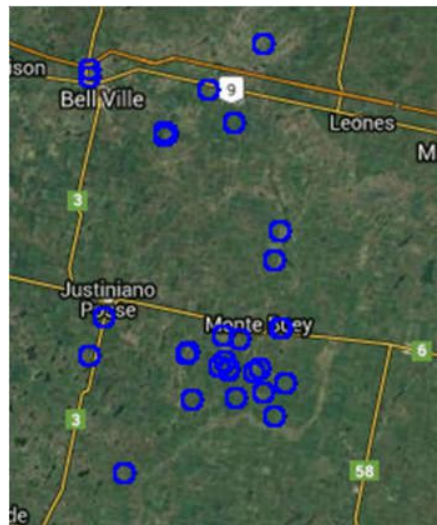


Figura 2: Ubicación de los sensores HydraProbe II en el área de estudio.

Metodo

En un primer análisis se compararon los promedios espaciales de las mediciones de HS y TS registradas en la superficie definida por un pixel con la información remotamente sensada. La bondad de las relaciones entre los estadísticos se cuantifica mediante los estadísticos coeficiente de determinación (R^2), raíz del error cuadrático medio (RMSE) y el sesgo (BIAS).

Para completar la validación se realizó el análisis estadístico Test de Hipótesis. Esta metodología se utiliza ampliamente para comparar dos grupos de observaciones con respecto a una variable numérica. La estructura del Test de Hipótesis se establece mediante una hipótesis nula (H_0), la cual se refiere a cualquier hipótesis que se desea probar. El rechazo de H_0 conduce a la aceptación de una hipótesis alternativa (H_1) que por lo general representa la pregunta o teoría a probar (Walpole et al., 2007). La regla de decisión se establece con anticipación. Luego con el estadístico p-valor se ejecuta la decisión a un nivel de significancia (α) adoptado:

$$\begin{aligned} \text{p-valor} \leq \alpha &\rightarrow \text{rechazar } H_0 \text{ al nivel de significancia } \alpha \\ \text{p-valor} > \alpha &\rightarrow \text{no rechazar } H_0 \text{ al nivel de significancia } \alpha \end{aligned} \quad (1)$$

En este trabajo, el análisis estadístico se aplicó para evaluar la diferencia de medias en las variables de estudio seleccionadas (HS y TS). Se utilizó un nivel de confianza del 95% ($\alpha = 0.05$). Las hipótesis establecidas se basan en la comparación de los datos *in situ* con las observaciones satelitales y se presentaron de la siguiente manera:

$$\begin{aligned} H_0: \mu_{\text{DATOS IN SITU}} &= \mu_{\text{DATOS SATELITALES}} \\ H_1: \mu_{\text{DATOS IN SITU}} &\neq \mu_{\text{DATOS SATELITALES}} \end{aligned} \quad (2)$$

Por otro lado, el segundo análisis consistió en caracterizar la variabilidad espacial de las observaciones de HS y TS para determinar la capacidad de las misiones para captar dicha heterogeneidad. Para este objetivo se constataron los datos *in situ*, puntuales y promediados dentro de un pixel, con las mediciones satelitales provistas por las misiones seleccionadas.

RESULTADOS

Las imágenes fueron pre-procesadas mediante el software de procesamiento ENVI (Environment for Visualizing Images). Se les aplicaron los procesos de geo-codificación (únicamente en la misión SMAP), recorte del área de estudio y reproyección.

Validación De Las Misiones

Los valores diarios de HS y TS *in situ* fueron promediados dentro de un pixel y comparados con las observaciones de las misiones SMOS y SMAP.

Para la variable HS se presentó una diferencia de medias de $0.0944 \text{ m}^3/\text{m}^3$ y $0.0205 \text{ m}^3/\text{m}^3$ entre las mediciones *in situ* y las observaciones de SMOS y SMAP correspondientemente. Se presentó una moderada relación ($R^2=0.30$) entre las estimaciones de SMOS y las mediciones *in situ*. El error encontrado (RMSE= $0.1219 \text{ m}^3/\text{m}^3$) representa el 29% del valor máximo de HS de los sensores de CONAE. El parámetro bias alcanzó el valor de $-0.130 \text{ m}^3/\text{m}^3$. Para la misión SMAP, se observó una mejor relación entre las mediciones ($R^2=0.66$). Se mostró un RMSE de $0.0617 \text{ m}^3/\text{m}^3$ (que representa un 14,4% del valor máximo) y un bias de $-0.020 \text{ m}^3/\text{m}^3$. Los resultados de las validaciones realizadas para HS se observan en la Figura 3.



XVII Simposio Internacional SELPER 2016

Geotecnologías, herramientas para la construcción de una nueva visión del cambio global y su transformación para un futuro sostenible

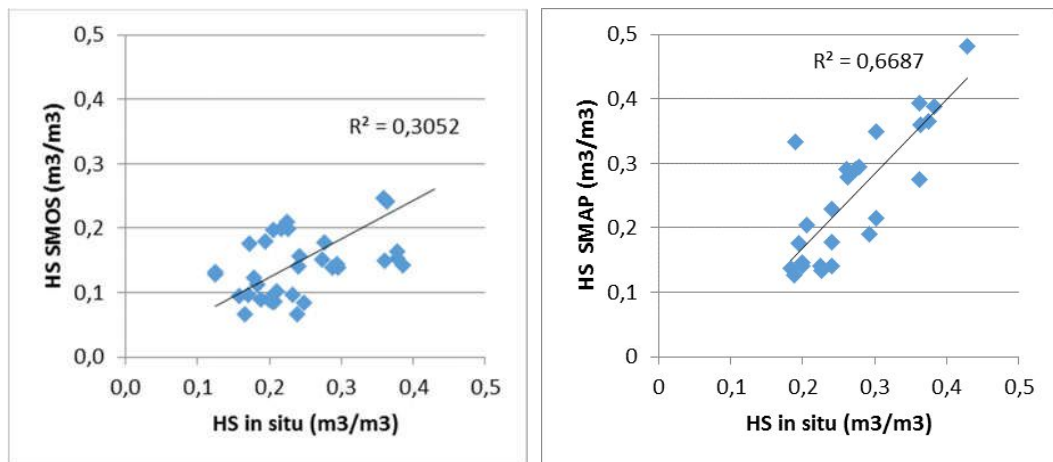


Figura 3: Comparación de los promedios espaciales de HS in situ con información de las misiones SMOS y SMAP respectivamente.

Para complementar la validación de los productos se graficó la evolución temporal de la HS observada *in situ* y los sensores satelitales. Para ello se utilizaron los valores promedios diarios, tal como se observa en la Figura 4.

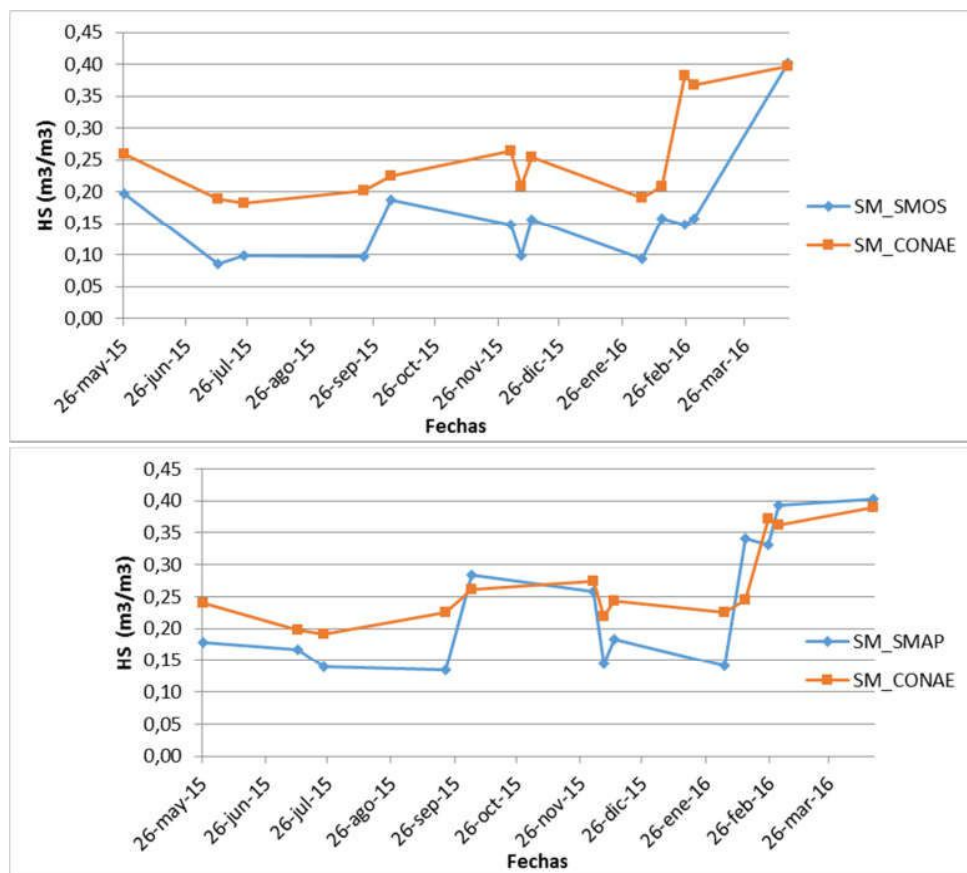


Figura 4: Evolución temporal de la HS registrada por los sensores de las misiones SMOS y SMAP y por la red de sensores terrestres de CONAE.

Para SMAP se realizó conjuntamente el análisis de validación de la variable TS. La diferencia de promedios alcanzó el valor de 1.581 °K. Los resultados mostraron una muy buena correlación para la estimación de TS por la misión, arrojando un R^2 de 0.91 (Figura 5). Se presentó un RMSE de 2.2744 °K que representa el 0.75% del valor máximo. El bias mostró un valor de -1.5805 °K.

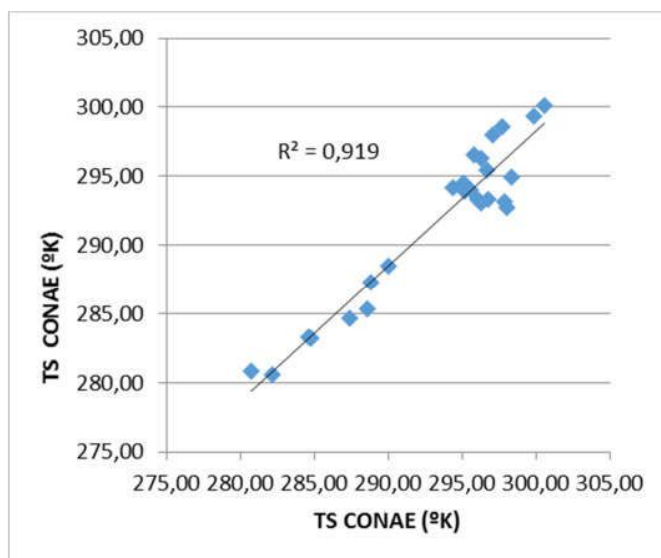


Figura 5: Comparación de los promedios espaciales de TS *in situ* con información de la misión SMAP.

En la Figura 6 se observa la evolución temporal de los promedios diarios de TS para los datos *in situ* y las mediciones de SMAP respectivamente.

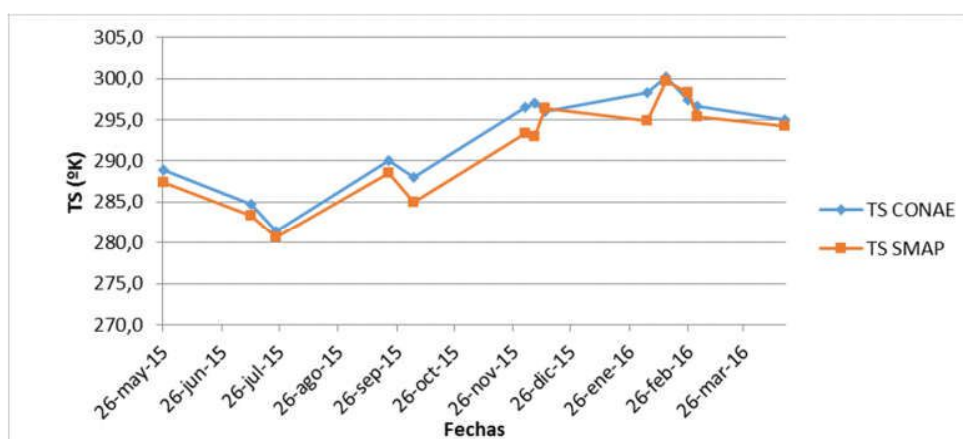


Figura 6: Evolución temporal de la TS registrada por el sensor de la misión SMAP y por la red de sensores terrestres de CONAE.

Por último, se realizó el análisis estadístico de comparación de las mediciones a través de un Test de Hipótesis para la diferencia de medias. Los resultados muestran diferencias de medias para la variable HS entre los datos de campo y los brindados por SMOS. El p-valor muestra un valor de 9.66e-06 en este análisis. Para la misma variable las mediciones *in situ* con las de SMAP no evidencian diferencias de medias, presentando un p-valor de 0.4047. Por último, se planteó el análisis para la TS de SMAP. El p-valor arrojó el resultado de 0.4262 demostrando que no hay diferencia de medias entre los valores estudiados.

Capacidad De Los Radiómetros En Captar Heterogeneidad

Se evaluó gráficamente la variabilidad espacial presente en las mediciones. La relación existente entre los datos *in situ* de los sensores, puntuales y promediados, con las observaciones satelitales de HS y TS, puede observarse en la Figura 7 y 8 respectivamente. En este caso, SMAP presenta mayor variabilidad respecto de SMOS para HS. Si bien la cantidad de datos disponibles para fecha procesada, no es suficiente para un análisis estadístico de rigor, los resultados sugieren que la dispersión de HS en el pixel es significativa. Por otra parte, el gráfico de TS evidencia una moderada variabilidad intra-pixel en esta variable.

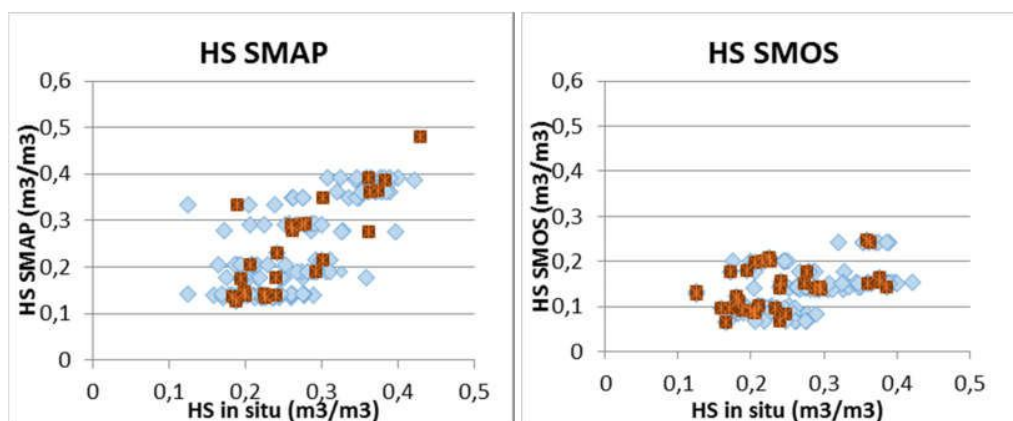


Figura 7: Relación entre los datos *in situ* puntuales (azul) y promediados dentro de un pixel (rojo) con las mediciones satelitales SMOS y SMAP para las variable HS.

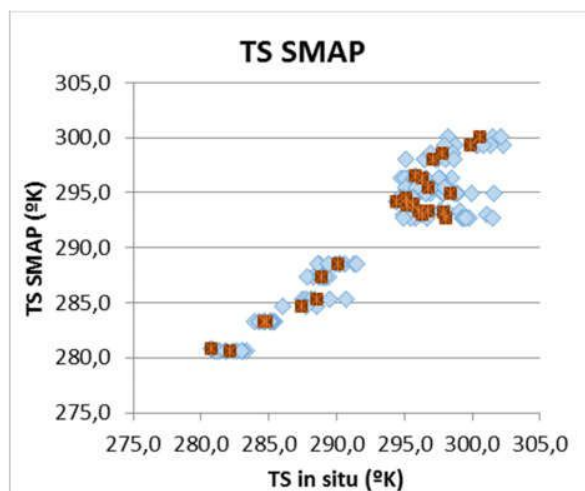


Figura 8: Relación entre los datos *in situ* puntuales (azul) y promediados dentro de un pixel (rojo) con las mediciones satelitales SMOS y SMAP para las variable TS.

CONCLUSIONES

La evaluación de los productos derivados de información satelital es de gran importancia para el entendimiento de los errores presentes en los mismos. En este estudio se propuso validar las observaciones de las misiones SMOS y SMAP comparándolos con datos *in situ* tomados en un área representativa de la Región Pampeana Argentina.

El análisis de HS derivado de SMOS presentó un $R^2=0.3052$, comparable al encontrado por otros autores (Jackson et al., 2012 y Albergel et al., 2012). Sin embargo se encuentra por debajo de la relación lograda por Sanchez et al. (2012) en la cuenca Duero de España ($R^2=0.532$). Dente et al. (2012) obtuvo un valor de $R^2=0.55$ en la región de Maqu, China y en un estudio similar realizado en la región Twente de los Países Bajos obtuvo un $R^2=0.51$. El error encontrado en este estudio se aleja de la precisión planteada por la misión ($0,04 \text{ m}^3/\text{m}^3$). No obstante ello, este error es considerablemente mejor al hallado por otros autores tales como Gherboudjt et al. (2012) que obtuvieron un RMSE de $0.18 \text{ m}^3/\text{m}^3$ en Canadá. El parámetro bias indica que las mediciones de SMOS subestiman la HS presente. Este comportamiento coincide con el ya expuesto en diferentes trabajos para diversas regiones de estudio (Gherboudjt et al., 2012; Albergel et al., 2012; Jackson et al., 2012). La evolución temporal demuestra que las estimaciones de SMOS acompañan las variaciones temporales de las mediciones *in situ*, aunque nunca superan a las mismas.

En este trabajo se observa que la misión SMAP tiene un comportamiento mejor al de la misión SMOS. La validación del producto HS de SMAP resultó sustancialmente superior, ($R^2=0.66687$), al obtenido para SMOS. El RMSE se acercó a la precisión esperada para la misión ($0,04 \text{ m}^3/\text{m}^3$) por el grupo científico de SMAP. El estadístico bias demostró una leve subestimación en la HS. En el análisis temporal de HS también se demuestra la buena relación entre los datos *in situ* y las observaciones de SMAP a lo largo del tiempo. La validación de HS de este estudio presentó resultados similares o mejores a los hallados por Zeng et al. (2016). La variable TS de SMAP también fue validada con los datos *in situ* de la región. Los resultados demuestran una excelente correlación entre ambas mediciones ($R^2=0.919$).

Mediante el análisis de comparación de las medias a través de un Test de hipótesis se completó la validación de las misiones. En el caso de la misión SMAP ambas variables (HS y TS) manifestaron que no hay diferencia significativa entre los datos *in situ* y las estimaciones provistas por la misión. El radiómetro de dicha misión representa correctamente la dinámica espacial de las variables analizadas. Para la misión satelital SMOS se presentan diferencias significativas para la HS al comparar los datos, lo que podría sugerir que los productos de este radiómetro deberían ser fusionados con los de otras fuentes de medición para lograr información con mayor precisión.

Por último, se analizó gráficamente la variabilidad presente en las mediciones. Dicha heterogeneidad puede observarse en la comparación puntual de los datos. En esta relación a cada valor proporcionado por la misión satelital le corresponde una serie de diferentes valores *in situ*, representado en las figuras por un alineamiento de los datos. La relación con los datos promediados dentro de un píxel mejora la correlación de ambas variables, pero no explicaría la variabilidad de las variables HS que se evidencia naturalmente.

Los resultados presentados en este trabajo indicarían que la información provista por estas misiones es capaz de captar y registrar la variación temporal de la HS y TS. No obstante ello, la variabilidad intra-píxel no sería adecuadamente representada por las mismas.

AGRADECIMIENTOS

Los autores desean agradecer a la Comisión Nacional de Actividades Espaciales (CONAE) por brindarnos las observaciones de campo proveniente de su red de sensores.



BIBLIOGRAFIA

- Albergel, C.; De Rosnay, P.; Gruhier, C.; Muñoz-Sabater, J.; Hasenauer, S.; Isaksen, L.; Kerr Y.; Wagner, W. 2012. "Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations". *Remote Sensing of Environment*, Volume 118, Pages 215-226.
- Anderson, M. C.; Norman, J. M.; Kustas, W. P., Houborg, R.; Starks, P. J.; Agam, N. 2008. "A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales". *Remote Sensing of Environment*, Volume 112, Pages 4227-4241.
- Dente, L.; Su, Z.; Wen, J. 2012. "Validation of SMOS Soil Moisture Products over the Maqu and Twente regions". *Sensors Journal*, Volume 12, Pages 9965-9986.
- Entekhabi, D.; Njoku, E.; O'Neill, P.; Kellogg, K.; Crow, W.; Edelstein, W.; Entin, J.; Goodman, S.; Jackson, T.; Johnson, J.; Kimball, J.; Piepmeier, J.; Koster, R.; Martin, N.; McDonald, K.; Moghaddam, M.; Moran, S.; Reichle, R.; Shi, J.; Spencer, M.; Thurman, S.; Tsang, L.; Zyl, J. 2010. "The soil moisture active passive (SMAP) mission". *Proc. IEEE*, Volume 98, Pages 704-716.
- Entin, J. K.; Robock, A.; Vinnikov, K.; Hollinger, S.; Liu, S.; Namkhai, A. 2000. "Temporal and spatial scales of observed soil moisture variations in the extratropics". *J. Geophys. Res.*, Volume 105, Pages 11865-11877.
- Gebremichael, M.; Vivoni, E. 2008. "Spatial sampling uncertainty in SMEX04 soil moisture fields: A data-based resampling experiment," *Remote Sens. Environ.*, Volume 112, Pages 326-336.
- Gherboudj, I.; Magagi, R.; Goita, K.; Berg, A.; Toth, B.; Walker, A. 2012. "Validation of SMOS data over agricultural and boreal forest areas in Canada". *IEEE Transactions on Geoscience and Remote Sensing*, Volume 50, Pages 1623-1635.
- Houser, P.R. 2005. "Infiltration and soil moisture processes". In: Keeley, J. (Ed.), *Water Encyclopedia*. John Wiley & Sons, Inc., New York, Pages 493-506.
- Jackson, T. J.; Chang, A.; Schmugge, T. J. 1981. "Active microwave soil measurements for estimating soil moisture". *Photogramm. Eng. Remote Sens.* Volume 47, Pages 801-805.
- Jackson, T.; Le Vine, D. 1996. "Mapping surface soil moisture using an aircraft-based passive microwave instrument: Algorithm and example". *J. Hydrol.* Volume 184, Pages 85-99.
- Jackson, T. J.; Bindlish, R.; Cosh, M. H.; Zhao, T.; Starks, P. J.; Bosch, D. D., Seyfried, M.; Moran, M.; Goodrich, D.; Kerr, Y.; Leroux, D. 2012. "Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the US". *IEEE Transactions on Geoscience and Remote Sensing*, Volume 50, Pages 1530-1543.
- Kerr, Y.; Njoku, E. 1990. "A semi-empirical model for interpreting microwave emission from semiarid land surfaces as seen from space". *IEEE TGRS*, Volume 28, Pages 384-393.
- Kerr, Y.; Waldteufel, P.; Wigneron, J.; Martinuzzi, J.; Font, J.; Berger, M. 2001. "Soil Moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission". *IEEE T. Geosci. Remote*, Volume 39, Pages 1729-1736.

- Kerr, Y.; Waldteufel, P.; Wigneron, J.; Delwart, S.; Cabot, F.; Boutin, J.; Escorihuela, M.; Font, J., Reul, N.; Gruhier, C.; Juglea, S.; Drinkwater, M.; Hahne, A., Martin-Neira, M.; Meckelenburg, S. 2010a. "The SMOS mission: new tool for monitoring key elements of the global water cycle". Proc. IEEE Volume 98, Pages 666-687.
- Narayan, U.; Njoku. V 2004. "Retrieval of soil moisture from passive and active L/S band sensor (PALS) observations during the Soil Moisture Experiment in 2002 (SMEX02)". Remote Sensing of Environment, Volume 92, Pages 483-496.
- Njoku, E.; Li. L. 1999. "Retrieval of land surface parameters using passive microwave measurements at 6-18 GHz". IEEE Trans. Geosci. Remote Sens., Volume 37, Pages 79-33.
- Njoku, E.; Jackson. T.; Lakshmi, V.; Chan, T.; Nghiem. S. 2003. "Soil moisture retrieval from AMSR-E". IEEE Trans. Geosci. Remote Sens., Volume 41, Pages 215- 229.
- Owe, M.; De Jeu, R.; Holmes. T. 2008. "Multisensor historical climatology of satellite-derived global land surface moisture". J. Geophys. Res., Volume 113, Page 01002.
- Sanchez, N.; Martínez-Fernández, J.; Scaini, A.; Perez-Gutierrez. C. 2012. "Validation of the SMOS L2 soil moisture data in the REMEDHUS network (Spain)". IEEE Transactions on Geoscience and Remote Sensing, Volume 50, Pages 1602-1611.
- Schmugge. T. J. 1983. "Remote sensing of soil moisture: Recent Advances". IEEE T. Geosci. Remote, Volume 21, Pages 145-146.
- Sobrinho, J. A.; Franch, B.; Mattar, C.; Jimenez-Muñoz, J.; Corbari. C. 2012. "A method to estimate soil moisture from Airborne Hyperspectral Scanner (AHS) and ASTER data: application to SEN2FLEX and SEN3EXP campaigns". Remote Sens. Environ., Volume 117, Pages 415-428.
- Ulaby, F.; Batlivala, P.; Dobson. M. 1978. "Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part I-bare soil". IEEE Trans. Geosci. Electro. Volume 16, Pages 286-295.
- Walpole, R. E.; Myers, R. H.; Myers, S. L.; Ye, K. 2007. "Probabilidad y estadística para ingeniería y ciencias (No. TA430. P76 2012)". Pearson Educación.
- Wigneron, J.; Chanzy, A.; Calvet, J.; Bruguier. N. 1995. "A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields". Remote sens. Environ., Volume 51, Pages 331-341.
- Zeng, J.; Chen, K. S.; Bi, H.; Chen. Q. 2016. "A Preliminary Evaluation of the SMAP Radiometer Soil Moisture Product Over United States and Europe Using Ground-Based Measurements". IEEE Transactions on Geoscience and Remote Sensing, Volume 54, Pages 4929-4940.

Anexo E

Soil moisture estimation over flat lands in Argentinean Pampas region using Sentinel-1A data and non-parametric methods - *International Journal of Remote Sensing*

Artículo enviado a la revista *International Journal of Remote Sensing* que se encuentra en la segunda etapa de evaluación.

1

2 Soil moisture estimation over flat lands in Argentinean Pampas 3 region using Sentinel-1A data and non-parametric methods

4 G. García^{a,b}, V. Venturini^b, M. Brogioni^c, E. Walker^{a,b} and L. Rodriguez^b

5 ^aConsejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina;

6 ^bCentro de Estudios Hidro-Ambientales (CENEHA), Facultad de Ingeniería y Ciencias
7 Hídricas (FICH), Universidad Nacional del Litoral (UNL). Ciudad Universitaria. Ruta
8 Nacional N° 168 - Km 472,4. (3000) Santa Fe, Argentina

9 ^c Consiglio Nazionale delle Ricerche (CNR), Istituto di Fisica Applicata “N. Carrara”
10 (IFAC). Florence, Italy.

11 **ARTICLE HISTORY**

12 Compiled July 12, 2018

13 **ABSTRACT**

14 A procedure for soil moisture (SM) estimation over flat lands in the Argentinean
15 Pampas region, using the water balance equation that considers SM to be the result
16 of the water inlets and outflows to the soil system, is presented. In recent years,
17 data remotely sensed with SAR and radiometer sensors have been used to develop
18 different methodologies to obtain SM maps. Thus, a variety of methodologies with
19 different levels of complexity are available nowadays. These models require soil
20 information such as soil physical properties and mineral composition, not readily
21 available in Argentina and many other remote areas of the world. The procedure
22 presented in this paper takes into account water input and output processes of the
23 soil system and represents them with different hydro-environmental variables and
24 SAR data. The water balance equation was solved with multiple linear regression,
25 multivariate adaptive regression and artificial neural network statistical models, fed
26 with readily available data over CONAE core site located in Cordoba province,
27 Argentina. The resulting models were obtained with precipitation, air temperature
28 and relative humidity observations and with SAR data from the Sentinel-1A satellite
29 mission. The error of the model estimates represents 10 % of the observed measured
30 values of SM and is in line with state of the art algorithms. The similar error reached
31 for the models, suggests that any of them can be used, although the MLR method
32 allows analyzing and quantifying the errors introduced by the variables.

33 **KEYWORDS**

34 soil moisture; Sentinel-1A images; multiple linear regression; artificial neural
35 network; multivariate adaptive regression; flat lands

36 **1. Introduction**

37 The hydrological cycle is dominated by processes such as infiltration, runoff,
38 percolation, and evapotranspiration (ET). Soil moisture (SM) is a key state variable
39 in this cycle because it influences the redistribution of water among these processes.
40 SM can be defined as temporary storage within a shallow layer of the earth that is
41 limited to the aeration zone (Engman 1991) and depends on soil depth, texture, and

CONTACT G. García. Email: ggarcia@fich.unl.edu.ar

42 structure. This “storage” may range from about 30-50 mm in shallow sandy soils to
43 more than 200 mm in deep loamy and clay-loamy soils (Verheye 2004).

44 SM estimation is important for hydrological, meteorological and agricultural
45 applications, for instance in the prediction of natural disasters such as floods and
46 droughts. Despite its importance, soil moisture monitoring has proven to be a difficult
47 task due to its spatial and temporal variability. Traditionally, SM observation is carried
48 out by conventional point measurements, which are available at a limited number
49 of sites only. The data collected with this method are usually poorly representative
50 and inefficient, also scheduled field surveys are an unfeasible approach. Moreover,
51 continuous *in situ* observations are expensive and regional sampling is time-consuming
52 because it requires repeated sampling at observation points.

53 A feasible approach to the routinely monitoring of SM is represented by the use of
54 satellite sensors, in particular microwave remote sensing systems which have shown
55 great potential in SM estimation on a regional and global scale. The sensitivity of
56 active (Synthetic Aperture Radar - SAR and scatterometer) and passive (Radiometer)
57 microwave sensors to SM has been successfully demonstrated in several studies
58 (Paloscia et al. 2006; Notarnicola et al. 2006, 2008; Pathe et al. 2009; Mladenova
59 et al. 2014). Indeed microwaves can penetrate the soil, and the thickness of the soil
60 layer sensed strongly depends on their wavelength (other than soil properties): in
61 general longer wavelengths can penetrate deeper than shorter ones. Also, the extinction
62 of electromagnetic waves is smaller for dry soil than for the wet ones, making the
63 penetration depth inversely related to the soil moisture (Behari 2005).

64 At present, SM is estimated operationally by using space borne radiometers and
65 scatterometers at low microwave frequency which has proven to be highly sensitive
66 to SM and quite insensitive to nuisance land characteristics like soil roughness,
67 composition, short vegetation coverage, etc. (Kerr et al. 2001, 2010; Entekhabi et al.
68 2010; Wagner et al. 2013).

69 Radiometers measure the energy emitted by the Earth’s surface in the microwave
70 band. However, the antennas deployed on the spacecrafts are relatively small for the
71 involved wavelength, resulting in a spatial resolution on the order of tens of kilometers.
72 The state of the art radiometers, such as the one of the Soil Moisture and Ocean
73 Salinity (SMOS) mission of European Space Agency (ESA) or the one of the Soil
74 Moisture Active and Passive (SMAP) mission of NASA, have a spatial resolution of
75 approximately 40 km (Kerr et al. 2010; Entekhabi et al. 2010).

76 Amongst active sensors, the European scatterometer (ESCAT) on board the
77 European Remote Sensing (ERS) and Advanced SCATterometer (ASCAT) on board
78 METeorological OPERational (METOP) satellites have proved useful for detecting SM
79 temporal changes (Wagner et al. 1999; Naeimi et al. 2009). The ASCAT Surface
80 Soil Moisture (SSM) product provides daily estimates of surface SM and shows high
81 potential for drought monitoring on a global scale (Zhang et al. 2015). The major
82 disadvantage of spaceborne microwave scatterometers is their low spatial resolution
83 ($\approx 25 - 50$ km), similar to that of a passive microwave sensor. The coarse spatial
84 resolution limits the applicability of both types of sensors on a smaller scale than
85 the watershed scale. In contrast, SAR allows monitoring surface characteristics, such
86 as SM, at spatial resolutions of meters to tens of meters under almost all weather
87 conditions (Pathe et al. 2009; Aubert et al. 2013; Pierdicca et al. 2013; Stamenkovic
88 et al. 2017).

89 In this sense, the technological evolution of SAR over the past ten years allowed the
90 development of several satellite missions to provide information in different microwave
91 frequencies. For instance, ALOS-1/-2 and SMAP at L-band (1.25 GHz - approximately

92 20 cm of wavelength); ERS-1, ERS-2, Envisat, Sentinel-1A, RADARSAT-1 and
93 RADARSAT-2 missions operating at C-band (5 GHz - 5 cm of wavelength);
94 TerraSAR-X and Cosmo-SkyMed missions operate on X-band (9 GHz - about 3 cm
95 of wavelength).

96 The main limitation in obtaining SM from SAR backscattering is related to the
97 inherent complexity of the radar signal (Shi et al. 1997; Loew and Mauser 2006;
98 Mattia et al. 2006). Indeed, backscattering is a nonlinear function of SM, surface
99 roughness; vegetation cover and incidence angle, are among other factors. In particular,
100 for bare or poorly vegetated soils, the received signal is dominated by the SM and
101 roughness. On bare soils, several SM retrieval algorithms were developed based on
102 backscattering models and inversion methods. In general, backscattering models can
103 be classified into theoretical, empirical and semiempirical models (Kornelsen and
104 Coulibaly 2013). Change detection methods, optimization of cost function methods,
105 linear and non-linear interpolation methods, statistical probability methods and
106 computational intelligence methods are examples of the classes above (Kornelsen
107 and Coulibaly 2013). Among computational intelligence models, Multiple Linear
108 Regression (MLR), Artificial Neuronal Network (ANN) and Multivariate Adaptive
109 Regression Splines (MARS) are the most well-known algorithms. The basic idea behind
110 the MLR model is to find the relationship between several independent (“predictor”)
111 variables and a dependent (“criterion”) variable (Weisberg 2005). Simple MLR models
112 have been applied since 1982 by Ulaby et al. (1982) to relate soil physical properties
113 to radar backscattering. After that, other authors described the relationship between
114 SM and radiometer brightness temperature (Narayan et al. 2004; Paloscia et al. 2006;
115 Sobrino et al. 2012; Sánchez et al. 2014). MARS (Friedman 1991) is an extension
116 of the linear regression method that handles nonlinearities and the dependence
117 between datasets. In contrast with MLR, MARS has not widely been used in soil
118 moisture-related studies.

119 Most of the models have been used in areas with moderate to steep slopes, where
120 horizontal water fluxes, such as runoff, dominate SM variability (Forte Lay et al.
121 2008). Ranney et al. (2015) described the influence of porosity, hydraulic conductivity,
122 texture, and soil depth on SM variability, concluding that the impact of soil properties
123 may be greater during wet conditions. These authors studied the Tarrawarra basin in
124 Australia, where horizontal fluxes of water control SM patterns during wet periods,
125 while during dry periods, vertical water fluxes control SM. Studies that estimated SM
126 in flat areas have related SM variability with parameters that characterize the soil,
127 but they have overlooked vertical water flows, such as infiltration and ET, dominant
128 in flat relief regions (Paloscia et al. 2008; Baghdadi et al. 2012).

129 The central region of Argentina is one of the most productive areas of South
130 America where large rivers such as Paraná River and La Plata River are located.
131 The importance of SM estimates in this country is reflected in the objectives of the
132 SACOM Mission (<http://www.conae.gov.ar/index.php/espanol/2012-09-13-18-04-18>)
133 , i.e. “To obtain specific products derived from SAR information, in particular soil
134 moisture maps, which are of great help for agriculture, hydrology and for the health
135 area, due to its proven socio-economic impact”. SM models developed and evaluated in
136 Argentina are rare and require information on the physical properties of the soil and/or
137 its mineral composition and/or vegetation cover (Salgado et al. 2001; Del Valle et al.
138 2008; Grings et al. 2015). Soil properties are scarce in some regions of Argentina while
139 meteorological data are accessible from different government agencies, universities and
140 private agents. This fact suggested that models based on meteorological data are more
141 appropriate than models based only on soil data. It is also well known that in Argentina

142 the acquisition of SAR images is difficult due to scarce funding sources, so that any free
143 SAR image has a significant added value in the country. The Copernicus program, with
144 its Sentinel satellites, managed by ESA offers a unique opportunity to have free SAR
145 images for the scientific community, with the limitation that these are only distributed
146 in simple polarization VV or in the best case VV-VH in our region.

147 The inversion algorithms are more focused on exploiting the backscattering
148 information (i.e. intensity and texture) than on considering punctual data as
149 meteorological variables. Nevertheless, the water input/output fluxes to the soil system
150 should be considered so as to explain SM variability. In fact, water balance models
151 relate SM variability to hydro-meteorological processes by non-linear relationships
152 (Alley 1984; Liang et al. 1994; Brocca et al. 2014). Thus, this paper aims at developing
153 a procedure to estimate SM using a water balance model incorporating water flux
154 variables from readily available data in remote areas. The balance equation was solved
155 using the statistical models MLR, MARS, and ANN. These algorithms use as inputs
156 SAR backscattering and meteorological data, and are run over flat land with sparse
157 vegetation or bare soil. The simple nonparametric techniques suggested in this paper
158 require similar input data such as hydrological and agricultural assessment methods,
159 making them appealing to hydrologists and agronomists operating in regions where
160 the soil is poorly characterized. Finally, the MLR and MARS methods' performances
161 are evaluated by contrasting the results to those of the ANN.

162 This paper is organized into six sections. The next section presents a description
163 of the procedure applied for estimating SM. Section 3 introduces the study area,
164 available ground observations and satellite data used. Section 4 shows the preliminary
165 SM estimates achieved with the proposed methodology. Section 5 presents a discussion
166 of results. Finally, in section 6 the main conclusions are outlined.

167 **2. Methodology**

168 In the previous section the need to quantify the water inlets and outlets of the
169 soil system has been presented through a water balance model accounting for all
170 of the processes that influence SM variability. The water balance model used and its
171 relationship with the SM will be described hereinafter. Then the statistical methods
172 proposed to estimate the SM will be detailed.

173 **2.1. Water balance model**

174 Water balance models express soil moisture temporal variation as a superposition of
175 precipitation (PP), ET and flow rates as shown in equation 1 (Orth et al. 2013).

$$\frac{\Delta SM}{\Delta t} = PP - ET - Q \quad (1)$$

176 where Q represents surface runoff. All variables are expressed in mm/day (or in an
177 equivalent unit).

178 The relationships between each process and SM may take on different forms, but
179 the authors agree upon the fact that they are nonlinear of the form of equation 2 (Orth
180 et al. 2013; Brocca et al. 2014).

$$Z = \alpha(SM)^\gamma \quad (2)$$

181 where Z represents either ET or runoff. Thus,

$$\log(Z) \approx \gamma \log(SM) \quad (3)$$

182 On the other hand, different methods have been developed to calculate ET under
183 different conditions of surface humidity and spatio-temporal scales. Among the most
184 used methods are Thornthwaite (Thornthwaite 1948), Penman-Monteith (Penman
185 1948) and Priestley and Taylor (Priestley and Taylor 1972). These methodologies
186 require data such as air temperature (T_a), air water vapor pressure (e_a), net radiation
187 (R_{net}), surface temperature (T_s) and relative humidity (RH) (Venturini 2007; Orth
188 et al. 2013). Besides, surface runoff varies with surface roughness, soil properties,
189 and PP (Orth et al. 2013). Consequently, the water balance equation expresses SM
190 variability as a superposition of nonlinear functions that depend on meteorological
191 variables and surface properties.

192 This allows one to deduce that the logarithm of the variables involved in ET are
193 related to the second member of equation 3. Also, Q varies with the surface roughness,
194 the properties of the soil and the PP , so these variables express the relationship
195 between Q and SM (Orth et al. 2013). Therefore, whatever the equations used to
196 estimate ET and Q , all lead to a water balance where the variability of the SM is
197 expressed as a superposition of nonlinear functions that involve the variables above.

198 It is assumed that there is no change in SM during the time of acquisition of the
199 satellite products to solve the water balance equation. Also, the water balance is
200 solved within a pixel, where it is assumed that the horizontal movements of water,
201 runoff, are cancelled within a pixel, and only the vertical movements of water prevail.
202 Thus, the simplified balance equation will be solved by using statistical methods
203 that allow one to establish the relationships between the SM and the variables that
204 describe ET and Q . The reason for this choice is that the definition of the model is
205 simpler with statistical approaches than with numerical methods, where the calibration
206 of physically significant parameters can be complex for large areas (Kornelsen and
207 Coulibaly 2013). The hydrologic models found in the literature to estimate the water
208 balance between the income of water from PP and snow melt and the outflow of water
209 by ET , streamflow, and ground water re-charge are the models such as Soil and Water
210 Assessment Tool (SWAT), Variable Infiltration Capacity (VIC), among others (Kang
211 and Sridhar 2018). These models differ in the treatment of the ET , the accounting of
212 SM and the recharge of the aquifer.

213 The statistical methods proposed in this paper are MLR, MARS, and ANN. MLR
214 is a well-known method that seeks the coefficients that relate all input variables to
215 the unknown variable as a simple linear combination. Although this approach is not
216 innovative, it is more flexible than numerical methods (which are not more recent than
217 linear regression) to solve this type of relationship and does not require expert users.
218 On the other hand, ANNs offer a more precise way of solving non-linear equations
219 with complicated rules (Ali et al. 2015).

220 **2.2. Regression methods**

221 *2.2.1. Multiple Linear Regression (MLR)*

222 A MLR model computes the dependent variable Y using a linear combination of
223 independent variables X_k . The linear regression has the following form (Weisberg

224 2005):

$$Y_i = \beta_0 + \sum_{k=1}^n \beta_k X_{k,i} + \epsilon \quad (4)$$

225 where β_0 , $\beta_k(k > 0)$, and ϵ are constant terms representing the intersection, the
226 regression coefficients, and the model error, respectively.

227 To estimate β_j , the least square approach is followed, to find the minimum of the
228 sum (S) of the squared residuals defined in equation 5:

$$S = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i,1} - \dots - \beta_p X_{i,p})^2 \quad (5)$$

229 The regression model requires that input variables (X_i) explain the output variable
230 (Y), and furthermore that each input variable adds different information, i.e., they
231 should not be collinear. For input variable selection, the backward elimination
232 methodology is applied using p-value statistics to determine the significance of each
233 variable (Weisberg 2005). Subsequently, once a set of input variables representative
234 of the behavior of the output variable is selected, the correlation between them is
235 analyzed. This task is performed to investigate variable linearity by resorting to the
236 Variance Inflation Factor (VIF) (Harrell 2001; Fox 2002). VIF values greater than
237 four suggest the existence of multicollinearity, VIF values greater than ten indicate
238 the presence of severe multicollinearity (Montgomery et al. 1992; Fox 2002; Cohen et al.
239 2003). Other authors suggest that the VIF limit is between five and ten (Marquardt
240 1970). Since both statistics are relevant for the model variable selection, in this work a
241 combination of p-value and VIF criteria is applied. Thus, variables with large p-value
242 and which are collinear (VIF greater than 5) are first discarded.

243 The size of the sample needed to calibrate the model can be optimized by the
244 learning curve method. In turn, this optimization procedure will improve the model
245 performance. The learning curve is obtained by splitting the input dataset into two
246 sets and computing the model performance with each of them (Beleites et al. 2013).
247 Different set sizes, varying from 10 % to 90% for the calibration dataset are used to
248 generate the curve. Similar errors are encountered with both set sizes when sufficient
249 amount of data for calibration have been employed.

250 After the variable selection, the calibration of the model is done, and the coefficients
251 β_i computed for each variable.

252 2.2.2. Multivariate adaptive regression splines (MARS)

253 MARS is a procedure for fitting adaptive non-linear regression that uses piecewise
254 basis functions (Bfs) to define relationships between a response variable and some set
255 of predictors (Friedman 1991).

256 The MARS algorithm consists of two phases of forward and backward stepwise
257 procedures. The forward phase places candidate knots at random positions within
258 the range of each predictor variable to define a pair of Bfs. At each step, the model
259 adapts the knot and its corresponding pair of Bfs to give the maximum reduction in
260 the error. This process of adding Bfs continues until the maximum number is reached.
261 The developed model in the forward phase will overfit with the training data and
262 therefore require a backward stepwise selection to eliminate ineffective Bfs. The least

263 effective Bfs in the mean square sense are eliminated until the change in prediction
 264 error is small (Afshar and Yilmaz 2017). The backward phase, in fact, prunes the
 265 model to create a more generalized model with better abilities.

266 The general MARS function can be represented by using the following equation
 267 (Friedman 1991):

$$\hat{f}(x) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{k=1}^{k_m} [S_{km}(x_{v(k,m)} - t_{km})] \quad (6)$$

268 where β_0 and β_m are fitting parameters, M is the number of Bfs, K_m is the number
 269 of knots, S_{km} takes on values of either 1 or -1 and indicates the right/left sense of
 270 the associated step function, $v(k, m)$ is the label of the independent variable, and t_{km}
 271 indicates the knot location.

272 For sake of simplicity, only the piecewise linear function is expressed and considered
 273 in this paper. Piecewise linear functions follow the form $Max(0, x - t)$ with a knot
 274 defined at value t . Formally,

$$Max(0, x - t) = \begin{cases} x - t & \text{if } x \geq t \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

275 **2.3. Artificial Neural Networks (ANNs)**

276 ANNs are statistical models set up from a set of training data mimicking the learning
 277 capacity of the human being. They are resistant to noisy data and can approximate
 278 multivariate nonlinear relationships between the variables (Ahmad et al. 2010). A
 279 commonly used type of ANN is the Multi-Layer Perceptron (MLP). MLP can be
 280 trained to represent arbitrary input-output relationships (Hornik et al. 1989).

281 The architecture of the MLP is characterized by neurons grouped in layers of
 282 different levels. Three different types of layers are distinguished: the input layer, the
 283 hidden layers, and the output layer. Figure 1 depicts the architecture of the MLP
 284 neural network.

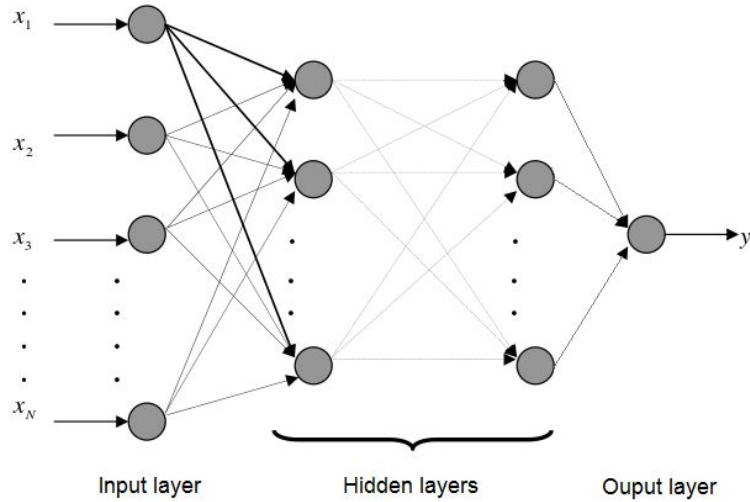


Figure 1.: Architecture of the MLP neural network model used in the prediction of SM

285 The MLP model can be mathematically formulated as:

$$\hat{Y} \approx F \left(\sum_{j=1}^m w_{kj} F \left(\sum_{i=1}^n w_{ji} X_i + B_j \right) + B_k \right) \quad (8)$$

286 where w_{kj} are weights between hidden and output layers; w_{ji} are weights between
 287 input and hidden layers; X_i are input variables; m is the number of neurons in a
 288 hidden layer; n is the number of neurons in an input layer, B_j and B_k are the bias
 289 values of the neurons in the hidden and output layers, respectively; F is the transfer
 290 function; and \hat{Y} is the output.

291 The MLP is trained by a large set of training samples based on a supervised learning
 292 procedure. The basic learning method for the MLP is called backpropagation or
 293 backward propagation of errors (Rumelhart et al. 1986). The net training process with
 294 this algorithm involves three phases: the forward calculation with the input patterns,
 295 the calculation of the error committed and the backward propagation of that error
 296 with the corresponding adjustment of the weights of the model.

297 The inputs must be normalized to ensure proper operation of the backpropagation
 298 algorithm. Furthermore, in order to accelerate the learning process, the entries should
 299 not be correlated and should be scaled so that their covariances are approximately
 300 equal, thus ensuring that the different synaptic weights in the network learn at
 301 approximately the same speed (Haykin 2009).

302 2.4. Model validation

303 After the retrieval models have been calibrated, the next step is to evaluate their
 304 performances. This operation entails applying the methodology to a data set that was
 305 not used in the definition of the models. Then, the estimated SM values are compared
 306 with observed SM values.

307 To quantify the methodology accuracy, the determination coefficient (R^2), the root

308 mean square error (RMSE), and the bias are computed. R^2 describes the proportion
 309 of the total variance in the observed data that can be explained by the model
 310 (Legates and McCabe Jr. 1999), RMSE quantifies the difference between observed and
 311 modeled values (Sousa et al. 2007), and the bias indicates whether observed values are
 312 overestimated or underestimated (Albergaria et al. 2014).

313 3. Study area and data set

314 3.1. Study area

315 The study area is located in the southeast of Córdoba province, Argentina, around a
 316 small village named Monte Buey (-32° 55', -62° 27') (Figure 2). Part of this province
 317 is within the Pampas region, an area of approximately 840000 km² characterized by
 318 slopes lower than 0.5% (Bedano et al. 2016).

319 It is a flat, gently sloping area. During normal water years, vertical water flows
 320 predominate over horizontal ones. The region has a continental climate with an average
 321 annual temperature of 17.4 °C. Monthly mean temperature ranges from 24.3 °C in
 322 January to 10.5 °C in July. Average annual *PP* is 920 mm (1960-2010), mostly
 323 accumulated from September to April (Thibeault et al. 2015). According to Duval
 324 et al. (2016) the soil is classified as typic Argiudoll. The Monte Buey area is mainly
 325 characterized by agricultural land and dairy farms and, to a lesser extent, livestock.
 326 Predominant crops are soybean and corn. Sometimes, these crops are replaced by
 327 sorghum, alfalfa, and pastures (Thibeault et al. 2015).

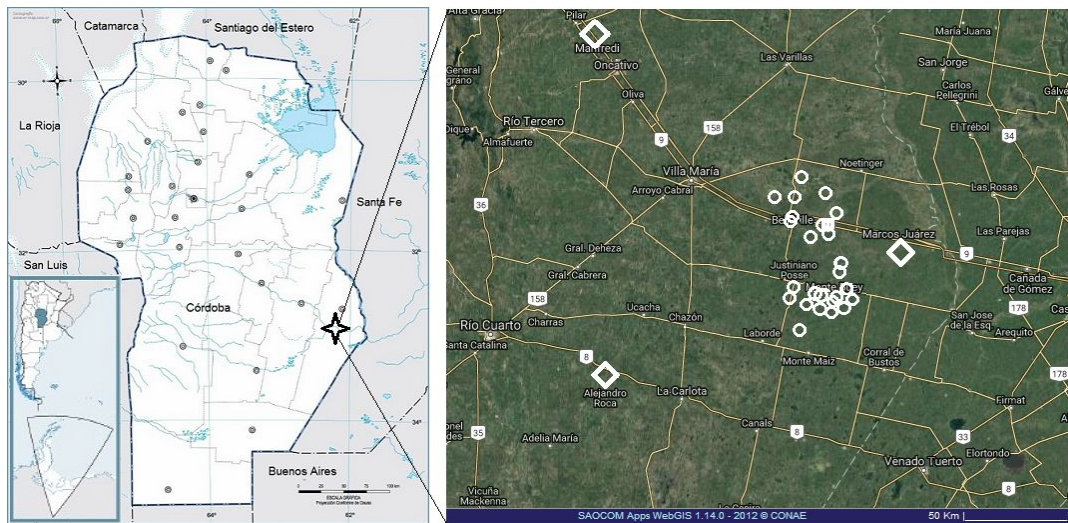


Figure 2.: Left: Location of study area; Right: *in situ* measurements locations; meteorological data (diamond); SM data (circles)

328 3.2. Dataset

329 The dataset used in this paper spans over the period between December 2014 and
 330 April 2016. It is composed of: optical satellite data, SAR data, *in situ* SM and T_s ,
 331 meteorological data.

332 Areas without vegetation or with sparse vegetation are identified by using optical

333 images. To carry out this task NDVI data derived from the MODIS sensor on board
 334 of the Terra satellite mission (EOS AM-1) was used. MOD13Q1 the MODIS product
 335 provides NDVI data at 250 m spatial resolution every sixteen days. This product
 336 is readily available; unlike products from other missions that have a better spatial
 337 resolution but a much larger revisit time.

338 Sentinel-1A mission has a SAR sensor on board that works at C band (~ 5 cm
 339 wavelength). This mission provides free Single Look Complex (SLC) and Ground
 340 Range Detected (GRD) products, both of them in Interferometric Wide Swath mode
 341 at VV polarization (this is the acquisition mode planned by ESA over the study
 342 region). The images are processed by using the Sentinel Application Platform (SNAP)
 343 tool provided by ESA. SLC products are processed by performing the multilooking
 344 which reduces *Speckle* noise and modifies its original spatial resolution to 30 m; then
 345 they are calibrated and georeferenced using the WSG84 geographic system and a
 346 digital elevation model from the Shuttle Radar Topography Mission, with a 90 m flat
 347 resolution. Finally, SAR data are converted to decibels. GRD products are processed
 348 with the same software performing the following operations: multilooking, calibration;
 349 geocoding and dB conversion. Details on SAR and optical satellite data, such as
 350 acquisition date and product type, available for this work are provided in Table 1.

Table 1.: Inventory of optical and SAR images.

Terra-MOD13Q1	Sentinel-1A			
Date (dd/mm/yy)	Date (dd/mm/yy)	Product	Orbit	Pass
19/12/2014	21/12/2014	GRD	76	Ascending
17/01/2015	14/01/2015	GRD	76	Ascending
02/02/2015	07/02/2015	GRD	76	Ascending
22/03/2015	15/03/2015	GRD	76	Ascending
07/04/2015	08/04/2015	GRD	76	Ascending
23/04/2015	18/04/2015	GRD	39	Descending
09/05/2015	02/05/2015	GRD	76	Ascending
09/05/2015	12/05/2015	GRD	39	Descending
25/05/2015	26/05/2015	SLC	76	Ascending
10/06/2015	05/06/2015	GRD	39	Descending
26/06/2015	19/06/2015	SLC	76	Ascending
28/07/2015	23/07/2015	SLC	39	Descending
13/08/2015	16/08/2015	GRD	39	Descending
29/08/2015	30/08/2015	SLC	76	Ascending
14/09/2015	09/09/2015	GRD	39	Descending
30/09/2015	23/09/2015	SLC	76	Ascending
16/10/2015	27/10/2015	GRD	39	Descending
17/11/2015	10/11/2015	GRD	76	Ascending
17/11/2015	20/11/2015	SLC	39	Descending
19/12/2015	14/12/2015	SLC	39	Descending
17/01/2016	07/01/2016	GRD	39	Descending
17/01/2016	21/01/2016	SLC	76	Ascending
02/02/2016	31/01/2016	GRD	39	Descending
18/02/2016	14/02/2016	SLC	76	Ascending
18/02/2016	24/02/2016	GRD	39	Descending
05/03/2016	09/03/2016	SLC	76	Ascending
21/03/2016	19/03/2016	GRD	39	Descending
06/04/2016	02/04/2016	SLC	76	Ascending
06/04/2016	12/04/2016	GRD	39	Descending
22/04/2016	26/04/2016	GRD	76	Ascending
08/05/2016	20/05/2016	SLC	76	Ascending

351 *In situ* gravimetric SM and T_s were provided by Comisión Nacional de Actividades
 352 Espaciales (CONAE). CONAE operates a sensor network implemented to assist in the
 353 process of calibrating SM maps produced by current missions, such as SMAP and a
 354 future Argentine mission named “Satélite Argentino de Observación Con Microondas”

355 (SAOCOM). The network is located in the fields near to Monte Buey city, where each
 356 site is equipped with HydraProbe II sensors (Stevens Water Monitoring System, Inc.)
 357 installed for continuous moisture, temperature and salinity of the soil monitoring at
 358 a depth of 5 cm (Thibeault et al. 2015). The stations have been properly distributed
 359 in the study area as shown in Figure 2 (circles) with the aim of covering the spatial
 360 variability of the surface parameters of the area.

361 Meteorological data are available from weather stations run by The Instituto
 362 Nacional de Tecnología Agropecuaria (INTA). Stations near the study area are:
 363 Alejandro Roca, Manfredi, and Marcos Juárez all in Córdoba province marked with
 364 diamonds in Figure 2. The data provided by INTA stations include T_a , e_a , PP and
 365 RH . Cumulated seven-day PP and seven-day average T_a , e_a , and RH are computed
 366 as model input variables. These point data are interpolated using the Inverse Distance
 367 Weighting (IDW) method to obtain a map of each variable in the study area.

368 Finally, SM , T_s , T_a , e_a , PP , RH and σ^0 for Sentinel-1A dates shown in Table 1 are
 369 available for the calibration of the statistical models.

370 3.3. Pre-processing data

371 A dataset of satellite, meteorological, and *in situ* data is used as input to the models.
 372 The fact that data are generated with different technologies, with distinct spatial and
 373 temporal resolutions, is not a constraint on development of the models. Scaling and
 374 interpolation techniques are applied to obtain compatible satellite and meteorological
 375 data. The different sources of information with their spatial resolutions and scaling
 376 operations are presented in Figure 3.

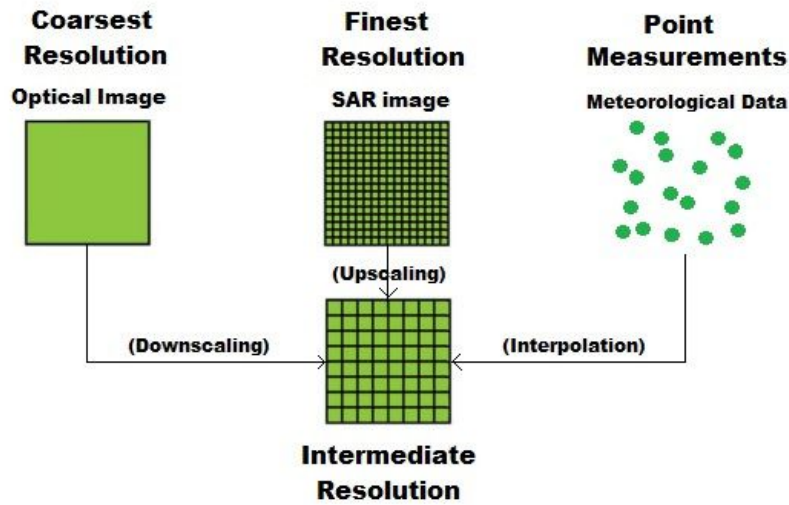


Figure 3.: Data pre-processing to integrate different spatial resolutions

377 Optical satellite data usually have a coarser spatial resolution than SAR data.
 378 Therefore, to be able to overlap both sources of data, the spatial resolution of one
 379 of them, or both, must be modified. In this work, the spatial resolution of optical data
 380 was downsampled by using a 2-D discrete wavelet transform, implemented through an
 381 extension of the pyramid algorithm proposed by Mallat (Mallat 1989). This algorithm
 382 consists of an up-sampling operator along with filtering operations using 1-D low-pass
 383 and high-pass filters over the rows and columns of the image. In this paper, MODIS

384 NDVI image original spatial resolution (250 m) was downscaled to approximately 30
385 m. Different wavelet families (i.e., *Daubechies*, *Coiflet*, *Symlets*, and *Haar*) with order
386 two and four were used and the errors of the resulting downscaled images analyzed.
387 The NDVI downscaled images were contrasted with NDVI images from Landsat 8
388 satellite. The smaller RMSE, approximately 18 % of the maximum value of NDVI,
389 was achieved with *Coiflets* family of order two. On the other hand, in this work, the
390 MODIS NDVI product, with the closer acquisition date to that of Sentinel-1, has
391 been used. Then, SAR images spatial resolution was upscaled, applying multilooking
392 operation at 30 m.

393 Meteorological data are obtained from automatic weather stations. Therefore, these
394 point observations must be interpolated to obtain a spatial distribution over the study
395 area to create regional maps. To this aim, the Inverse Distance Weighting (IDW)
396 interpolation method is used (Watson and Philip 1985). IDW predicted a data value
397 for a given location using data collected in the surroundings and assuming that each
398 measured location has a local effect that diminishes with distance. Consequently, the
399 closest observed values to the prediction location have more influence on the predicted
400 value than those farther away.

401 Once the spatial resolutions of the input variables were unified, maps of T_a , e_a , PP ,
402 RH , σ^0 , T_s , and NDVI were stacked, and the values of each variable at the CONAE
403 *in situ* stations were extracted. Then, the processed data were organized in an input
404 matrix where each row is a set of input variables related to a single SM observation,
405 and each column is a distinct variable (T_a , e_a , PP , RH , σ^0 , T_s and NDVI), as shown
406 in equation 9,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_i \\ \vdots \\ Y_n \end{bmatrix} = \mathcal{S} \left\{ f \left(\begin{bmatrix} X_{1,1} \\ X_{2,1} \\ X_{i,1} \\ \vdots \\ X_{n,1} \end{bmatrix} \right), g \left(\begin{bmatrix} X_{1,2} \\ X_{2,2} \\ X_{i,2} \\ \vdots \\ X_{n,2} \end{bmatrix} \right), \dots \left(\begin{bmatrix} \dots \\ \dots \\ \dots \\ \vdots \\ \dots \end{bmatrix} \right), k \left(\begin{bmatrix} X_{1,m} \\ X_{2,m} \\ X_{i,m} \\ \vdots \\ X_{n,m} \end{bmatrix} \right) \right\} \quad (9)$$

407 where \mathcal{S} is the retrieval algorithm, Y_i represents the i -th soil moisture sample, $X_{i,j}$
408 represents the j -th input variable for the i -th sample, and they are related by the
409 functions f , g , h , i , j and k . In this work, the resulting \mathbf{X} matrix has 248 samples and
410 seven columns.

411 Then, a data scrubbing operation is applied to the input matrix. This process
412 consisted in eliminating the row of the matrix when any variable in it is outside of
413 the respective range. The variability range of SM was defined by analyzing the field
414 data. The range of SM observed in the study area by CONAE varied from 5 to 45
415 %Vol.; therefore, this was the valid range adopted for simulation (Thibeault et al.
416 2015). An analysis of the variation of σ^0 for different land covers was used to define
417 the valid range σ^0 , i.e., -18 and -4 dB (Liu et al. 2016). The criterion used to eliminate
418 the outliers in the other variables was based on the percentile statistic, which allows
419 selection of the most representative data block of each variable (Helsel and Hirsch
420 2002). Here, the scrubbing mining was performed using the 95 % percentile of T_a , e_a ,
421 PP , RH and T_s . After applying this operation, the number of samples is reduced to
422 184.

423 Figure 4 shows the temporal variation of the variable, where each dot represents
424 a spatially averaged value over the whole study area. PP (Cumulated seven-day

425 precipitation) time series shows two seasons: a wet season from September to April
426 and a dry season from May through August. A similar seasonality is also observed in
427 SM series, while T_a , T_s , and e_a present an annual cycle and they seem to be correlated
428 to each another.

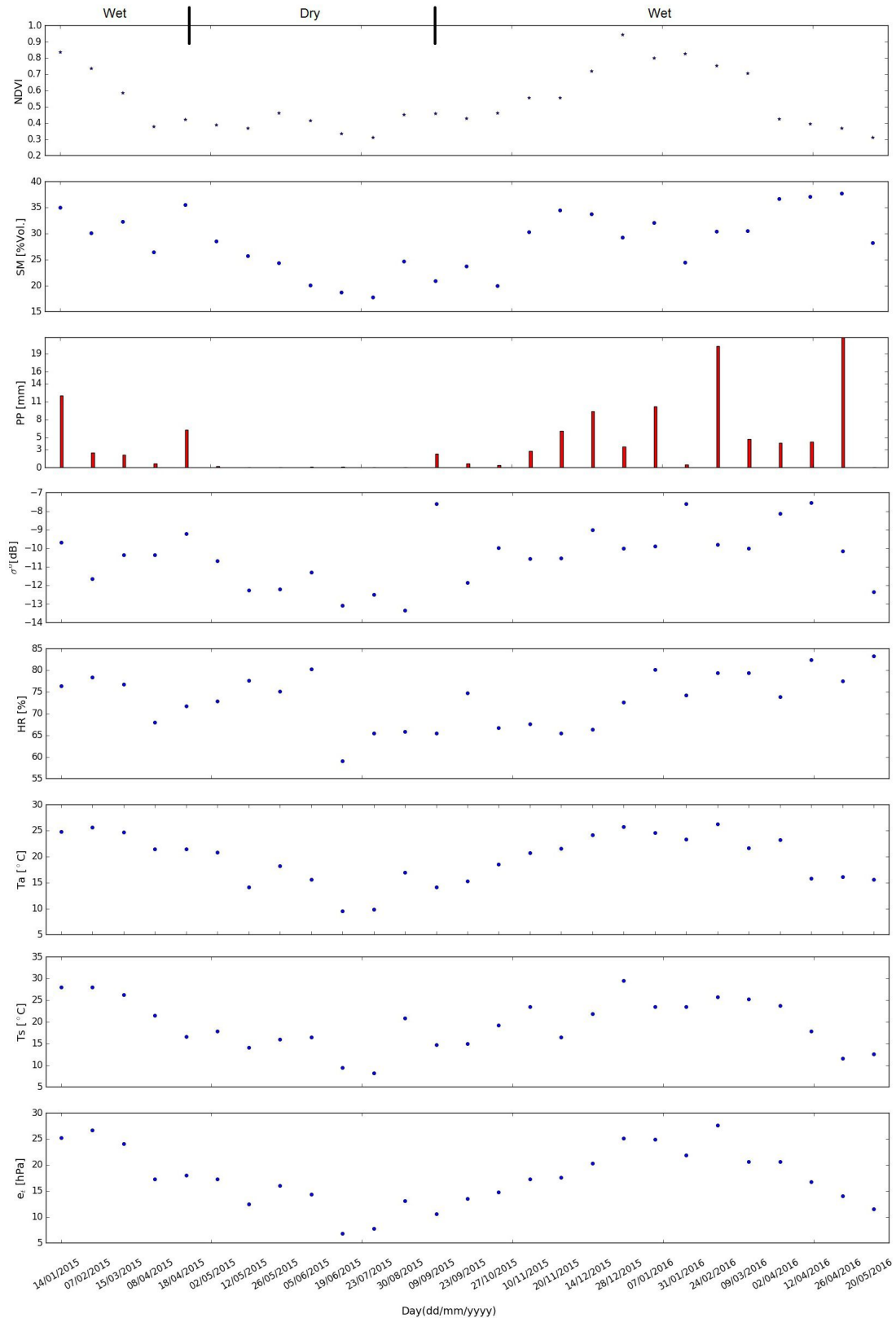


Figure 4.: Time variation of each regional averaged variable

429 The reduced matrix is then filtered by NDVI in order to select data collected on
430 bare and sparsely vegetated soils ($0 < \text{NDVI} < 0.5$). The number of samples in the final
431 input \mathbf{X} matrix for calibrating and validating the selected methods is 111.

432 All the proposed methodologies require statistically independent variables. Hence,
433 the mutual dependency of the input variables was established using the Pearson
434 correlation (r) coefficient, as shown in Figure 5. This Figure suggests that there is
435 a linear relationship (so-called co-linearity) between T_s and T_a with a r greater than
436 0.8. This finding is supported by Lin et al. (2016), who showed that T_s increases with
437 increasing T_a . On the other hand, e_a is related to T_a by an exponential function (Buck
438 1981), and a similar relation is observed in the scatter plot between both variables,
439 with r greater than 0.9. Therefore, e_a is also related to T_s , the r coefficient obtained
440 is equal to 0.83. The variable σ^0 did not show any clear linear relationship with other
441 variables. The scatter plot between σ^0 and SM does not suggest a linear trend between
442 both variables ($r = 0.3$), even though SM increases with increasing σ^0 , indicating there
443 is a non-linear relationship. It should be noted that σ^0 presented very low values of r ,
444 while T_s has a smaller value of r .

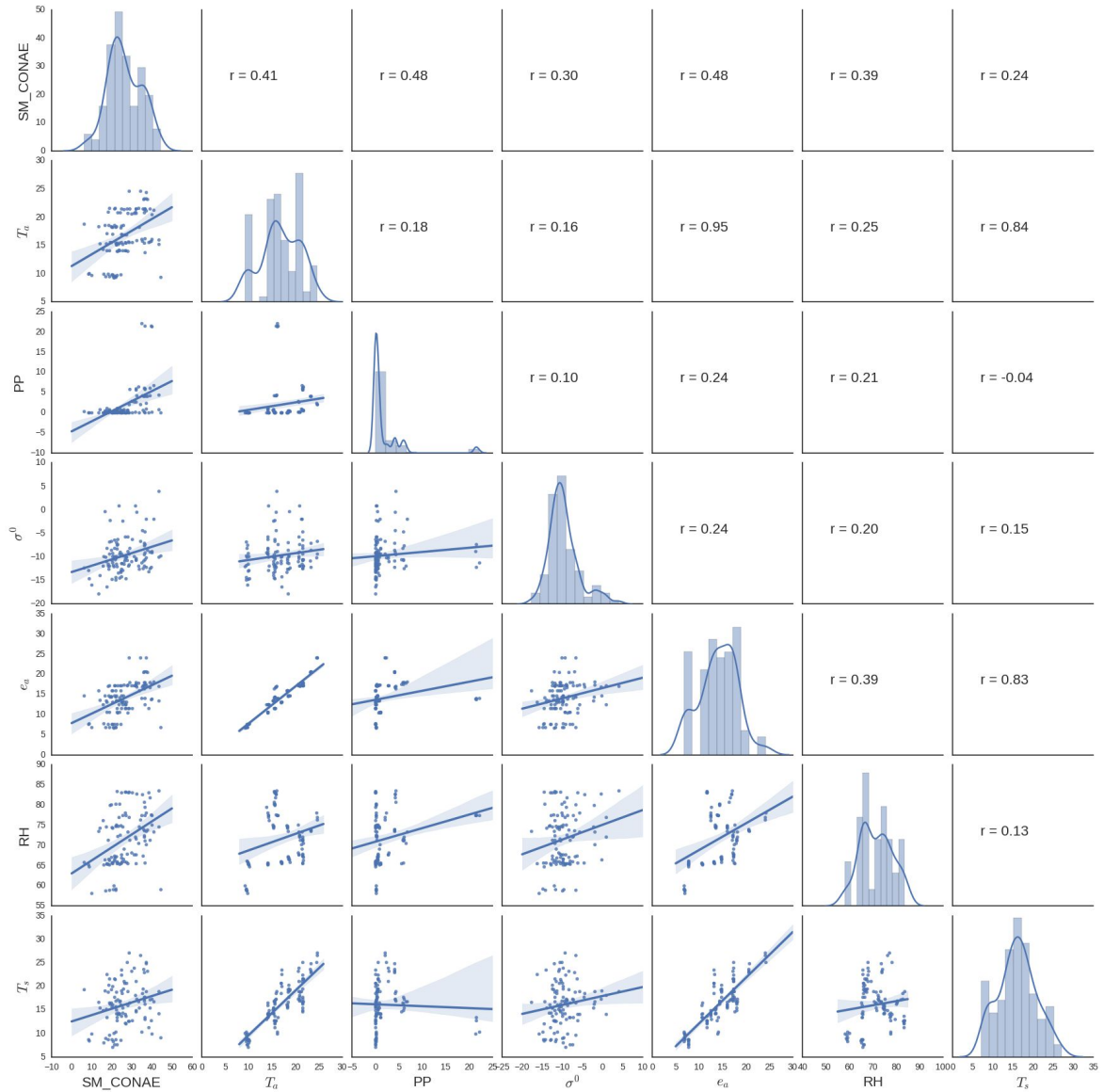


Figure 5.: Correlation matrix between SM CONAE, T_a , PP, σ^0 , e_a , RH and T_s .

445 Table 2 shows the statistic variables computed with samples plotted in Figure 5.

Table 2.: Input statistic variables, minimum (Min), maximum (Max), mean and standard deviation (S)

Variable	Units	Min	Max	Mean	S
SM	%Vol.	6.25	43.36	26.26	8
σ^0	dB	-17.82	-4.53	-10.78	3
T_a	$^{\circ}C$	9.25	24.7	17	4.15
T_s	$^{\circ}C$	7	28	16.28	4.75
e_a	hPa	6.76	24.2	14.26	4
PP	mm	0	22	1.96	4.23
RH	%	58	83.63	71.74	6.64
NDVI		0.24	0.5	0.37	0.07

446 Given that the vegetation roughness affects the backscattered signal, the vegetation

447 phenology limits the model application, thus in this work the data-set was filtered to
448 represent bare soil and sparse vegetation. Figure 6 shows the vegetation cover in the
449 study area, during the four seasons, i.e., winter, spring, summer, and autumn. Each
450 NDVI map has been classified using four classes: water, bare soil, sparse vegetation
451 and dense vegetation. This classification was achieved by using threshold on the NDVI
452 values for each cluster. Hence, areas with sparse vegetation has moderate NDVI values
453 of approximately 0.2 - 0.5. In bare soil areas NDVI takes values close to zero (0 to
454 0.2), and for water bodies, NDVI presents negative values (Holben 1986; Lillesand and
455 Kiefer 1994).

456 It should be noted that during winter time (Figure 6-a) dense vegetation covered
457 only 10.6 % of the area, while pixels with NDVI greater than 0.5 increased from 28.3
458 % in spring to 99 % in summer (Figure 6-b and 6-c). Then, by the end of summer
459 (Figure 6-d) dense vegetation dropped to 77 %. Based on vegetation cover variation
460 presented in Figure 6, the proposed methodology can be applied to the study area
461 during the seeding season and rainy months (from April to September).

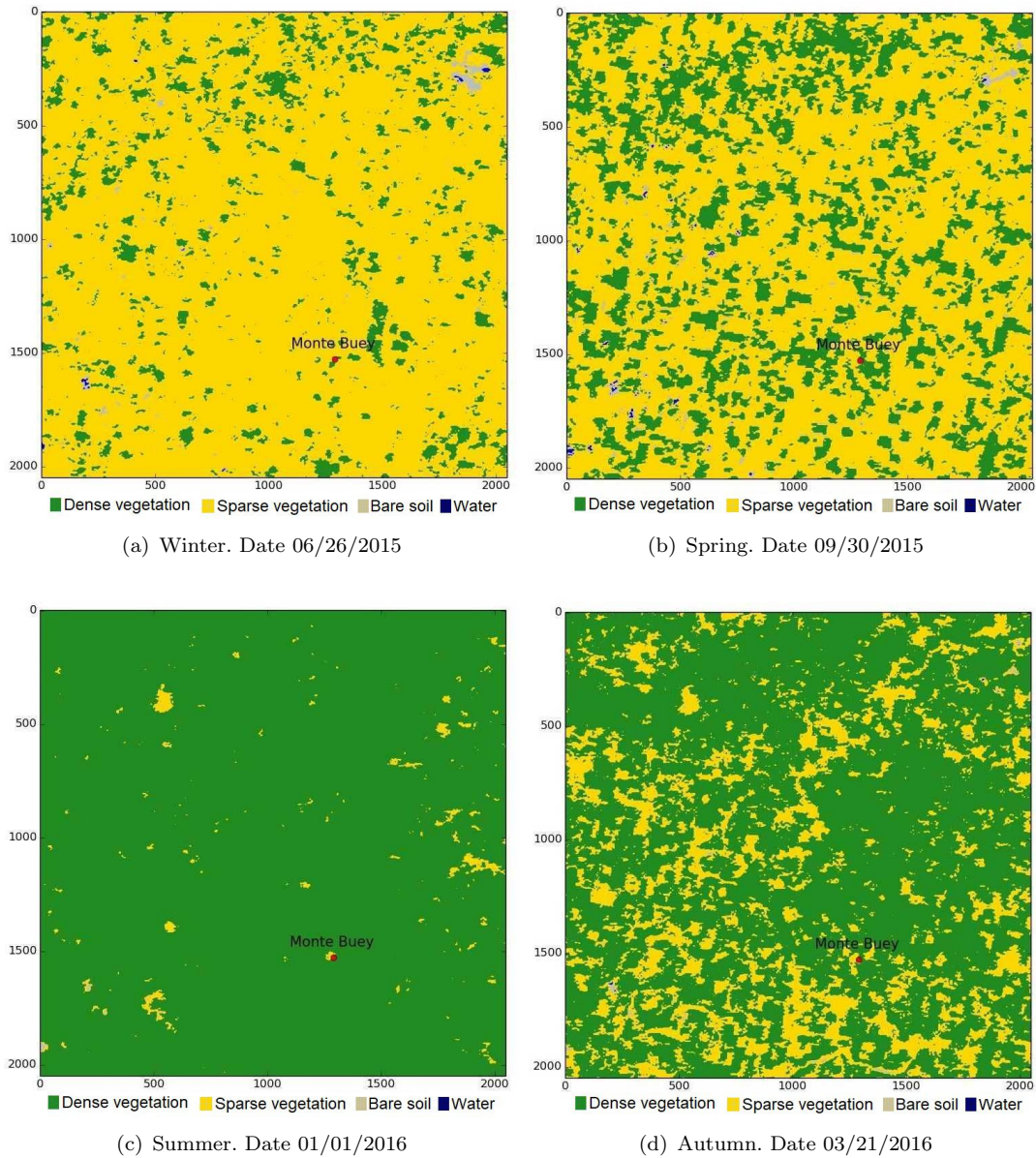


Figure 6.: Land cover classification based on NDVI MOD13Q1 product: water(Blue); bare soil(Gray); sparse vegetation (Yellow); dense vegetation (Green)

462 **4. Results and analysis**

463 This section presents the retrieval model calibration and the performances that they
 464 provide. The balance equation was solved using the statistical methods MLR, MARS,
 465 and MLP, which were implemented using the Python programming language.

466 **4.1. MLR model**

467 For the Multiple Linear Regression algorithm available input variables in the cleaned
 468 matrix were scaled to the range [0,1]. Then, the matrix is modified by applying
 469 logarithmic function to SM , T_a , RH , T_s , e_a . Indeed it should be noted that σ^0 is
 470 related to SM by a logarithmic function (e.g. see Zribi et al. (2005)). Thus σ^0 is
 471 not modified into the input matrix. In the literature the relationship reported
 472 between SM and PP is not uniform. In the literature the relationship reported
 473 derived from the observed data is $SM = 0.022 e^{0.093PP}$ was empirically
 474 without further processing.

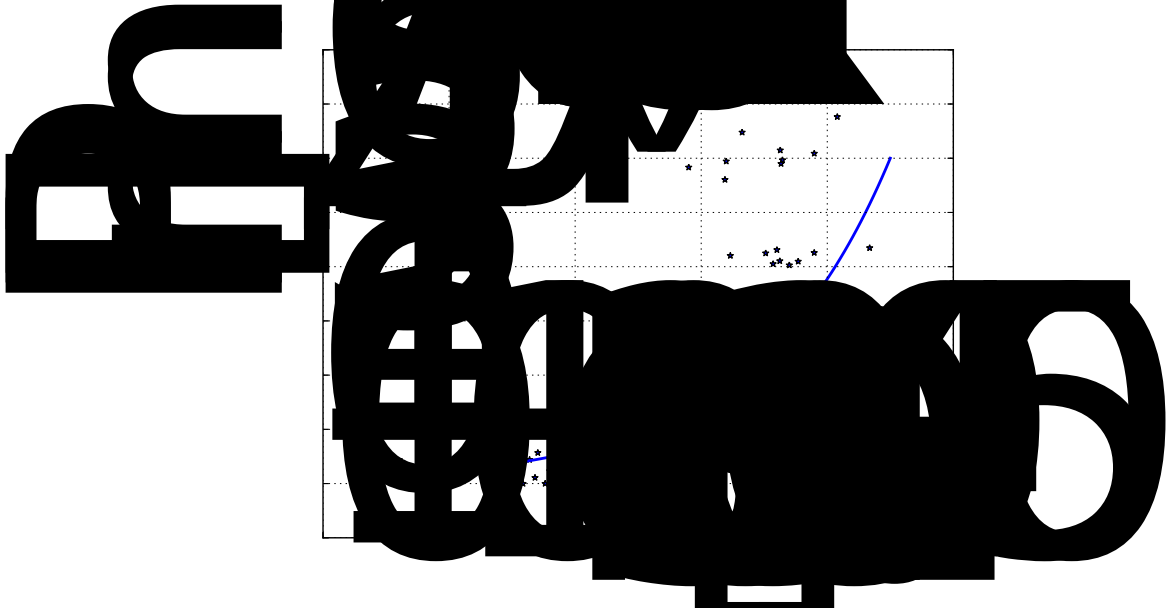


Figure 7.: Relationship between PP and SM

475 The selection of the variables that better explain SM variability is accomplished with
 476 the backward elimination procedure. The first MLR model created with the scrubbed
 477 matrix yielded an R^2 of about 0.42. The p-value statistic for σ^0 , RH and PP were
 478 significant (p-value < 0.05). The p-value for e_a , T_s , and T_a was larger than 0.05, e_a
 479 being the most critical. The analysis of the VIF statistic, on the other hand, indicated
 480 e_a as the most collinear variable followed by T_a . Applying p-value and VIF criteria, e_a
 481 was removed from the dataset, and the MLR was rerun. New results show that there
 482 was no collinearity between the variables but T_s has p-value equal to 0.538, thus it
 483 was eliminated and the MLR run again. The remaining variables yielded acceptable
 484 p-value and VIF values.

485 The final working matrix was composed of PP , $\log(T_a)$, $\log(RH)$ and σ^0 , which
 486 was split into two matrices: one for calibration and one for validation. The optimum
 487 number of data to calibrate and validate the model was determined by the learning
 488 curve analysis using the RMSE. The optimum data size is met with an RMSE equal
 489 to 6 %Vol. for a 75 % of calibration set size.

490 The calibration set used to calculate the MLR coefficients was composed of 83
 491 samples. The model obtained has an R^2 equal to 0.4 and RMSE of 6.36 %Vol. (see
 492 Figure 8). Table 3 summarizes MRL model coefficients, p-value and VIF statistics. It
 493 is worth noticing that selected variables remain significant (p-values lower than 0.05)

494 and without collinearity (VIF lower than 5).

Table 3.: Statistics of the selected input variables.

Parameter	Coefficient	p-value	VIF
β_0	1.52	0.0	-
σ^0	0.17	0.009	1.2
PP	0.11	0.008	1.1
$\log(T_a)$	0.15	0.03	1.3
$\log(RH)$	0.82	0.009	1.1

495 Finally, the MLR model resulted in being:

$$\widehat{SM} = 1.52 + 0.17 \sigma^0 + 0.11 PP + 0.15 \log(T_a) + 0.82 \log(RH) \quad (10)$$

496 As described previously (Section 1), the vertical movement of water in the
 497 soil-atmosphere system occurs spontaneously by energy gradients from the ground
 498 to the surrounding air. The latent heat flux, representing ET, causes the cooling
 499 process in the soil-atmosphere system and incorporates water vapor into the air. This
 500 process explains the relationship between T_a , T_s and RH and would justify T_a and
 501 RH selection (Venturini et al. 2012).

502 The model structure highlighted several relationships. Lin et al. (2016) showed that
 503 T_s increases with increasing T_a . This fact might be the reason why T_s is eliminated
 504 from the input matrix. On the other hand, e_a is related to T_a by an exponential
 505 function (Buck 1981), therefore it is also related to T_s . Furthermore, Kosmas et al.
 506 (2001) demonstrated the relationship between e_a , RH and ET . The collinearity of e_a
 507 with RH and T_a was demonstrated in many meteorological equations which explains
 508 the elimination of e_a from the dataset (Petterssen 1968).

509 4.2. *MARS model*

510 The Multivariate Adaptive Regression Splines modeling is a data-driven process
 511 derived from the MLR method which uses linear functions to describe nonlinear
 512 processes. The forward and backward phases are performed with the calibration data
 513 set to find the splines functions of the model, this is the reason why the raw datasets
 514 described in section 3.3 are required to construct the MARS model.

515 The calibrated MARS model has an R^2 equal to 0.58 and RMSE of 5.48 %Vol.
 516 (see Figure 8). The obtained basis functions and variable selection are summarized in
 517 Table 4. It is observed that RH , PP , T_a and σ^0 play important roles in deciding the
 518 MARS SM estimation model. The knots are located at $T_a = 9.9$, 18.42 and 21.22°C,
 519 $PP = 4.05$ mm and $\sigma^0 = -8.55$ dB. These knots delimit the variable range into the
 520 following intervals: T_a range is cut into four intervals, and the PP and σ^0 ranges each
 521 into two intervals, where different linear relationships are identified.

Table 4.: Coefficients and corresponding basis functions of MARS model.

Coefficient	Equation name	Equation
-27.77	Intercept	
0.55	Bf_1	RH
-2.15	Bf_2	$\text{Max}(0, 4.05 - PP)$
-0.56	Bf_3	$\text{Max}(0, -8.55 - \sigma^0)$
2.34	Bf_4	$\text{Max}(0, T_a - 9.9)$
2.54	Bf_5	$\text{Max}(0, 18.4199 - T_a)$
-3.81	Bf_6	$\text{Max}(0, T_a - 21.22)$

*Bf: Basic Function

522 Finally, the MARS prediction function resulted in being:

$$\widehat{SM} = -27.77 + 0.55 Bf_1 - 2.15 Bf_2 - 0.56 Bf_3 + 2.34 Bf_4 + 2.54 Bf_5 - 3.81 Bf_6 \quad (11)$$

523 According to the above relationships, when PP is less than 4.05 mm, SM will be
 524 decreased by 2.15 times the PP more than the threshold value of 4.05 mm (Bf_2).
 525 If σ^0 is less than -8.55 dB the SM decreases by 0.56 times the amount less than the
 526 threshold (Bf_3). When the T_a variable is bigger than the threshold value of 9.9 °C and
 527 less than 18.42 °C, SM will increase by 2.34 (Bf_4) and 2.54 (Bf_5) times the amount
 528 bigger/smaller than the threshold respectively, while for T_a values greater than the
 529 threshold value of 21.22 °C, the SM decreases by 3.81 times the amount bigger than
 530 the threshold (Bf_6).

531 4.3. MLP ANN model

532 For the MultiLayer Perceptron Artificial Neural Network algorithm, available inputs
 533 were scaled to the range [-1,1]. Then, the standardized dataset was randomly divided
 534 into calibration (75%) and test (25%) subsets.

535 In order to define the appropriate MLP structure for this work, a trial and error
 536 scheme has been used. The number of input neurons are four, one for each of the
 537 proposed variables. The number of hidden layers and the neurons in each hidden
 538 layer have been varied to obtain the best configuration that guaranteed an optimum
 539 model performance (minimum error and maximum convergence), avoiding any possible
 540 overfitting. The number of hidden neurons increased from 1 to 8 in steps of one in
 541 each trial. The network was calibrated and tested for each trial looking for the lowest
 542 RMSE.

543 The resulting MLP architecture has four layers, i.e. one input layer with four
 544 neurons; two hidden layers with three and four neurons respectively, and one output
 545 layer with one neuron. The transfer function for this model was the REctified Linear
 546 Unit (ReLU) function for the hidden layers. This function had the best network
 547 performance and performed better than the sigmoid function. The R^2 and RMSE for
 548 this configuration were 0.58 and 4.2 %Vol respectively. Therefore, the adopted MLP
 549 model architecture has a configuration of 4-3-4-1 neurons. The derived MLP model
 550 was calibrated with an R^2 equal to 0.54 and RMSE of 5.73 % Vol. (Figure 8). This
 551 configuration of the layers is similar to the one used by Paloscia et al. (2008, 2010),
 552 who found that a MLP configuration with two hidden layers was the optimal one for
 553 SM estimation from σ^0 and SM ground measurements. Also, Paloscia et al. (2013)
 554 proposed an algorithm based on an ANN with two hidden layers in order to ingest
 555 different inputs based on the possible configurations of the Sentinel-1A sensor, thus

556 obtaining a SM product in near-real-time. The accuracy (RMSE) of the algorithm
 557 does indeed range from around 2% of SM, in HV polarization, to 6% of SM when only
 558 VV polarization is present.

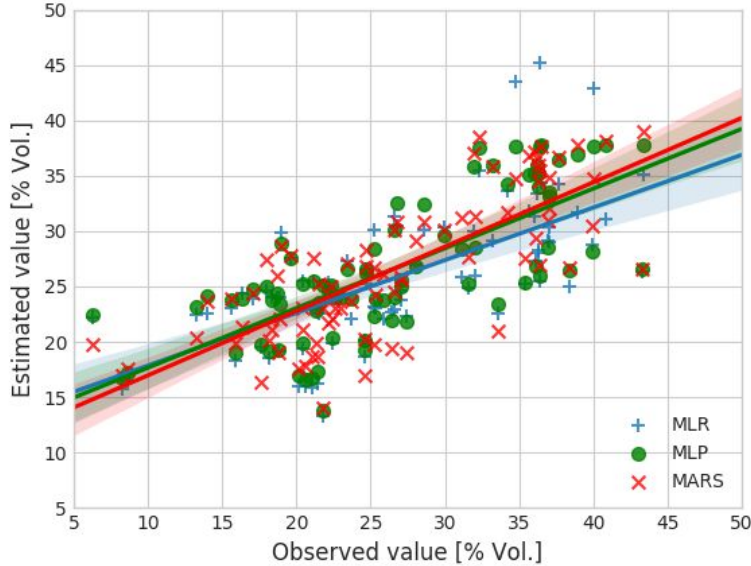


Figure 8.: Comparison between the observed and estimated SM values that result in the calibration phase for the MLR, MARS and MLP methods.

559 **4.4. Model Validation**

560 Subsequently, calibrated models were validated to evaluate the efficacy of the proposed
 561 methodologies. This operation implies applying each algorithm to a set of data that was
 562 not used in the definition of the MLR coefficients, MARS splines, and the MLP weights.
 563 The validation dataset was thus composed of 28 samples. Finally, the estimated soil
 564 moisture values were compared with the observed ones.

565 For the MLR method, the comparison between estimated and observed SM values
 566 yielded R^2 equal to 0.58, RMSE of 4.25 %Vol. and a bias of 0.25 %Vol. For the MARS
 567 method, the validation yielded R^2 equal to 0.64, RMSE of 4.17 %Vol. and a bias of
 568 -0.71 %Vol. The MLP ANN method yielded R^2 equal to 0.65, RMSE of 4.16 %Vol.
 569 and a bias = -1.25 %Vol. (See Figure 9). The error obtained by models represents 10
 570 % of the observed values of SM. Figure 9 shows that SM values around 25 %Vol. were
 571 well simulated for all the models, however the models overestimated SM observations
 572 lower than 25 %Vol.

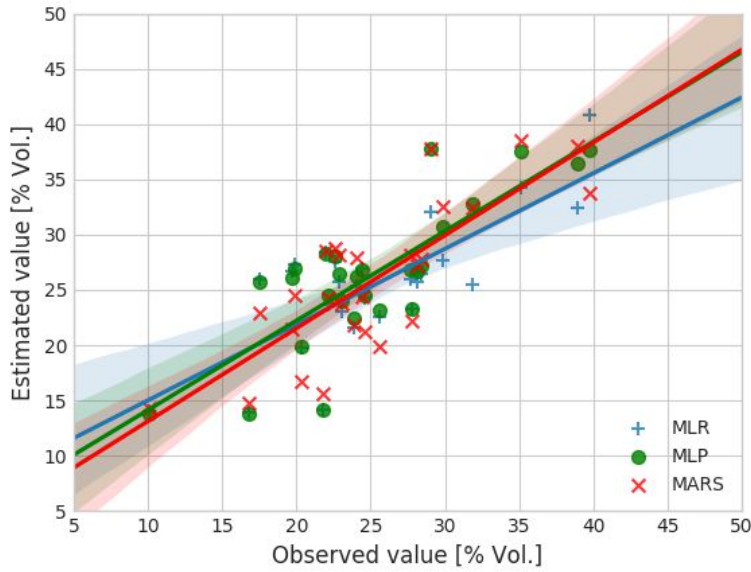


Figure 9.: Comparison between the observed and estimated SM values that result in the validation phase for the MLR, MARS and MLP methods.

573 The implemented models yielded results comparable to those published by other
 574 authors using different approaches (see Table 5 that summarizes the accuracy of the
 575 methods available in the literature). In general, the error varies from 3 to 20 %Vol.
 576 It is the Support Vector Regression (SVR) method that is best performed over steep
 577 terrains. Inversion models present a broad range of error which basically depends
 578 on the land cover. The simple MLR model proposed here, yields errors of about 4
 579 %Vol. comparable with inversion models and statistical approaches. In particular,
 580 the results presented in this paper agree with those published by Paloscia et al.
 581 (2008), who compared several algorithms for retrieving SM over a flat agricultural
 582 area in Northern Italy. The authors found errors between 3 and 9 %Vol. The results
 583 achieved by Lievens and Verhoest (2012) with a physically-based approach, using
 584 RADARSAT-2 images in flat topography, are similar to the ones obtained in this work
 585 with hydro-meteorological data. However, the methodology proposed here has a much
 586 simple structure, and it makes use of widely available data from weather stations and
 587 SAR data (here we used Sentinel-1A VV pol backscattering). The proposed models
 588 make possible SM estimation in regions where soil ancillary data (e.g. soil properties)
 589 are scarce because of their remoteness or their inaccessible locations or their vast
 590 dimensions. In particular, the method incorporates the available variables and relates
 591 them automatically to the processes of water inlets and outlets to the soil system.
 592 Although the final accuracy is slightly lower compared to other methods, the developed
 593 approach allows the obtaining of SM estimations in areas where other methodologies
 594 cannot be used.

Table 5.: Summary of SM retrieval methodologies from SAR data.

Methodology	Band	Authors	Topography	Land cover	RMSE (%Vol.)
ANNs	C	Paloscia et al. (2008)	Flat land	vegetated	9
		Paloscia et al. (2010)	Mountain		10
	X	Kseneman et al. (2012)	Flat land	bare soil	14
SVR	C	Pasolli et al. (2011)	Mountain	vegetated	3
		Bertoldi et al. (2014)			5
Inversion	C	Paloscia et al. (2008)	Flat land	vegetated	12
		Merzouki et al. (2011)			6
		Gherboudj et al. (2011)			20
		Lievens and Verhoest (2012)			4
	X	Baghdadi et al. (2012)	bare soil	3	
Bayes	C	Paloscia et al. (2008)	Flat land	vegetated	11
		Notarnicola et al. (2008)		bare soil	<10
		Pierdicca et al. (2010)		vegetated	4

595 **4.5. Uncertainty analysis of the model**

596 A first-order analysis is appropriate for quantifying the uncertainties introduced in
 597 a dependent variable by independent variables and parameters due to the inherent
 598 uncertainties in measurements and estimates (Chow et al. 1988).

599 This method assumes that a variable Y is expressed as a function of x , such that
 600 $Y = f(x)$. If the model f is correct, the uncertainties of Y are associated with x
 601 measurement errors. Knowing that \bar{x} is the mean of x , the mean of Y , \bar{Y} is calculated
 602 as $\bar{Y} = f(\bar{x})$. If the true value of x differs from \bar{x} by a certain amount, the effect of
 603 this disagreement on Y can be examined by expanding $f(x)$ as a Taylor series around
 604 $x = \bar{x}$. If the high order terms of the Taylor series are neglected, the expression of the
 605 error is approximated, as shown in equation 12

$$Y - \bar{Y} \approx \left. \frac{df}{dx} \right|_{x=\bar{x}} (x - \bar{x}) \quad (12)$$

606 The error variance of Y is expressed as:

$$S_Y^2 \approx \left(\left. \frac{df}{dx} \right|_{x=\bar{x}} \right)^2 S_x^2 \quad (13)$$

607 where S_Y^2 and S_x^2 are the variances of Y and x respectively. This methodology can be
 608 extended to several independent variables (Chow et al. 1988).

609 The knowledge of the analytical expression of the models (i.e. the function f) is
 610 necessary for applying this analysis that is based on the derivatives. For this reason,
 611 the uncertainty analysis in this paper was carried out only for the MLR model, due
 612 to the complexity of analytically expressing the MLP and MARS model equations.

613 By applying equation 13 to equation 10, the SM error variance is as follows:

$$S_{SM}^2 = \left(\frac{\overline{SM}\beta_{\sigma^0}}{\log(e)} \right)^2 S_{\sigma^0}^2 + \left(\frac{\overline{SM}\beta_{PP}}{\log(e)} \right)^2 S_{PP}^2 + \left(\frac{\overline{SM}\beta_{T_a}}{T_a} \right)^2 S_{T_a}^2 + \left(\frac{\overline{SM}\beta_{RH}}{RH} \right)^2 S_{RH}^2 \quad (14)$$

614 Replacing the average values and MLR coefficients for each variable, the following

615 is obtained:

$$S_{SM}^2 = 105.8 S_{\sigma^0}^2 + 44.3 S_{PP}^2 + 57.4 S_{T_a}^2 + 689.6 S_{RH}^2 \quad (15)$$

616 In order to examine equation 15, a uniform noise of 10 % and 20 % of the S value
617 are added to each variable, as shown in Table 6.

Table 6.: Mean and S of standardized input variables for the first order analysis.

Variable	Mean	S	S 10 %	S 20 %
σ^0	0.52	0.22	0.24	0.26
PP	0.09	0.19	0.21	0.23
T_a	0.52	0.21	0.23	0.25
<i>RH</i>	0.82	0.1	0.11	0.12

618 The uncertainty introduced by σ^0 in the SM variance varies from 6.3 to 7.4, with
619 an S of 10% and 20%, respectively. *RH* is the variable with the largest nominal value
620 adding an error of about 6.9 to 9.6, with an S of 10% and 20 % respectively. Summing
621 all variable variance, SM uncertainty ranges from 18% to 23% of the mean SM.

622 This analysis showed that errors in SAR image processing would not affect the SM
623 estimation as much as errors in *RH* would.

624 4.6. Soil moisture maps

625 After model calibration and validation at point scale, the methodology was extended
626 to the entire study area for generating SM maps. The SAR image was segmented to
627 eliminate areas in which the model cannot be used as vegetated regions, water bodies,
628 and cities. For this task, optical data were processed to define a mask image. Then,
629 the models MLR, MARS, and MLP were applied to the remaining pixels representing
630 bare soil and sparse vegetation areas.

631 In developing the algorithms, the MODIS MOD13Q1 product was used because
632 Landsat 8 has a poorer revisit time: this did not allow the derivation of NDVI
633 maps close enough to all of the available Sentinel-1A acquisitions. However, for map
634 generation, NDVI was derived from Landsat 8 images which have a finer spatial
635 resolution than MODIS. Then Sentinel-1A images were collected close to Landsat 8
636 overpasses. Table 7 lists the acquisition date of the Sentinel-1A and Landsat 8
637 images used for the map generation.

Table 7.: Inventory of images for generating SM maps.

Date (dd/mm/yyyy)	
Sentinel-1A	Landsat 8
29/06/2015	18/06/2015
03/10/2015	08/10/2015
28/12/2015	27/12/2015
19/03/2016	16/03/2016

638 The classification mask was derived for each SAR acquisition with the NDVI
639 classes previously described. Also, urban areas were obtained by using a supervised
640 classification. The resulting class images are shown in Figure 10. Masked pixels are
641 presented in black, while pixels where the the retrieval applies are in white. The

642 evolution of the vegetation cover (black color) and how the plots change according to
 643 the state of the crop can be observed in detail in these images. For the acquisitions
 644 collected in the winter-spring period (Figure 10 - a and b) the pixels with NDVI greater
 645 than 0.5 (black color) could be associated with pastures or wheat. In the study area,
 646 the main crops were carried out during the summer period (Figure 10 - c and d), the
 647 most important being soy and corn.

648 Sentinel-1A images were upscaled to a pixel of 30 m to reduce the *Speckle* noise
 649 and match optical data resolution. Meteorological point data were interpolated on the
 650 same grid. Finally, the models were applied to the data of σ^0 , PP , T_a and RH of the
 651 pixels of bare soil and sparse vegetation (white pixels in Figure 10).

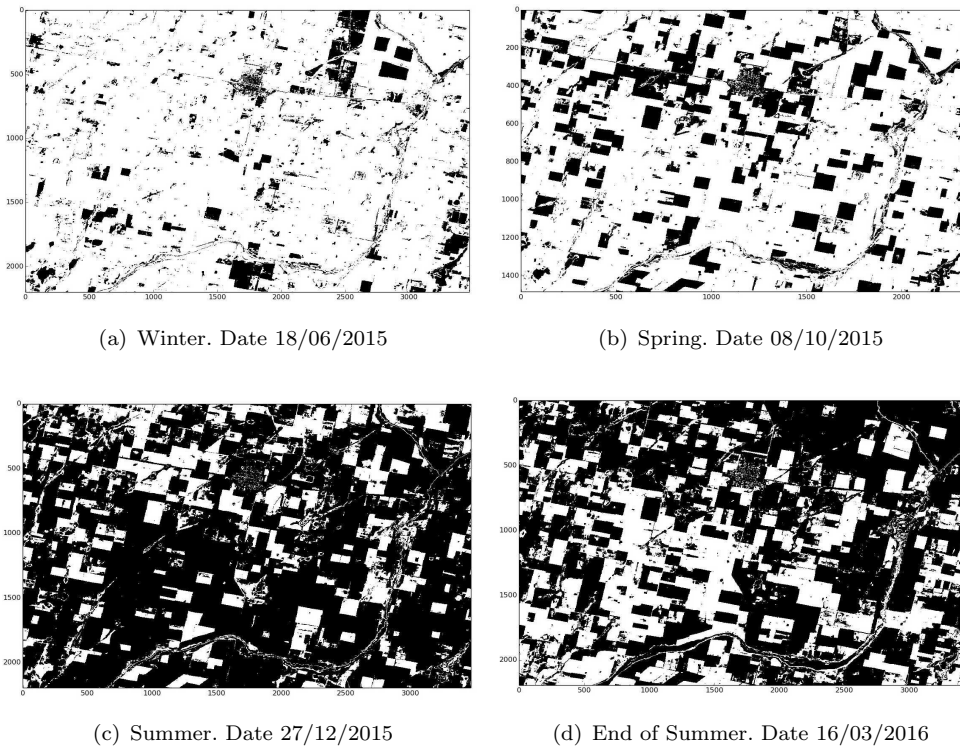


Figure 10.: Class images for different Landsat 8 acquisitions. Black: unusable pixels; White: pixels where the models were applied

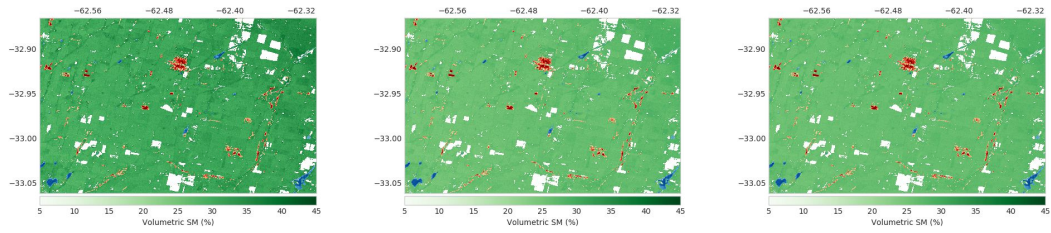
652 SM maps are depicted in Figure 11. The first column shows SM estimated with
 653 the MLR method, the second column represents the maps obtained with MARS, and
 654 the third one the maps elaborated with the MLP method. In these maps the green
 655 scale represents SM. Blue is used for water pixels, red for cities and white for densely
 656 vegetated areas.

657 The mean, minimum (Min), maximum (Max) and S of SM maps are presented
 658 in Table 8. In general, estimated SM expressed in %Vol., varies between 23.5 and
 659 58.5 for the MLR method, between 23.2 and 34 for MARS and between 21.3 and
 660 39.2 for the MLP method. There is a large range of variability in the regional mean
 661 value between the methods. The largest difference among the methods is observed
 662 for summer dates (i.e. 28/12/2015 and 16/03/2016) when the MLR method reached a
 663 maximum of 58.5 %Vol. while MARS and MLP reached 34 and 39 %Vol., respectively.
 664 These differences for the summer dates can be related to the relative weight that each

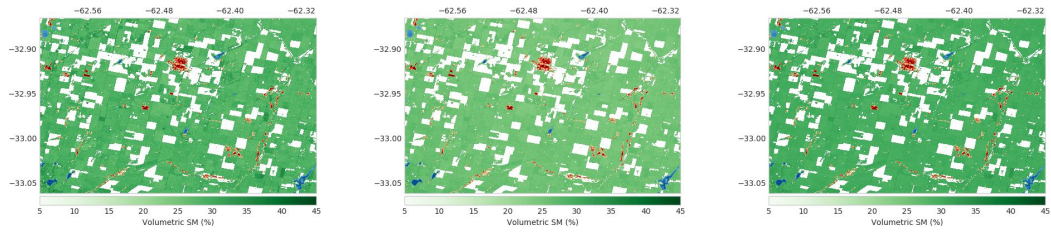
665 method assigns to PP, since it is in this season when large precipitations take place.
666 The range of estimated SM depends on the input variable values and standardization
667 function applied to solve the non-linearity problem of the balance equation. In order
668 to analyze the quality of method estimation, the RMSE for each map was calculated
669 by contrasting them against CONAE in situ data. The largest RMSE was obtained
670 with the MLR method for all dates. In particular for date 16/03/2016 the maximum
671 error yielded was 17 %Vol. approximately. These RMSE were reduced by using the
672 MLP method but the smallest errors were yielded by the MARS method.

673 Although the mean SM values obtained with the methods are different, a visual
674 comparison indicated that the maps attained with the MLR, MARS and MLP methods
675 provided similar spatial variability.

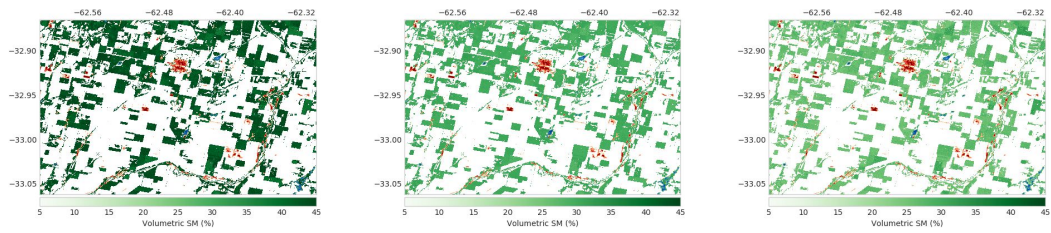
Winter. Date 29/06/2015



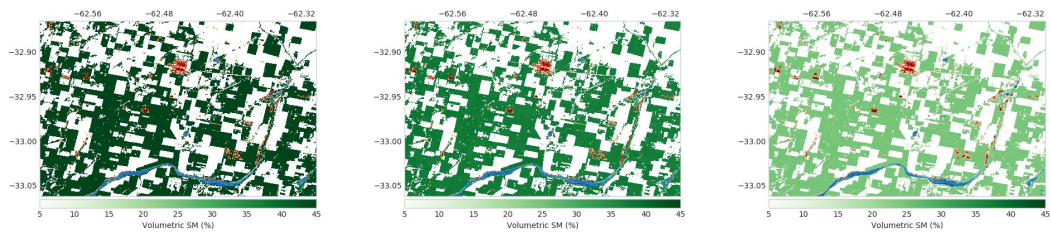
Spring. Date 03/10/2015



Summer. Date 28/12/2015



End of Summer. Date 16/03/2016



(a)

(b)

(c)

Figure 11.: SM maps on the study area: (a) SM obtained with MLR, (b) SM using MARS, (c) SM using MLP; SM (Green scale); Water bodies (Blue); Cities (Red); Masked pixels where models do not apply (White).

Table 8.: SM maps statistical analysis

Date	Method	Max %Vol.	Min %Vol.	Mean %Vol.	S %Vol.	RMSE %Vol.
29/06/2015	MLR	41.9	25.6	31.3	2	14.3
	MARS	29.8	22.8	26.4	1.1	6.9
	MLP	31.3	23.7	26.6	1	7.6
03/10/2015	MLR	37.1	23.5	28.7	1.7	9.5
	MARS	28.3	21.3	25.2	1	6.2
	MLP	33.6	26.6	29.6	0.9	8.6
28/12/2015	MLR	56.5	35.9	43.5	1.7	14.9
	MARS	32.8	26.5	30.1	1	1.9
	MLP	34	23.2	26.4	1.7	4.2
16/03/2016	MLR	58.5	37.6	48.5	2.2	17.4
	MARS	39.2	32.1	37.3	0.8	2
	MLP	25.9	23.2	24.3	0.3	11.8

676 5. Discussion

677 Many authors studied how to exploit SAR data by merging them with other different
678 sources of information. It is well known that SAR technology has a great potential for
679 regional refine-scale applications due to its finer spatial resolution than those of the
680 scatterometer and radiometer. However, SAR data are affected by the *Speckle* noise
681 that needs to be reduced by applying the multilook operation that worsens the original
682 spatial resolution. However, these applications are limited by SAR coarse temporal
683 resolution, especially in the case of the study area of this paper, where only about
684 2-3 images per month are collected by Sentinel-1 satellites. The results presented here
685 demonstrated that it is possible to obtain SM maps with adequate spatial resolution
686 integrating different sources of data. i.e., Sentinel-1 and field observations, with simple
687 machine learning methods.

688 In this article, SM was estimated by solving the nonlinear water balance equation
689 proposed in the methodology section. It has been shown that MLR, MARS, and
690 ANN-MLP methods are suitable for nonlinear equation approximation, even though all
691 of them have advantages and disadvantages. The main advantage of the MLR is that
692 it approximated the water balance with a physically meaningful relationship between
693 the input variables, with a reasonable amount of data. In this sense, MARS is also able
694 to derive a physically understandable relationship between input and output variables.
695 Furthermore, MARS weighs each input variable in different ranges of values, resulting
696 in a smart linearization of the nonlinear processes. In opposition ANN-MLP solves
697 any type of non-linearity; although a large number of samples is required. Moreover,
698 the resulting equation of any ANN method is not explicit. In general, MARS and
699 MLP resulting equations are more difficult to analyze, in terms of uncertainty and
700 total error, compared to those obtained with MLR. In other words, the MLR resulting
701 equation is a straightforward relationship between the inputs and outputs. In general,
702 the MLR method is simple and statistically manageable compared with the other two
703 methods, while the performance is comparable to the others.

704 In terms of the number of elements to calibrate, MLR has a parameter for each
705 input variable. However, in the MLP and MARS methods, multiple parameters must
706 be calibrated. For instance, in the MLP method the number of hidden layers, the

707 number of neurons per hidden layer, weights, function of activation, learning rate,
708 among other parameters, have to be calibrated; and in the MARS method the number
709 of base functions, the knot locations, the coefficient of each basis function, among
710 others feature, have to be calibrated. The errors reported here are comparable with
711 those published with SVR, ANNs, Bayes, among other more complex methods.

712 Indeed, it is well documented in the literature that backscattering at C-band is
713 more sensitive to the vegetation cover than L-band. Also, it was demonstrated that
714 VV polarization is not the best polarization for SM retrieval. However, the results
715 discussed here showed that Sentinel-1 images at VV polarization yield acceptable
716 errors, though the estimates should be improved if HH polarization images are used.
717 Even so, Sentinel-1 free of charge data provided by ESA represent a key opportunity for
718 researchers to develop SM product with lower cost than those derived with expensive
719 suitable polarization and/or frequency.

720 6. Conclusions

721 In the last decades, several models based on artificial neural networks, fuzzy logic,
722 support vector machine, and Bayesian inversion approaches have been developed to
723 estimate SM at high spatial resolution. Many of these methods use SAR backscattering
724 as the main proxy for deriving soil properties, such as SM, neglecting water inputs.
725 They also need other sources of information like soil texture and structure, topographic
726 features and land cover patterns to solve the ill-posed retrieval problem. Unfortunately,
727 these data are not available in several parts of the world as they are in the emerging
728 countries. However, weather stations are a “low-tech” instrument that can provide
729 useful information for the soil moisture estimation since the later variable is the result
730 of processes like precipitation, evapotranspiration, etc., which a ground station can
731 monitor.

732 The presented methodology for estimating SM was based on the water balance
733 equation that considers SM as the result of water inputs and outputs to the soil system.
734 The water balance equation was solved by using the statistical methods MLR, MARS,
735 and ANN-MLP. It is worth mentioning that in the proposed method the microwave
736 backscattering it is considered to be surrogate of soil property variables.

737 MLR, MARS and ANN-MLP yielded similar results, with errors of approximately
738 10 % of the observed average value of SM. These errors are comparable with those
739 published by other authors with different methodologies, such as inversion models,
740 Bayesian models, and SVR.

741 The *pros* of the proposed models stem from the diversity of variables that could
742 be used within the same equation: i.e. meteorological variables, surface properties,
743 hydrological variables and remotely sensed data. The mix of different variables enables
744 building simple water balance models that explain SM variability.

745 In summary, non-parametric models have been widely used in the past to estimate
746 SM with soil property variables. However, there is very little previous work that applied
747 the proposed methodologies with hydro-meteorological data to model SM. In this
748 sense, the MLR and MARS models presented here yield promising results worth further
749 exploring with other data sources.

750 Acknowledgement(s)

751 The authors wish to thank Argentinas's Comisión Nacional de Actividades Espaciales
752 (CONAE) for kindly providing the *in situ* data belonging to its network of automatic
753 stations, as well as the European Space Agency (ESA) for providing the Sentinel-1A
754 data. Thanks to the Consejo Nacional de Investigaciones Científicas y Técnicas
755 (CONICET) of Argentina that supports Gabriel García PhD studies.

756 References

- 757 Afshar, M. H., and M. T. Yilmaz. 2017. "The added utility of nonlinear methods compared to
758 linear methods in rescaling soil moisture products." *Remote Sensing of Environment* 196:
759 224–237.
- 760 Ahmad, S., A. Kalra, and H. Stephen. 2010. "Estimating soil moisture using remote sensing
761 data: A machine learning approach." *Advances in Water Resources* 33 (1): 69–80.
- 762 Albergaria, J. T., F. G. Martins, and C. Alvim-Ferraz, M. C. M. and Delerue-Matos. 2014.
763 "Multiple linear regression and artificial neural networks to predict time and efficiency of
764 soil vapor extraction." *Water, Air, and Soil Pollution* 225 (8).
- 765 Ali, I., F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola. 2015. "Review of
766 machine learning approaches for biomass and soil moisture retrievals from remote sensing
767 data." *Remote Sensing* 7 (12): 16398–16421.
- 768 Alley, W. M. 1984. "On the Treatment of Evapotranspiration, Soil Moisture Accounting, and
769 Aquifer Recharge in Monthly Water Balance Models." *Water Resources Research* 20 (8):
770 1137–1149.
- 771 Aubert, M., N. Baghdadi, M. Zribi, K. Ose, M. E. Hajj, E. Vaudour, and E. Gonzalez-Sosa.
772 2013. "Toward an Operational Bare Soil Moisture Mapping Using TerraSAR-X Data
773 Acquired Over Agricultural Areas." *IEEE Journal of Selected Topics in Applied Earth
774 Observations and Remote Sensing* 6 (2): 900–916.
- 775 Baghdadi, N., M. Aubert, and M. Zribi. 2012. "Use of TerraSAR-X data to retrieve soil
776 moisture over bare soil agricultural fields." *IEEE Geoscience and Remote Sensing Letters* 9
777 (3): 512–516.
- 778 Bedano, J. C., A. Domínguez, R. Arolfo, and L. Wall. 2016. "Effect of Good Agricultural
779 Practices under no-till on litter and soil invertebrates in areas with different soil types." *Soil
780 and Tillage Research* 158: 100–109.
- 781 Behari, J. 2005. *Microwave Dielectric Behavior of Wet Soils*. Vol. 8 of *Remote Sensing and
782 Digital Image Processing*. Dordrecht: Springer Netherlands.
- 783 Beleites, C., U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp. 2013. "Sample size planning
784 for classification models." *Analytica Chimica Acta* 760 (June 2012): 25–33.
- 785 Bertoldi, G., S. Della Chiesa, C. Notarnicola, L. Pasolli, G. Niedrist, and U. Tappeiner.
786 2014. "Estimation of soil moisture patterns in mountain grasslands by means of SAR
787 RADARSAT2 images and hydrological modeling." *Journal of Hydrology* 516: 245–257.
- 788 Brocca, L., S. Camici, F. Melone, T. Moramarco, J. Martínez-Fernández, J. F. Didon-Lescot,
789 and R. Morbidelli. 2014. "Improving the representation of soil moisture by using a
790 semi-analytical infiltration model." *Hydrological Processes* 28 (4): 2103–2115.
- 791 Buck, A. L. 1981. "New Equations for Computing Vapor Pressure and Enhancement Factor."
792 *Journal of Applied Meteorology* 20 (12): 1527–1532.
- 793 Chow, V. T, D. R. Maidment, and L. W. Mays. 1988. *Applied Hydrology*. McGraw-Hill
794 Publishing Company.
- 795 Cohen, J., P. Cohen, S. G. West, and L. Aiken. 2003. *Applied Multiple Regression / Correlation
796 Analysis for the Behavioral Sciences*. Vol. Third Edit.
- 797 Del Valle, H. F., P. D. Blanco, W. Sione, C. M Rostagno, and N. O. Elissalde. 2008.
798 "Assessment of Salt-Affected Soils Using Multisensor Radar Data: A Case Study from

- 799 Northeastern Patagonia (Argentina).” In *Remote Sensing of Soil Salinization: Impact on*
800 *Land Management*, edited by FL-USA. Metternicht, Graciela And Zinck, Alfred CRC Press,
801 Boca Ratón, Chap. 9, 155–173.
- 802 Duval, M. E., J. A. Galantini, J. M. Martínez, F. M. López, and L. G. Wall. 2016. “Sensitivity
803 of different soil quality indicators to assess sustainable land management: Influence of site
804 features and seasonality.” *Soil and Tillage Research* 159: 9–22.
- 805 Engman, E. T. 1991. “Applications of microwave remote sensing of soil moisture for water
806 resources and agriculture.” *Remote Sensing of Environment* 35 (2-3): 213–226.
- 807 Entekhabi, D., E. G. Njoku, P. E. O’Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K.
808 Entin, et al. 2010. “The soil moisture active passive (SMAP) mission.” *Proceedings of the*
809 *IEEE* 98 (5): 704–716.
- 810 Forte Lay, J., O. Scarpati, and A. Capriolo. 2008. “Precipitation variability and soil water
811 content in Pampean Flatlands (Argentina).” *Geofísica Internacional* 47 (4): 341–354.
- 812 Fox, J. 2002. *R and S-Plus Companion to Applied Regression*. Sage Publications, Inc. Thousand
813 Oaks, CA, USA ©2002.
- 814 Friedman, J. H. 1991. “Multivariate Adaptive Regression Splines.” *The Annals of Statistics.*
815 *Institute of Mathematical Statistics* 19 (1): 1–67.
- 816 Gherboudj, I., R. Magagi, A. A. Berg, and B. Toth. 2011. “Soil moisture retrieval over
817 agricultural fields from multi-polarized and multi-angular RADARSAT-2 SAR data.”
818 *Remote Sensing of Environment* 115 (1): 33–43.
- 819 Grings, F., C. A. Bruscantini, E. Smucler, F. Carballo, M. E. Dillon, E. A. Collini,
820 M. Salvia, and H. Karszenbaum. 2015. “Validation Strategies for Satellite-Based Soil
821 Moisture Products over Argentine Pampas.” *IEEE Journal of Selected Topics in Applied*
822 *Earth Observations and Remote Sensing* 8 (8): 4094–4105.
- 823 Harrell, F. 2001. *With Applications to Linear Models, Logistic Regression, and Survival*
824 *Analysis*. First edit ed.
- 825 Haykin, S. S. 2009. *Neural networks and learning machines*. 3rd ed. Pearson Education.
- 826 Helsel, D. R., and R. M. Hirsch. 2002. “Statistical Methods in Water Resources.”
827 *Technometrics* 36 (3): 323.
- 828 Holben, B. N. 1986. “Characteristics of maximum-value composite images from temporal
829 AVHRR data.” *International Journal of Remote Sensing* 7 (11): 1417–1434.
- 830 Hornik, K., M. Stinchcombe, and H. White. 1989. “Multilayer feedforward networks are
831 universal approximators.” *Neural Networks* 2 (5): 359–366.
- 832 Kang, H., and V. Sridhar. 2018. “Improved Drought Prediction Using Near Real-Time Climate
833 Forecasts and Simulated Hydrologic Conditions.” *Sustainability* 10 (6): 1799.
- 834 Kerr, Y. H., P. Waldteufel, J. P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M. J. Escorihuela,
835 et al. 2010. “The SMOS Mission: New Tool for Monitoring Key Elements of the Global
836 Water Cycle.” *Proceedings of the IEEE* 98 (5): 666–687.
- 837 Kerr, Y. H., P. Waldteufel, J. P. Wigneron, J. M. Martinuzzi, J. Font, and M. Berger. 2001.
838 “Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission.”
839 *IEEE Transactions on Geoscience and Remote Sensing* 39 (8): 1729–1735.
- 840 Kornelsen, K. C., and P. Coulibaly. 2013. “Advances in soil moisture retrieval from synthetic
841 aperture radar and hydrological applications.” *Journal of Hydrology* 476: 460–489.
- 842 Kosmas, C., M. Marathianou, S. Gerontidis, V. Detsis, M. Tsara, and J. Poesen. 2001.
843 “Parameters affecting water vapor adsorption by the soil under semi-arid climatic
844 conditions.” *Agricultural Water Management* 48 (1): 61–78.
- 845 Kseneman, M., D. Gleich, and B. Potočník. 2012. “Soil-moisture estimation from TerraSAR-X
846 data using neural networks.” *Machine Vision and Applications* 23 (5): 937–952.
- 847 Legates, D. R., and G. J. McCabe Jr. 1999. “Evaluating the Use of ” Goodness of Fit ” Measures
848 in Hydrologic and Hydroclimatic Model Validation.” *Water Resources Research* 35 (1):
849 233–241.
- 850 Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges. 1994. “A simple hydrologically
851 based model of land surface water and energy fluxes for general circulation models.” *Journal*
852 *of Geophysical Research: Atmospheres* 99 (D7): 14415–14428.

- 853 Lievens, H., and N. E. C. Verhoest. 2012. "Spatial and temporal soil moisture estimation from
854 RADARSAT-2 imagery over Flevoland, The Netherlands." *Journal of Hydrology* 456-457:
855 44-56.
- 856 Lillesand, T., and R. Kiefer. 1994. *Remote Sensing and Image Interpretation*. John Wiley (&
857 Sons, New York; 3rd edition.
- 858 Lin, X., W. Zhang, Y. Huang, W. Sun, P. Han, L. Yu, and F. Sun. 2016. "Empirical
859 estimation of near-surface air temperature in China from MODIS LST data by considering
860 physiographic features." *Remote Sensing* 8 (8): 1-15.
- 861 Liu, D., A. K. Mishra, and Z. Yu. 2016. "Evaluating uncertainties in multi-layer soil moisture
862 estimation with support vector machines and ensemble Kalman filtering." *Journal of*
863 *Hydrology* 538: 243-255.
- 864 Loew, A., and W. Mauser. 2006. "A Semiempirical Surface Backscattering Model for Bare Soil
865 Surfaces Based on a Generalized." *East* 44 (4): 1022-1035.
- 866 Mallat, G. 1989. "A Theory for Multiresolution Signal Decomposition : The Wavelet
867 Representation." II (7).
- 868 Marquardt, D. W. 1970. "Generalized Inverses, Ridge Regression, Biased Linear Estimation,
869 and Nonlinear Estimation." *Technometrics* 12 (3): 591.
- 870 Mattia, F., G. Satalino, L. Dente, and G. Pasquariello. 2006. "Using a priori information to
871 improve soil moisture retrieval from ENVISAT ASAR AP data in semiarid regions." *IEEE*
872 *Transactions on Geoscience and Remote Sensing* 44 (4): 900-911.
- 873 Merzouki, A., H. Mcnairn, and A. Pacheco. 2011. "Evaluation of the Dubois , Oh , and IEM
874 radar backscatter models over agricultural fields using C-band RADARSAT-2 SAR image
875 data." 36 (October 2009).
- 876 Mladenova, I. E., T. J. Jackson, E. Njoku, R. Bindlish, S. Chan, M. H. Cosh, T. R.H. Holmes,
877 et al. 2014. "Remote monitoring of soil moisture using passive microwave-based techniques
878 - Theoretical basis and overview of selected algorithms for AMSR-E." *Remote Sensing of*
879 *Environment* 144: 197-213.
- 880 Montgomery, D. C., E. A. Peck, and G. G. Vining. 1992. *Introduction to Linear Regression*
881 *Analysis*. Wiley-Interscience; 2 edition.
- 882 Naeimi, V., K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner. 2009. "An Improved Soil
883 Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations." *IEEE*
884 *Transactions on Geoscience and Remote Sensing* 47 (7): 1999-2013.
- 885 Narayan, U., V. Lakshmi, and E. G. Njoku. 2004. "Retrieval of soil moisture from passive and
886 active L/S band sensor (PALS) observations during the Soil Moisture Experiment in 2002
887 (SMEX02)." *Remote Sensing of Environment* 92 (4): 483-496.
- 888 Notarnicola, C., M. Angiulli, and F. Posa. 2006. "Use of radar and optical remotely sensed
889 data for soil moisture retrieval over vegetated areas." *IEEE Transactions on Geoscience and*
890 *Remote Sensing* 44 (4): 925-934.
- 891 Notarnicola, C., M. Angiulli, and F. Posa. 2008. "Soil Moisture Retrieval From Remotely
892 Sensed Data: Neural Network Approach Versus Bayesian Method." *IEEE Transactions on*
893 *Geoscience and Remote Sensing* 46 (2): 547-557.
- 894 Orth, R., R. D. Koster, and S. I. Seneviratne. 2013. "Inferring Soil Moisture Memory
895 from Streamflow Observations Using a Simple Water Balance Model." *Journal of*
896 *Hydrometeorology* 14 (6): 1773-1790.
- 897 Paloscia, S., G. Macelloni, and E. Santi. 2006. "Soil moisture estimates from AMSR-E
898 brightness temperatures by using a dual-frequency algorithm." *IEEE Transactions on*
899 *Geoscience and Remote Sensing* 44 (11): 3135-3144.
- 900 Paloscia, S., P. Pampaloni, S. Pettinato, and E. Santi. 2008. "A Comparison of Algorithms for
901 Retrieving Soil Moisture from ENVISAT/ASAR Images." *IEEE Transactions on Geoscience*
902 *and Remote Sensing* 46 (10): 3274-3284.
- 903 Paloscia, S., P. Pampaloni, S. Pettinato, and E. Santi. 2010. "Generation of soil moisture maps
904 from ENVISAT/ASAR images in mountainous areas: a case study." *International Journal*
905 *of Remote Sensing* 31 (9): 2265-2276.
- 906 Paloscia, S., S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, and A. Reppucci. 2013. "Soil

- 907 moisture mapping using Sentinel-1 images: Algorithm and preliminary validation.” *Remote*
908 *Sensing of Environment* 134: 234–248.
- 909 Pasolli, L., C. Notarnicola, L. Bruzzone, G. Bertoldi, S. Della Chiesa, V. Hell, G. Niedrist, et al.
910 2011. “Estimation of Soil Moisture in an Alpine Catchment with RADARSAT2 Images.”
911 *Applied and Environmental Soil Science* 2011: 1–12.
- 912 Pathe, C., W. Wagner, D. Sabel, M. Doubkova, and Basara J.B. 2009. “Surface Soil Moisture
913 Retrieval Using Envisat ASAR Global Mode Data.” *IEEE Transactions on Geoscience and*
914 *Remote Sensing* 47 (2): 194.
- 915 Penman, H. L. 1948. “Natural Evaporation from Open Water, Bare Soil and Grass.”
916 *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 193
917 (1032): 120–145.
- 918 Petterssen, S. 1968. *Introducción a la meteorología*. Nueva ciencia, nueva técnica. Espasa-Calpe.
- 919 Pierdicca, N., L. Pulvirenti, and C. Bignami. 2010. “Soil moisture estimation over vegetated
920 terrains using multitemporal remote sensing data.” *Remote Sensing of Environment* 114
921 (2): 440–448.
- 922 Pierdicca, N., L. Pulvirenti, C. Bignami, and F. Ticconi. 2013. “Monitoring Soil Moisture in an
923 Agricultural Test Site Using SAR Data: Design and Test of a Pre-Operational Procedure.”
924 *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (3):
925 1199–1210.
- 926 Priestley, C. H. B., and R. J. Taylor. 1972. “On the Assessment of Surface Heat Flux and
927 Evaporation Using Large-Scale Parameters.” *Monthly Weather Review* 100 (2): 81–92.
- 928 Ranney, K. J., J. D. Niemann, B. M. Lehman, T. R. Green, and A. S. Jones. 2015. “A method
929 to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data.”
930 *Advances in Water Resources* 76: 81–96.
- 931 Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. “Learning Internal Representations
932 by Error Propagation.” .
- 933 Salgado, H., L. Génova, B. Brisco, and M. Bernie. 2001. “Surface soil moisture estimation
934 in argentina using RADARSAT-1 imagery.” *Canadian Journal of Remote Sensing* 27 (6):
935 685–690.
- 936 Sánchez, N., M. Piles, J. Martínez-Fernández, M. Vall-llossera, L. Pipia, A. Camps, A. Aguasca,
937 F. Pérez-Aragüés, and C. M. Herrero-Jiménez. 2014. “Hyperspectral Optical, Thermal,
938 and Microwave L-Band Observations For Soil Moisture Retrieval at Very High Spatial
939 Resolution.” *Photogrammetric Engineering & Remote Sensing* 80 (8): 745–755.
- 940 Shi, J., J. Wang, A. Y. Hsu, P. E. O’Neill, and E. T. Engman. 1997. “Estimation of bare
941 surface soil moisture and surface roughness parameter using L-band SAR image data.”
942 *IEEE Transactions on Geoscience and Remote Sensing* 35 (5): 1254–1266.
- 943 Sobrino, J. A., B. Franch, C. Mattar, J. C. Jiménez-Muñoz, and C. Corbari. 2012. “A method
944 to estimate soil moisture from Airborne Hyperspectral Scanner (AHS) and ASTER data:
945 Application to SEN2FLEX and SEN3EXP campaigns.” *Remote Sensing of Environment*
946 117: 415–428.
- 947 Sousa, S., F. Martins, M. Alvimferraz, and M. Pereira. 2007. “Multiple Linear Regression
948 and Artificial Neural Networks Based on Principal Components to Predict Ozone
949 Concentrations.” *Environmental Modelling & Software* 22 (1): 97–103.
- 950 Stamenkovic, J., L. Guerriero, P. Ferrazzoli, C. Notarnicola, F. Greifeneder, and J. Thiran.
951 2017. “Soil Moisture Estimation by SAR in Alpine Fields Using Gaussian Process Regressor
952 Trained by Model Simulations.” *IEEE Transactions on Geoscience and Remote Sensing* 55
953 (9): 4899–4912.
- 954 Thibeault, M., J.M. Cáceres, D. Dadamia, A.G. Soldano, M. Uriburu Quirno, J.M. Guerrieri,
955 R. Edrosa, et al. 2015. “Spatial and temporal analysis of the monte buey saocom and snap
956 core site.” In *IGARSS 2015*, 969–971.
- 957 Thornthwaite, C. W. 1948. “An Approach toward a Rational Classification of Climate.”
958 *American Geographical Society* 38: 55–94.
- 959 Ulaby, F. T., R. K. Moore, and A. K. Fung. 1982. *Microwave Remote Sensing, Volumes 1, 2,*
960 *and 3*. Addison Wesley, Reading, MA.

- 961 Venturini, V. 2007. "Estimation of evaporative fraction and evapotranspiration from remotely
962 sensed data using a complementary relationship." PhD diss., Universidad nacional del litoral.
- 963 Venturini, V., C. Krepper, and L. Rodriguez. 2012. "Evapotranspiration Estimation Based on
964 the Complementary Relationships, Evapotranspiration - Remote Sensing and Modeling."
965 Dr. ayse i ed., Chap. 2. Intech.
- 966 Verheyne, W. H. 2004. *Agro-Climatic-Based Land Evaluation Systems*. Vol. II.
- 967 Wagner, W., S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, P. De Rosnay, A. Jann, et al. 2013.
968 "The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and
969 Emerging Applications." 22 (1): 5–33.
- 970 Wagner, W., G. Lemoine, and H. Rott. 1999. "A Method for Estimating Soil Moisture from
971 ERS Scatterometer and Soil Data." *Remote Sensing of Environment* 70: 191–207.
- 972 Watson, D.F., and G. M. Philip. 1985. "A Refinement of Inverse Distance Weighted
973 Interpolation." *Geoprocessing* (2): 315–327.
- 974 Weisberg, S. 2005. *Applied Linear Regression*. Third edit ed. Minneapolis, Minnesota: A John
975 Wiley & Sons, Inc., Publication.
- 976 Zhang, J., I. Becker-Reshef, and C. Justice. 2015. "Evaluation of the ASCAT surface soil
977 moisture product for agricultural drought monitoring in USA." *2015 IEEE International
978 Geoscience and Remote Sensing Symposium (IGARSS)* 669–672.
- 979 Zribi, M., N. Baghdadi, N. Holah, and O. Fafin. 2005. *New methodology for soil surface moisture
980 estimation and its application to ENVISAT-ASAR multi-incidence data inversion*. Technical
981 Report 3-4.