Universidad de Concepción
Dirección de Postgrado
Facultad de Agronomía - Programa de Doctorado en Ciencias de la Agronomía

Uso de Teledetección y Carbono Orgánico del suelo para la Evaluación de la Degradación del Bosque de *Nothofagus Obliqua* de Chile.

Tesis para optar al grado de Doctor en Ciencias de la Agronomía

LIZARDO MAURICIO REYNA BOWEN
CHILLAN-CHILE
2017

Profesor Guía: Erick Zagal Venegas, PhD.
Dpto. de Suelos y Recursos Naturales, Facultad de Agronomía
Universidad de Concepción

Uso de Teledetección y Carbono Orgánico del Suelo para la Evaluación del la Degradación del Bosque de *Nothofagus Obliqua* de Chile.

Aprobada por:

Erick Zagal Venegas _____
Ing. Agrónomo, M.Sc., PhD. Profesor Guía

Francis Dube _____
Ing. Forestal, M.Sc., PhD. Profesor Guía

Juan Barrera Berrocal _____
Ing. Agrónomo, PhD. Profesor co-guía

Gonzalo Silva Aguayo _____
Ing. Agrónomo, M.Sc., PhD Director de Programa

Agradecimientos

Primero quiero agradecer a los profesores Erick Zagal, Francis Dube y Juan A. Barrera, por su amable y esencial ayuda en la realización de este trabajo de investigación. Ha sido un placer conocerlos.

Expreso mi agradecimiento también al personal del laboratorio de suelos de la Facultad de Agronomía por su colaboración en el análisis de las muestras utilizadas en este estudio. Al proyecto FONDECYT No. 11121279 que brindó acceso a las parcelas experimentales utilizadas en este estudio.

A la Universidad Técnica de Manabí por la confianza en mi propósito de buscar nuevos conocimientos y experiencias con el objetivo de mejorar su nivel académico y aportar en mayor medida al desarrollo de la región.

A la Secretaría de Educación Superior, Ciencia y Tecnología (SENESCYT - Ecuador) por otorgarme una beca que permitió financiar mis estudios en la Universidad de Concepción - Chillán, Chile.

A mis familiares y amigos... han hecho que mis días en Chile sean inolvidables!

Tabla de Contenido

Lista de Figuras

Lista de cuadros

RESUMEN

El carbono (C) orgánico es un factor clave en la lucha contra el calentamiento global. Las mayores reservas de C en los ecosistemas terrestres se encuentran en los bosques y el suelo. Sin embargo, la creciente demanda de recursos derivados de los bosques en especial la madera junto con el amento de la superficie de plantaciones exóticas ponen en riesgo la conservación de los bosques. El bosque nativo de *Nothofagus obliqua* en Chile, es uno de los remanentes en América de Sur y se encuentra en constante amenaza debido a sobreexplotación de recursos naturales y servicios ecosistémicos que este provee como la extracción de madera y la gandería. La teledetección ha demostrado ser una eficaz y económica herramienta para la evaluación de los bosques a escalas regionales, nacionales y mundiales, permitiendo a través del análisis de imágenes de satélite cartografiar los cambias en la cobertura del suelo incluyendo los bosques.

En esta investigación intentamos establecer una metodología para el monitoreo del bosque de *Nothogagus obliqua* del Centro Sur de Chile en diferentes estados de degradación. El estudio se basó en el análisis de la reflectancia; una propiedad de la variable a estudiar que está en función de las condiciones en la que se encuentra, este caso la biomasa foliar y el C del suelo en diferentes estados de degradación del bosque. Para el estudio de C del suelo se utilizaron un total de 70 muestras de dos tipos de suelo (Andisol y Ultisol) con baja densidad aparente (< 1 g cm^{-3}) y alto contenido de C orgánico ($> 5\%$ en los primeros 5 cm de profundidad). Una parte de cada muestra fue utilizada para los análisis del carbono en el laboratorio por el método de combustión seca. Otra parte fue colocada en cajas petri para ser escaneadas con un radiómetro en el rango del Visible e Infrarrojo cercano (VIS-NIR) utilizando iluminación natural. El análisis de los datos fue realizado utilizando el lenguaje de programación Python junto con varias librería especializadas. Se utilizó el filtro Savitzky-Golay para eliminar el ruido aleatorio en las curvas espectrales. Para la calibración del modelo se utilizaron varios pre-tratamientos y se evaluaron de acuerdo al rendimiento que tuvieron en la regresión. La técnica de validación cruzada fue aplicada al conjunto de datos para determinar el óptimo número de variables latentes en la regresión y para evaluar la consistencia de los datos. La regresión por mínimos cuadrados parciales fue utilizada tanto para la validación

como para la calibración del modelo final. Resultados contradictorios fueron observados entre dos procedimientos. En la validación cruzada el coeficiente de determinación R^2 resultó negativo en la mayoría de los casos, lo que indicó que el conjunto de datos no es consistente y por tanto las curvas espectrales no pueden ser utilizadas para la calibración de un modelo. Mientras que cuando se evitó el proceso de validación cruzada y se calibró el modelo directamente los resultados indicaron una buena correlación (R^2 = 0.82 y $RMSE$ = 0.61%) entre los valores de carbono predichos por el modelo y los medidos en el laboratorio. Para calibrar el modelo, el conjunto de datos fue dividido en dos, el 80% para calibrar el modelo y el 20% para validarlo. El análisis posterior permitió concluir que: el proceso de validación cruzada debe utilizarse como técnica de exploración de la consistencia del conjunto de datos; sin embargo, en suelos con baja densidad aparente y utilizando iluminación natural, la correlación entre la reflectancia difusa y el carbono orgánico del suelo se puede distosionar a niveles que dificultad el análisis.

En esta investigación aplicamos un enfoque no espacial al análisis de la reflectancia de la biomasa foliar del bosque de *Nothofagus obliqua* para detectar cambios en la biomasa aérea del bosque causados por procesos de degradación. Para el efecto se evaluaron doce parcelas experimentales bajo distintos estados de degradación del bosque de segundo crecimiento ubicado en el Rango Andino de Chile. Se utilizó una serie temporal de imágenes del Satélite Landsat 8 (OLI) de las cuales se obtuvieron los siguientes productos: Índice Normalizado de Vegetación (NDVI por sus siglas en Ingles), la reflectancia del infrarrojo (NIR banda 5) y el infrarrojo de la vegetación (NIRv). Los resultados mostraron que en todos los productos los diferentes estados de degradación pueden generar valor similares de NDVI, NIR and NIRv. Sin embargo, aplicando un análisis de comparación múltiple con un 95% de probabilidad, fue posible establecer valores referenciales de cada una de las condiciones del bosque. Dichos valores, podrían ser comparados con los calculados en un periodo posterior al del estudio y determinar si ha existido algún procesos de degradación o recuperación del bosque.

Demostramos también que en terrenos de topografía irregular como en la pre-cordillera Andina, el efecto de la exposición de las pendientes tiene una influencia sobre el Índice de Área Foliar (LAI por sus sigla en Ingles) siendo mayor en el lado sur especialmente en la condición de bosque no intervenido o intacto. Además se confirman los problemas de sombra generados en las pendientes con exposición sur en las imágenes de satélite. Sin embargo,

mediante la comparación de las medias de los valores, se observa que este método tiene potencial para ser establecido como método de seguimiento de la evolución de este tipo de bosque. Por otra parte, nuestro estudio advierte de un posible sobreajuste en la calibración de modelos matemáticos generados para predecir el C del suelo por espectroscopia, cuando las fuentes de variación de la reflectancia no son controladas de forma adecuada. Estos resultados son de gran interés para la efectiva aplicación de la teledetección y la espectroscopia.

SUMMARY

Organic carbon (C) is a key factor in the fight against global warming. The largest reserves of C in terrestrial ecosystems are found in forests and soil. However, the increasing demand for forest-derived resources, especially timber along with the increase in the area of exotic plantations, threatens the conservation of forests. The native *Nothofagus obliqua* forest in Chile, is one of the remnants in South America and is under constant threat. Remote sensing has proven to be an effective and economical tool for the assessment of forests at regional, national and global scales. In this research we tried to establish a methodology for the monitoring of the *Nothogagus obliqua* forest in southern Chile under different levels of degradation.

The study was based on the reflectance analysis; a property of the variable to be studied that is function of the conditions of the target. Here we evaluated the forest aerial biomass and the soil organic C (SOC) under different levels of forest degradation. A total of 70 samples of two types of soil (Andisol and Ultisol) with low bulk density (<1 g cm $^{-3}$) and high organic C content (> 5% at 5 cm in depth). A portion of each sample was used for the carbon analysis in the laboratory using the dry combustion method. Another part was placed in petri dishes to be scanned with a spectroradiometer in the range of Visible and Near Infrared (VIS-NIR) using natural illumination. The data analysis was performed using the Python programming language along with several specialized libraries. The Savitzky-Golay filter was used to eliminate random noise in the spectral curves. For the calibration of the model several pre-treatments were used and they were evaluated according to the performance obtained in the regression. The cross-validation technique was applied to the data set to determine the optimal number of latent variables or factors in the regression and to evaluate the consistency of the data. Partial least squares regression was used for both validation and calibration of the final model. Contradictory results were observed between the two procedures. In the cross-validation, the determination coefficient R^2 was negative in most cases, which indicated that the data set is not consistent and therefore the spectral curves can not be used for the calibration of a model. However, when the cross-validation process was avoided and the model was calibrated directly, the results indicated a good correlation (R^2 = 0.82 and RMSE = 0.61%) between the predicted values of the model and those measured in the laboratory. To calibrate the model, the data set was divided into two, 80% to calibrate the model and 20% to validate

it. The subsequent analysis allowed to conclude that: the process of cross-validation should be used as a technique for exploring the consistency of the data set; however, in soils with low bulk density and using natural light the diffuse reflectance can be distorted to levels that the correlation with soil organic C is difficult to stablish.

In this research we applied a non-spatial approach to the analysis of the reflectance of the foliar biomass of the *Nothofagus obliqua* forest to detect changes in forest biomass caused by degradation processes. For this purpose, twelve experimental plots were evaluated under different stages of degradation of the second growth forest located in the Andean Range of Chile. A temporal series of images of the Landsat 8 Satellite (OLI) were used, from which the following products were obtained: Normalized Vegetation Index (NDVI), infrared reflectance (NIR band 5) and infrared vegetation (NIRv). The results showed that in all products the different states of degradation can generate similar values of NDVI, NIR and NIRv. However, applying a multiple comparison analysis with 95% probability, it was possible to establish reference values for each of the forest conditions. These values could be compared with those calculated in a period after the study and determine if there have been any processes of forest degradation or recovery processes.

We also demonstrated that in terrain with irregular topography, such as in the Andean Foothills, the effect of slope exposure has an influence on the Leaf Area Index (LAI) being greater on the south side especially in undisturbed or intact forest conditions. In addition, shadow problems are generated on slopes with southern exposure in the satellite images. However, comparing the means of the values, we observed that method has a potential to be established as a method of monitoring the evolution of this type of forest. On the other hand, our study warns us of a possible overfitting in model calibration generated to predict the soil C spectroscopy when the sources of reflectance variation are not adequately controlled. These results are of great interest for the effective application of remote sensing and spectroscopy.

CAPÍTULO 1
Introducción General

1.1 Introducción General

Los bosques proveen entre el 52 y 72% del total de la producción primaria neta (PPN) en ecosistemas terrestres (Fekete, et al., 2014), y almacenan 298 Gt de carbono en su biomasa (FAO, 2010). Estos transforman el CO_2 presente en la atmósfera en carbono orgánico a través de la fotosíntesis y de forma subsecuente, una parte es incorporada al suelo por descomposición, y el restante retorna a la atmósfera por la respiración de plantas y microorganismos del suelo (Melillo, et al., 1993; Patenaude, Milne, & Dawson, 2005). Este dinámico proceso, convierte al suelo en el reservorio más importante de los ecosistemas terretres (Stevenson & Cole, 1999). Sin embargo, los cambios en el uso del suelo especialmente por deforestación, están reduciendo las reservas de C en diferentes regiones del mundo (Batjes, 1996; Dorji, et al., 2014; Chhabra, Palria, & Dadhwal, 2003; Sharma, et al., 2014). Estudios recientes indican que la deforestación y degradación de los bosques son la segunda fuente de emisión de CO_2 a la atmósfera (van der Werf, et al., 2009), alcanzando cerca del 17% del total por año (UN-REDD, 2011).

El bosque templado de Chile se encuentra entre los 35 Hotspots de biodiversidad a nivel mundial, y ocupa la mayor área de este tipo de bosque en América del Sur (Miranda, et al., 2015). En la región de la Araucanía, la expansión de las plantaciones exóticas es la principal causa de la pérdida del bosque nativo. (Nahuelhual, et al., 2012; Aguayo, et al., 2009). El cambio de cobertura más evidente fue observado en el rango costero donde cerca del 47% del paisaje es ahora dominado por plantaciones exóticas forestales (Miranda, et al., 2015). Patrones similares pero de menor intensidad han sido reportados para el rango Andino, el cual conserva el 70% del bosque nativo con respecto al periodo de 1973-1987 (Miranda, et al., 2015). Los bosques brindan una efectiva forma de captura de C, por tanto, su dinámica espacial debe ser evaluada para evitar pérdidas abruptas.

La reducción de las emisiones de C en los bosques es un desafío internacional (McBratney, et al., 2014). El programa REDD+ (El programa REDD+ ha sido desarrollado por las partes de la Convención Marco de las Naciones Unidas sobre el Cambio Climático) ha definido la captura de C como factor de alta prioridad para mitigar el calentamiento global y ha implementado un mecanismo económico para mitigar los impactos de la deforestación y degradación de los bosques en los países en vía de desarrollo (UN-REDD, 2011). Los métodos convencionales de campo para calcular la deforestación, pueden causar errores significativos

en las estimaciones (Steininger, et al., 2001). En ese contexto, es esencial contar con un método que permita estimar las tasas de deforestación y degradación de los bosques con mayor exactitud.

Las reservas de C en los bosques, se obtienen usualmente extrapolando ecuaciones alométricas obtenidas de mediciones de campo (Gallaun, et al., 2010; FAO, 2012). La Teledetección permite generar mapas de bosques a diferentes escalas (Patenaude, Milne, & Dawson, 2005); los cuales son usados después para estimar la reserva de C en la biomasa verde. Sin embargo, algunos resultados inconsistentes reportados entre diferentes métodos, sugieren que se necesita mayor investigación en el tema (Gutierrez-Velez & Pontius, 2012; Kim, Sexton, & Townshend, 2015).

El C orgánico del suelo (COS) es un indicador clave de la calidad del suelo (Stevenson & Cole, 1999; Zagal, et al., 2012). El interés en investigar el COS está aumentando debido a su potencial para mejorar los suelos, algunas funciones ecosistémicas y mitigar el impacto del cambio climático (McBratney, et al., 2014). El manejo de los bosques puede incrementar las reservas de C en los suelos (Dube, et al., 2012; Lal, 2005). Sin embargo la relación entre el COS y los diferentes estados de bosque, permanece incierta (Kirsten, et al., 2015). Debido a las pequeñas variaciones en las reservas de C en los bosques, se dificulta establecer valores límites para definir la degradación (Morales-Barquero, et al., 2014). Por tanto, es esencial contar con estimaciones oportunas y de las reservas de C en los bosques para su manejo.

La espectroscopía de suelo ha probado ser una alternativa rápida y más económica para la estimación del COS frente a métodos tradicionales (Kuang & Mouazen, 2013; Bartholomeus, et al., 2008). Adicionalmente la reflectancia del suelo, permite obtener información de otras propiedades del suelo a través de funciones de pedotransferencia (Werban, et al., 2013; Minasny, et al., 2013; Chabrillat, et al., 2013). Esta tecnología puede reducir las incertidumbres en la cartografía del COS y reflejar de forma más precisa su distribución espacial (Köchy, Hiederer, & Freibauer, 2015).

Esta investigación se enfocó en utilizar la reflectancia espectral de la biomasa aérea y el suelo como herramientas para determinar signos de degradación del bosque nativo de *Nothofagus obliqua* del Centro Sur de Chile. Una serie temporal (2013-2016) de imágenes

Landsat 8 fue utilizada para el análisis de la biomasa. Para la estimación del C orgánico del suelo se aplicó la técnica de espectroscopía.

1.2 Ciclo del Carbono

El ciclo corto del C está dominado por la interacción entre la atmósfera, los océanos, el suelo, la vegetación, y la actividad humana. La deforestación y el cambio en el uso del suelo, han incrementado el flujo del CO_2 del suelo hacia la atmósfera. Sin embargo, la cantidad de C orgánico en el suelo es significativamente mayor la de la vegetación y la atmósfera juntos (Stevenson & Cole, 1999) (**Figura 1.1**).

Figura 1.1 Reservas de carbono e intercambio entre la atmósfera, la vegetación y el suelo. Gráfico adaptado de (Hillel, 2010)

La descomposición de residuos vegetales y animales, es la principal fuente de C en el suelo. Las plantas a través del proceso de fotosíntesis capturan el CO_2 de la atmósfera, que pasa a formar parte de su propio tejido dando origen a la biomasa vegetal. Esto se conoce como produccion primaria bruta (PPB) y es controlada por la capacidad del suelo para liberar nutrientes mediante la descomposición de la materia orgánica. En este complejo proceso, parte del C es devuelto a la atmósfera como CO_2 debido a la respiración de las plantas y de los

microorganismos del suelo, el restante se conoce como producción primaria neta (PPN). La materia orgánica que es incorporada al suelo, es aprovechada por los microorganismos como fuente de energía, como resultado, son liberados nutrientes necesarios para el crecimiento de las plantas, esencialmente nitrógeno en forma de amonio NH_{4+} y nitrato NO_{3-}, otros minerales como el S, P y K, quedan también disponibles para las plantas, este proceso se denomina mineralización. La asimilación del C mediante síntesis de las células microbianas se llama inmovilización.

El humus está compuesto por sustancias húmicas que liberan el C y otros elementos de forma paulatina, el rango de tiempo varía según el tipo de suelo y el clima, desde semanas en los Histosoles hasta cientos de años en caso de los Gelisoles debido a la baja actividad microbiana (Buol, et al., 2011). En suelos donde las condiciones de humedad y temperatura son favorables para la actividad microbiana (bosques tropicales), el flujo de C entre el suelo y la atmósfera puede ser balanceado, no obstante, en bosques boreales por ejemplo, el contenido de C puede ser cientos de veces mayor al de la atmósfera (Horwath, 2006).

1.3 Espectroscopía de suelo

La espectroscopía de suelo es una técnica para determinar algunas propiedades del suelo en base al análisis de información espectral (Fystro, 2002; Wenjun, et al., 2014; Gomez, Viscarra Rossel, & McBratney, 2008; Brunet, et al., 2007; Zheng, et al., 2016). Consiste en generar un modelo matemático empírico a partir de los datos de reflectancia de un conjunto de muestras de suelo (conjunto base). Este conjunto de muestras debe representar el rango de valores entre los cuales oscila la variable a predecir. De este conjunto base, se debe tener los datos tanto de la variable de interés (e.g. pH) y las curvas de reflectancia de cada una de las muestras de suelo para la calibración del modelo. Existen varios métodos para establecer la relación matemática entre la variable de interés y la reflectancia, entre ellos los de primera generación que incluyen regresión múltiple, regresión por componentes principales, regresión step-wise (Haenlein & Kaplan, 2004). La regresión por mínimos cuadrados parciales es un método de análisis de segunda generación que tiende ser utilizado cuando se cuenta con un conjunto grande de predictores (e.g. las curvas espectrales). Este último método es el más utilizado en el campo de la espectroscopía de suelo. En años reciente se han incorporado técnicas basadas en 'aprendizaje' o por su terminología en Ingles 'Machine Learning'(Vohland, et al., 2011).

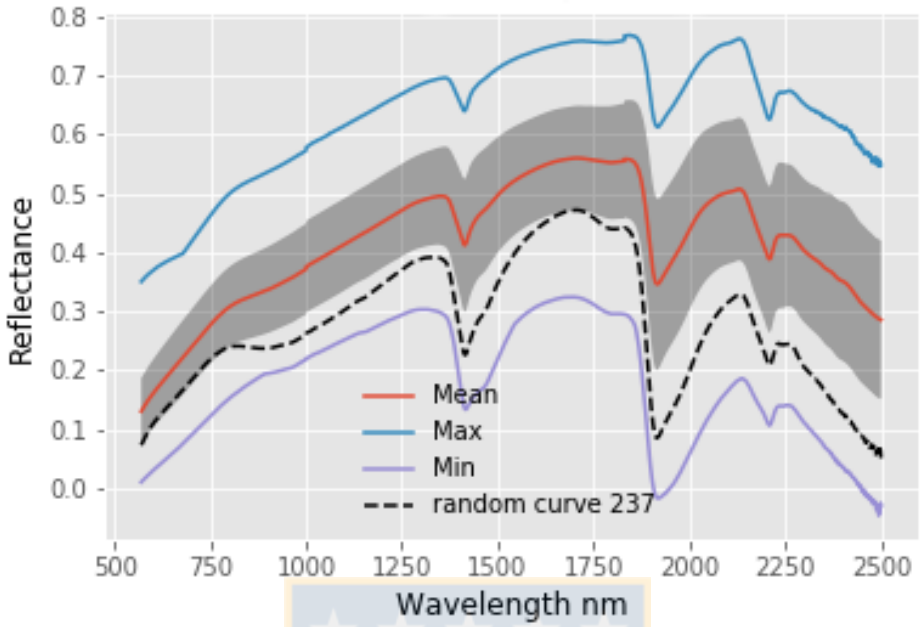

Figura 1.2 Ejemplo esquemático de la reflectancia del suelo con diferentes contenidos de C orgánico. A mayor contenido de C en la muestra, la reflectancia es menor. Así, la curva máxima 'Max' correspondería a la muestra de suelo con menor cantidad de C orgánico, y la curva 'Min' a la muestra con el mayor contenido de C orgánico. La curva 'Mean' representa la reflectancia media. La linea segmentada es una curva de una muestra seleccionada aleatoriamente del conjunto de datos. Este gráfico muestra el *espacio espectral* del conjunto base. A partir de esta información se genera un modelo matemático para predecir el C orgánico de la muestras de suelo que no fueron analizadas en el laboratorio. Este gráfico se elaboró a partir de datos no publicados generados en la Universidad de Córdoba de España.

La reflectancia espectral del suelo en el VIS-NIR (**Figura 1.2**)se ha utilizado para predecir el contenido de C (C) en diferentes tipos de suelo. Sarkhot, et al. (2011) reportaron altos valores de correlación para el C total y orgánico (R^2 = 0,85 y R^2 = 0,86, respectivamente) con un error de 5,33 g $\cdot$ kg^{-1} para C total y 2,88 g $\cdot$ kg^{-1} para C orgánico en un Entisol en la primera capa de 50 cm de suelo. Fontán, et al. (2010) encontraron valores de correlación más altos para el C inorgánico (R^2 = 0.76) que el C orgánico (R^2 = 0.67) en un vertisol hasta una profundidad de 90 cm. Otros estudios relacionados, muestran el desempeño de diferentes métodos multivariados en la calibración de los modelos (Vasques, Grunwald, & Sickman, 2008; Lucà, et al., 2017; Mouazen, et al., 2010). Los métodos clásicos para analizar el C del suelo requieren una preparación de la muestra rigurosa y productos químicos costosos (por ejemplo, el método por combustión seca). Como complemento (Sarkhot, et al., 2011) de estos

métodos, la espectroscopía del suelo puede utilizarse para construir modelos empíricos para predecir las propiedades del suelo a partir de datos espectrales. El potencial de esta técnica para predecir el C suelo se ha informado ampliamente (Fystro, 2002; Wenjun, et al., 2014; Gomez, Viscarra Rossel, & McBratney, 2008; Brunet, et al., 2007; Zheng, et al., 2016), y consideraciones importantes para la eficacia en la aplicación de la espectroscopía del suelo también han sido revisados por (Reeves III, et al., 2006; Knox, et al., 2015). Sin embargo, a nuestro conocimiento, no existen estudios que describan problemas potenciales como el ajuste excesivo en los procedimientos de calibración del modelo en un ejemplo de campo realista.

1.4 **Teledetección**

La teledetección es la ciencia que mide o infiere las características físicas de un objeto o medio, utilizando un sensor que se encuentra a una distancia del objeto o medio, es decir que no se encuentra en contacto directo con el objetivo (Woodhouse, 2013). Sin embargo, esta definición, aunque es correcta, resulta muy amplia, por ejemplo, el análisis de imágenes médicas, la espectroscopía, los sistemas de radares, el sondeo atmosférico, entre otros, son sistemas de teledetección. Por otra parte, en los últimos años, la aplicación de la teledetección junto con los sistemas de información geográfica (SIG), ha generado cierta confusión en cuanto a su definición.

Debido al vertiginoso crecimiento en la tecnología de adquisición de datos, como cámaras de alta resolución, sensores que pueden capturar regiones específicas del espectro y al desarrollo de tecnologías aero-espaciales y de posicionamiento geográfico, las imágenes satelitales han surgido como principal insumo para la teledetección, siendo el análisis de la cobertura la aplicación más difundida. Esto ha dado lugar al desarrollo de una disciplina que abarca básicamente cuatro áreas: procesamiento de imágenes, SIG, informática y teledetección. Es común encontrar textos y cursos, donde se refieren a estas cuatro áreas únicamente como teledetección, esto ha generado una percepción errada de su definición.

La teledetección puede dividirse en dos; pasiva y activa, diferenciándose esencialmente por la fuente de energía de la que obtiene la información. Los sensores satelitales multiespectrales son básicamente diseñados para teledetección pasiva, ya que capturan la energía que es reflejada (proveniente del sol) o emitida por los elementos de la superficie terrestre, nubes o

gases atmosféricos. Por ejemplo, los satélites Landsat, Spot, Rapideye, Ikonos y Worldview, utilizan sensores pasivos. Los sensores activos por su parte, emiten su propia energía, y a través de dispositivos receptores, capturan la señal de retorno. Adicionalmente, con sistemas muy precisos de mediciones de tiempo de retorno de la señal, calculan la distancia a la que se encuentra el objeto. Los sistemas radar (Radio Detection and Ranging) y LiDar (Light Detection and Ranging) son un ejemplo de sensores activos.

La aplicación de modelos matemáticos, es quizás la parte más compleja en teledetección. La razón principal es la abstracción conceptual de la interacción de la energía con los objetos naturales o medios, además de la relativa complejidad matemática necesaria para descifrar la información. Woodhouse (2013) menciona los modelos *directo* <<forward model>> e *inverso* <<inverse model>>.

1.4.1 Modelo directo

El modelo directo, relaciona las características físicas del objeto o medio con los datos registrados por un sensor (radiancia, temperatura de brillo, número digital etc.); en forma de píxeles en caso de las imágenes (2d), en forma de vector (1d) o datos puntuales. Este modelo requiere un acertado conocimiento de los procesos de interacción entre la radiación electromagnética y el objeto, así también, de las características del instrumento de medición, para determinar los posibles errores o ruidos en los datos. Podríamos plantear matemáticamente el modelo directo como la **Ecuación 1.1**

$$y = f(b, c, x) + e \tag{1.1}$$

La variable y está en función de f con los parámetros b, c, x, que representan las características del sensor, las variables medidas, y el estado del objeto respectivamente; e es el error aleatorio en los datos.

1.4.2 Modelo inverso

El modelo inverso, dada la **Ecuación 1.1**, intenta resolver dicha función para que a partir de y, encontrar x, es decir, el estado del objetivo. Esto resulta más complejo que el modelo directo. Una imagen satelital, registra la información en un determinado tamaño de píxel, el cual representa en una celda la cobertura de la superficie; que no necesariamente tiene

esa forma. A partir de esa información, más los datos tomados *in situ*, aplicamos el modelo directo y conseguimos los datos de reflectancia que corresponden con los datos, es decir, encontramos *y*.

Se debe tener presente que una imagen corresponde al estado del objeto en un momento determinado y bajo condiciones biofísicas específicas. Entonces, dada una imagen tomada en una fecha distinta a la de nuestros datos, podríamos determinar el estado del objeto? considerando que tenemos una función con infinitas soluciones, cual de todas sería la correcta?. Si analizamos hipotéticamente por ejemplo, el índice normalizado de vegetación (NDVI) de una masa vegetal, este puede responder de forma similar a la variación de múltiples variables biofísicas como el contenido de clorofila, el estrés hídrico, la fertilización, la infestación de plagas, el C orgánico de suelo, el clima, la exposición solar, entre otros. El objetivo principal de la teledetección, es a partir de una señal electromagnética, determinar el estado del objeto del cual proviene la señal. Desde luego, resulta complejo establecer una relación matemática apropiada entre los datos medidos y la realidad. El conocimiento avanzado de las propiedades físicas del objeto y de su interacción con la radiación, es fundamental para determinar la solución más aproximada.

1.5 Índice de vegetación

En teledetección, un índice de vegetación usualmente se utiliza para discriminar densidad, condición y vitalidad de la vegetación, aumentando el contraste entre la vegetación y el suelo (Jackson & Huete, 1991). Se calcula bajo el criterio de que la vegetación fotosintéticamente activa tiene máxima absorción de la radiación en el canal rojo debido a los pigmentos de clorofila, y tiene reflectividad máxima en el infrarrojo cercano debido a la estructura celular de la hoja (Karnieli, et al., 2001). La vegetación muerta o estresada, tiende a aumentar la reflectividad en el rojo y a disminuirla en el IR. Ese rango de transición en el espectro se conoce como borde rojo <<red edge>>. Otros tipos de cobertura como el suelo desnudo, tiene una reflectividad más uniforme en este rango.

El NDVI es quizás el índice más utilizado para evaluar la dinámica de la vegetación a partir de imágenes satelitales debido a su correlación con diferentes variables ecológicas como la densidad de biomasa, contenido de clorofila, vegetación fotosintéticamente activa, secuestro de C entre otras. Fue propuesto por (Rouse, et al., 1974), y consiste en una proporción

normalizada de la diferencia entre el canal rojo y el IR que resultan en valores entre -1 y 1, donde los valores cercanos a 0 representan <<sin vegetación verde>> y cercanos a 1 <<alta densidad de vegetación verde>> y los valores negativos corresponden a otros tipos de cobertura sin significado ecológico. Su interpretación biofísica es <<fracción de la radiación fotosintéticamente activa>>. Existen otros índices de vegetación con similar interpretación. (**Cuadro 1.1**).

Cuadro 1.1 Índices de vegetación y humedad derivados de imágenes satelitales. La reflectancia en el infrarrojo cercano es un componente esencial en el cálculo de los índices. El más utilizado en la evaluación de los bosques es el NDVI.

Nombre	Fórmula	Referencia
AFRI	$\frac{\rho_{nir}-0.5\rho_{2.1\mu m}}{\rho_{nir}+0.5_{\rho_{2.1\mu m}}}$	(Karnieli, et al., 2001)
LSWI	$\frac{\rho_{nir}-\rho_{1240nm}}{\rho_{nir}+\rho_{1240nm}}$	(Xiao, et al., 2004)
MSI	$\frac{\rho_{nir}}{\rho_{swir}}$	(Hunt Jr. & Rock, 1989)
NDVI	$\frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+\rho_{red}}$	(Rouse, et al., 1974)
NDWI	$\frac{\rho_{nir}-\rho_{1240nm}}{\rho_{nir}+\rho_{1240nm}}$	(Gao, 1996)
SR	$\frac{\rho_{nir}}{\rho_{red}}$	(Tucker, 1979)
SAVI	$\frac{\rho_{nir}-\rho_{red}}{\rho_{nir}+\rho_{red}+l}(1+l)$	(Huete, 1988)

AFRI = Aerosol free vegetation index

LSWI = Land surface water index

MSI = Moisture stress index

NDVI = Normalized difference vegetation index

NDWI = Normalized difference water index

SR = Simple ratio

SAVI = Soil adjusted vegetation index

1.6 Índice de área foliar (IAF)

El IAF or LAI (por sus siglas en Ingles), es un parámetro biofísico adimensional que está relacionado con varios procesos ecosistémicos de los bosques. El IAF representa la super-

fice foliar soportada por una determinada superfice de suelo (Aguirre-Salado, et al., 2011), y es un parámetro clave en procesos como la intercepción lumínica, intercambio de gases, fotosíntesis, ciclo de nutrientes y C e intercepción de la precipitación (Pu & Cheng, 2015). La determinación del IAF mediante métodos directos, resulta laborioso y costoso, lo que limita su utilización en grandes áreas (Aguirre-Salado, et al., 2011). Por otra parte, representar espacialmente el IAF, es un procedimiento crítico para una mejor comprención de los procesos físicos y biológicos en ecosistemas boscosos (Zheng & Moskal, 2009), y un factor clave para estudios multitemporales y a multiescala (local-regional). Adicionalemente el IAF puede considerarse como indicador relevante de estrés ambiental y de degradación de bosques (Dube & Stolpe, 2016).

La estimación del IAF a partir de datos espectrales, es una alternativa efectiva, rapida y de bajo costo. Sin embargo, ha sido un reto desde los inicios de la teledetección conseguir resultados exactos. No obstante, se han desarrollado varios métodos matemáticos y estadísticos que han permitido cartografiar el IAF a diferentes escalas para el estudio de los bosques (Aguirre-Salado, et al., 2011).

1.7 Programa Landsat

El programa Landsat tiene como misión, proveer de imágenes satelitales de mediana resolución a nivel global. Ha desarrollado desde 1972, una serie de satélites con el objetivo de adquirir imágenes de alta calidad que faciliten el estudio de la superficie terrestre. Posee la base de datos más grande del mundo de imágenes satelitales de mediana resolución, y las distribuye de forma gratuita a través de internet desde la pagina oficial del programa (**http://Landsat.usgs.gov/Landsat8.php**); el cual está dirigido por el U.S Geological Survey.

Landsat 8 es la continuación de la serie Landsat, desarrollado por la NASA desde 1972. Es el más reciente y mejorado satélite lanzado en febrero del 2013. El satélite opera en una órbita polar heliosincrónica a 705 km de altura en el ecuador. En esa órbita, sigue la ruta establecida en el world reference system (WRS-2), y captura escenas que representan 190x180 km en la superficie, aproximadamente 400 por día. El tiempo de revisita es de 16 días, y cruza el ecuador a las 10:11 a.m en el descenso (hora local). Las imágenes (se refiere a las escenas) son corregidas radiométricamente y registradas a una proyección cartográfica.

Para la captura de imágenes, Landsat 8 presenta notables mejorías respecto a la tecnología utilizada en sus antecesores TM y ETM+. Está equipado con dos sensores; Operational Land Imager (OLI) y Thermal Infrared Sensor (TIRS). El primero captura 9 bandas que cubren un rango del espectro entre 0.433-1.390 μm. El segundo captura dos bandas en el infrarrojo térmico. OLI incorpora dos nuevas bandas; coastal/aerosol como banda 1 para observaciones del color del océano y cirrus como banda 9 para detectar nubes poco densas. El principal cambio se efectuó en la banda 5 (infrarrojo cercano) para excluir la interferencia del vapor de agua atmosférico (**Figura 1.3**). Captura información térmica de la superficie en dos bandas, lo que representa un avance respecto a los anteriores sensores (USGS, 2015).

En OLI las bandas 1-7 y 9 tienen un tamaño de píxel de 30 metros y la banda pancromática (8) es de 15 metros. La información registrada en los píxeles, corresponde a números digitales, este valor representa una intensidad de brillo. Sin embargo, es necesario transformar esos números digitales a valores físicos como reflectancia o radiancia, para relacionarlos con las características de la superficie (**Ecuaciones 1.2** y **1.3**).

Radiancia:

$$L_\lambda = M_l Q_{cal} + A_l \tag{1.2}$$

L_λ = Radiancia $\quad \text{watt}/\text{m}^2\text{sr}\,\mu$
M_l = Factor de escala para radiancia multiplicativa para cada banda
A_l = Factor de escala para radiancia aditiva para cada banda
Q_{cal} = Valor del pixel

Reflectancia:

$$\rho'_\lambda = \frac{M_p Q_{cal} + A_p}{\sin\theta} \tag{1.3}$$

ρ_λ = Reflectacia TOA
M_p = Factor de escala para reflectancia multiplicativa
A_p = Factor de escala para reflectancia aditiva
Q_{cal} = Valor del pixel
θ = Ángulo senital solar

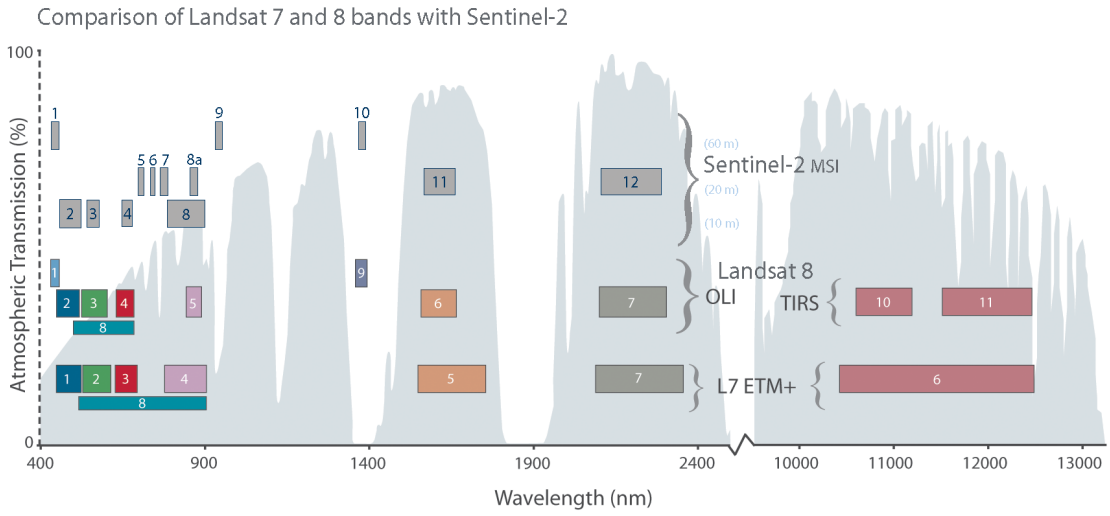

Figura 1.3 Comparación de las bandas de Landsat 7-8 y Sentinel-2. Los rectángulos representan el ancho de cada banda capturada por el sensor en los diferentes rangos del espectro electromagnético. La zona en gris, representa la transmisividad atmosférica en el espectro (USGS, 2015).

1.8 Modelo de Elevación Digital

Space Shuttle Topographic Mission (SRTM), fue una misión espacial liderada por la NASA en el año 2000. El objetivo fue utilizar la tecnología RADAR para obtener las elevaciones de todo el planeta, a través de pares de imágenes estereoscópicas. El resultado fue la generación del modelo digital de elevaciones (DEM por sus siglas en ingles) global. La resolución original de los datos es 1 arco-segundo, aproximadamente 30 metros. Sin embargo, para los sitios fuera del territorio de los Estados Unidos, los datos fueron escalados a 90 metros utilizando el método de interpolación de convolución cubica y distribuidos libremente a la comunidad científica a través de Internet. En el año 2014, el gobierno de los Estados Unidos decidió liberar los datos del proyecto SRTM en su resolución original para que puedan ser utilizado por la comunidad internacional con fines académicos y científicos.

Los datos del proyecto SRTM son administrados por el Centro U.S. Geological Survey's Earth Resources Observation and Science. Los datos están disponibles con diferentes niveles de procesamiento y en varios formatos, y pueden ser descargados desde el sito web **https://lta.cr.usgs.gov/SRTM**.

Un modelo de elevación digital es un arreglo matricial numérico, en el que cada celda representa el valor de la altura en un punto. Esta matriz se conoce como `raster`. Cada celda equivale a un pixel y es representado en el computador como una imagen en escala de grises. Un raster georeferenciado es un archivo al que cada celda está asignada a una coordenada geográfica. El formato más utilizado para la distribución de archivos raster georeferenciados es GeoTiff (Hengl & Evans, 2009).

Los modelos digitales de elevaciones son utilizados en diferentes áreas de la ciencia, sin embargo, son especialmente útiles en el estudio de los recursos naturales. A partir de un DEM se puede calcular la pendiente y la orientación, que son variables esenciales para realizar otros procesos como la extracción de cuencas y drenajes, modelar la radiación solar, analizar la erosión del suelo y deslizamientos de tierra. También son fundamentales para el estudio de la geomorfometría de la superficie de la tierra y la geología (Pike, Evans, & Hengl, 2009).

1.9 **Hipótesis**

- La reflectancia del suelo en el rango Visible e Infrarrojo cercano, permite generar un modelo estadístico capaz de predecir el carbono orgánico del suelo de muestras no analizadas del bosque de *Nothofagus obliqua* del Centro Sur de Chile.

- La variación interanual de la reflectancia de la biomasa foliar del bosque de *Nothofagus obliqua*, puede ser aplicada para determinar su estado de degradación.

1.10 **Objetivos**

- Determinar el carbono orgánico del suelo a través de espectroscopía de suelo en el bosque de *Nothofagus obliqua* del Centro Sur de Chile a tres profundidades, 0-5, 5-20 y 20-40 cm.

- Analizar desde el año 2013, la variación de la reflectancia de la biomasa foliar en los diferentes estados de degradación del bosque de *N. obliqua*.

- Evaluar un método estadístico basado en índices de vegetación para relacionar la variación de la biomasa foliar del bosque de *N. obliqua* y su efecto en la reflectancia.

1.11 **Referencias**

Aguayo, M., Pauchard, A., Azócar, G., & Parra, O. (2009). Cambio del uso del suelo en el centro sur de Chile a fines del siglo XX: Entendiendo la dinámica espacial y temporal del paisaje. *Revista chilena de historia natural*, *82*, 361 - 374. Available from http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0716-078X2009000300004&nrm=iso

Aguirre-Salado, C. A., Valdez-Lazalde, J. R., Ángeles-Pérez, G., M. de los Santos-Posadas, H., & Aguirre-Salado, A. (2011). Mapeo del índice de área foliar y cobertura arbórea mediante fotografía hemisférica y datos spot 5 hrg: regresión y k-nn. *Agrociencia*, *45*(1), 105-119.

Bartholomeus, H. M., Schaepman, M. E., Kooistra, L., Stevens, A., Hoogmoed, W. B., & Spaargaren, O. S. P. (2008). Spectral reflectance based indices for soil organic carbon quantification. *Geoderma*, *145*(1-2), 28–36. doi:10.1016/j.geoderma.2008.01.010

Batjes, N. -H. (1996). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, *47*, 151–163.

Brunet, D., Barthès, B. G., Chotte, J. -L., & Feller, C. (2007). Determination of Carbon and Nitrogen Contents in Alfisols, Oxisols and Ultisols from Africa and Brazil Using NIRS Analysis: Effects of Sample Grinding and Set Heterogeneity. *Geoderma*, *139*(1-2), 106–117. doi:10.1016/j.geoderma.2007.01.007

Buol, S. W., Southard, R. J., Graham, R. C., & McDaniel, P. A. (2011). *Soil Genesis and Classification*. (6th ed.). Wiley-Blackwell. Available from http://onlinelibrary.wiley.com/book/10.1002/9780470960622

Chabrillat, S., Ben-Dor, E., Rossel, R. A. V., Demattê, José, & M, A. (2013). Quantitative Soil Spectroscopy. *Applied and Environmental Soil Science*, *2013*, e616578. doi:10.1155/2013/616578

Chhabra, A., Palria, S., & Dadhwal, V. -K. (2003). Soil organic carbon pool in Indian forests. *Forest Ecology and Management*, *173*(1-3), 187–199. doi:10.1016/S0378-1127(02)00016-6

Dorji, T., Odeh, I. O. -A., Field, D. J., & Baillie, I. C. (2014). Digital soil mapping of soil organic carbon stocks under different land use and land cover types in montane ecosystems, Eastern Himalayas. *Forest Ecology and Management*, *318*, 91–102. doi:10.1016/j.foreco.2014.01.003

Dube, F., Espinosa, M., Stolpe, N. B., Zagal, E., Thevathasan, N. V., & Gordon, A. M. (2012). Productivity and carbon storage in silvopastoral systems with Pinus ponderosa and

Trifolium spp., plantations and pasture on an Andisol in Patagonia, Chile. *Agroforestry Systems*, *86*(2), 113–128.

Dube, F. & Stolpe, N. B. (2016). SOM and Biomass C Stocks in Degraded and Undisturbed Andean and Coastal Nothofagus Forests of Southwestern South America. *Forests*, *7*(12), 320. doi:10.3390/f7120320

FAO (2010). *Global Forest Resources Assessment*. (Technical report). Food and Agriculture Organization of the United Nations - Forestry Department.

_____ (2012). *Manual for building tree volume and biomass allometric equations*. (Technical report). Food and Agriculture Organization of the United Nations - Forestry Department.

Fekete, I., Kotroczó, Z., Varga, C., Nagy, P. T., Várbíró, G., Bowden, R. D., … Lajtha, K. (2014). Alterations in forest detritus inputs influence soil carbon concentration and soil respiration in a Central-European deciduous forest. *Soil Biology and Biochemistry*, *74*, 106–114. doi:10.1016/j.soilbio.2014.03.006

Fontán, J. M., Calvache, S., López-Bellido, R. J., & López-Bellido, L. (2010). Soil Carbon Measurement in Clods and Sieved Samples in a Mediterranean Vertisol by Visible and Near-Infrared Reflectance Spectroscopy. *Geoderma*, *156*(3-4), 93–98. doi:10.1016/j.geoderma.2010.02.001

Fystro, G. (2002). The Prediction of C and N Content and Their Potential Mineralisation in Heterogeneous Soil Samples Using Vis-NIR Spectroscopy and Comparative Methods. *Plant and Soil*, *246*(2), 139–149. doi:10.1023/A:1020612319014

Gallaun, H., Zanchi, G., Nabuurs, G. -J., Hengeveld, G., Schardt, M., & Verkerk, P. J. (2010). EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. *Forest Ecology and Management*, *260*(3), 252 - 261. doi:http://dx.doi.org/10.1016/j.foreco.2009.10.011

Gao, B. -C. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257-266. doi:10.1016/S0034-4257(96)00067-3

Gomez, C., Viscarra Rossel, R. A., & McBratney, A. B. (2008). Soil Organic Carbon Prediction by Hyperspectral Remote Sensing and Field Vis-NIR Spectroscopy: An Australian Case Study. *Geoderma*, *146*(3-4), 403–411. doi:10.1016/j.geoderma.2008.06.011

Gutierrez-Velez, V. H. & Pontius, R. G. J. R. (2012). Influence of carbon mapping and land change modelling on the prediction of carbon emissions from deforestation. *Environmental Conservation*, *39*(04), 325–336. doi:10.1017/S0376892912000173

Haenlein, M. & Kaplan, A. M. (2004). A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics*, *3*(4), 283–297. doi:10.1207/s15328031us0304_4

Hengl, T. & Evans, I. -S. (2009). Chapter 2 Mathematical and Digital Models of the Land Surface. In Hengl, T. & Reuter, H. I. (Eds.) GeomorphometryConcepts, Software, Applications. (Vol. 33, pp. 31 - 63). Elsevier. doi:http://dx.doi.org/10.1016/S0166-2481(08)00002-0

Horwath, W. (2006). Carbon cycling and formation of soils organic matter. In Eldor, A. P. (Ed.) Soil Microbiology, Ecology and Biochemistry. (pp. 303–364). Elsevier Inc.

Huete, A. -R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295-309. doi:10.1016/0034-4257(88)90106-X

Hunt Jr., E. -R. & Rock, B. -N. (1989). Detection of changes in leaf water content using Near- and Middle-Infrared reflectances. *Remote Sensing of Environment*, *30*(1), 43-54. doi:10.1016/0034-4257(89)90046-1

Jackson, R. D. & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, *11*(3), 185–200.

Karnieli, A., Kaufman, Y. J., Remer, L., & Wald, A. (2001). {AFRI} — aerosol free vegetation index. *Remote Sensivng of Environment*, *77*(1), 10 - 21. doi:http://dx.doi.org/10.1016/S0034-4257(01)00190-0

Kim, D. -H., Sexton, J. O., & Townshend, J. R. (2015). Accelerated Deforestation in the Humid Tropics from the 1990s to the 2000s. *Geophysical Research Letters*, n/a–n/a. doi:10.1002/2014GL062777

Kirsten, M., Kaaya, A., Klinger, T., & Feger, K. -H. (2015). Stocks of soil organic carbon in forest ecosystems of the Eastern Usambara Mountains, Tanzania. *{CATENA}*, (0), -. doi:http://dx.doi.org/10.1016/j.catena.2014.12.027

Knox, N. M., Grunwald, S., McDowell, M. L., Bruland, G. L., Myers, D. B., & Harris, W. G. (2015). Modelling Soil Carbon Fractions with Visible Near-Infrared (VNIR) and Mid-Infrared (MIR) Spectroscopy. *Geoderma*, *239*, 229–239. doi:10.1016/j.geoderma.2014.10.019

Köchy, M., Hiederer, R., & Freibauer, A. (2015). Global distribution of soil organic carbon – Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. *SOIL*, *1*(1), 351–365. doi:10.5194/soil-1-351-2015

Kuang, B. & Mouazen, A. M. (2013). Non-biased prediction of soil organic carbon and total nitrogen with vis–NIR spectroscopy, as affected by soil moisture content and texture. *Biosystems Engineering*, *114*(3), 249–258. doi:10.1016/j.biosystemseng.2013.01.005

Lal, R. (2005). Forest soils and carbon sequestration. *Forest Ecology and Management*, *220*(1–3), 242 - 258. doi:http://dx.doi.org/10.1016/j.foreco.2005.08.015

Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G., & Buttafuoco, G. (2017). Effect of Calibration Set Size on Prediction at Local Scale of Soil Carbon by Vis-NIR Spectroscopy. *Geoderma*, *288*, 175 – 183. doi:http://dx.doi.org/10.1016/j.geoderma.2016.11.015

McBratney, A. B., Stockmann, U., Angers, D. A., Minasny, B., & Field, D. J. (2014). Challenges for Soil Organic Carbon Research. In Hartemink, A. E. & McSweeney, K. (Eds.) Soil Carbon. (pp. 3-16). Springer International Publishing. doi:10.1007/978-3-319-04084-4_1

Melillo, Jerry. M., McGuire, A. David., Kicklighter, D. W., Moore, B., Vorosmarty, C. J., & Schloss, A. L. (1993). Global climate change and terrestrial net primary production. *Nature*, *363*, 234–240.

Minasny, B., Whelan, B. M., Triantafilis, J., & McBratney, Alex. B. (2013). Pedometrics Research in the Vadose Zone—Review and Perspectives. *Vadose Zone Journal*, *12*(4). doi:10.2136/vzj2012.0141

Miranda, A., Altamirano, A., Cayuela, L., Pincheira, F., & Lara, A. (2015). Different Times, Same Story: Native Forest Loss and Landscape Homogenization in Three Physiographical Areas of South-Central of Chile. *Applied Geography*, *60*(0), 20 – 28. doi:http://dx.doi.org/10.1016/j.apgeog.2015.02.016

Morales-Barquero, L., Skutsch, M., Jardel-PelÃ¡ez, E. J., Ghilardi, A., Kleinn, C., & Healey, J. R. (2014). Operationalizing the Definition of Forest Degradation for REDD+, with Application to Mexico. *Forests*, *5*(7), 1653–1681. doi:10.3390/f5071653

Mouazen, A. M., Kuang, B., De Baerdemaeker, J., & Ramon, H. (2010). Comparison among Principal Component, Partial Least Squares and Back Propagation Neural Network Analyses for Accuracy of Measurement of Selected Soil Properties with Visible and near Infrared Spectroscopy. *Geoderma*, *158*(1), 23–31. doi:10.1016/j.geoderma.2010.03.001

Nahuelhual, L., Carmona, A., Lara, A., Echeverría, C., & González, M. E. (2012). Land-cover change to forest plantations: Proximate causes and implications for the landscape in south-central Chile. *Landscape and Urban Planning*, *107*(1), 12 - 20. doi:http://dx.doi.org/10.1016/j.landurbplan.2012.04.006

Patenaude, G., Milne, R., & Dawson, T. P. (2005). Synthesis of remote sensing approaches for forest carbon estimation: reporting to the Kyoto Protocol. *Environmental Science & Policy*, *8*(2), 161–178. doi:10.1016/j.envsci.2004.12.010

Pike, R. -J., Evans, I. -S., & Hengl, T. (2009). Chapter 1 Geomorphometry: A Brief Guide. In Hengl, T. & Reuter, H. I. (Eds.) GeomorphometryConcepts, Software, Applications. (Vol. 33, pp. 3 - 30). Elsevier. doi:http://dx.doi.org/10.1016/S0166-2481(08)00001-9

Pu, R. & Cheng, J. (2015). Mapping forest leaf area index using reflectance and textural information derived from WorldView-2 imagery in a mixed natural forest area in Florida, {US}. *International Journal of Applied Earth Observation and Geoinformation*, *42*, 11 - 23. doi:http://dx.doi.org/10.1016/j.jag.2015.05.004

Reeves III, J. B., Follett, R. F., McCarty, G. W., & Kimble, J. M. (2006). Can Near or Mid-Infrared Diffuse Reflectance Spectroscopy Be Used to Determine Soil Carbon Pools?. *Communications in Soil Science and Plant Analysis*, *37*(15-20), 2307–2325. doi:10.1080/00103620600819461

Rouse, J. W., Jr, Deering, D. W., Schell, J. A., & Harlan, J. C. (1974). *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*. (Technical report, no. RSC 1978-2). Greenbelt, MD: Remote Sensing Center A&M University,College Station.

Sarkhot, D. V., Grunwald, S., Ge, Y., & Morgan, C. L. S. (2011). Comparison and Detection of Total and Available Soil Carbon Fractions Using Visible/near Infrared Diffuse Reflectance Spectroscopy. *Geoderma*, *164*(1-2), 22–32. doi:10.1016/j.geoderma .2011.05.006

Sharma, V., Hussain, S., Sharma, K. -R., & Arya, V. M. (2014). Labile carbon pools and soil organic carbon stocks in the foothill Himalayas under different land use systems. *Geoderma*, *232-234*, 81–87. doi:10.1016/j.geoderma.2014.04.039

Steininger, M. K., Tucker, C. J., Townshend, J. R. G., Killeen, T. J., Desch, A., Bell, V., & Ersts, P. (2001). Tropical deforestation in the Bolivian Amazon. *Environmental Conservation*, *null*, 127–134. doi:10.1017/S0376892901000133

Stevenson, F. -J. & Cole, M. -A. (1999). *Cycles of Soils*. (2 ed.). New York: Wiley.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127 - 150. doi:http://dx.doi.org/10 .1016/0034-4257(79)90013-0

UN-REDD (2011). *The UN-REDD Programme Strategy*. (Technical report). The United Nations Collaborative Programme on Reducing Emissions form Deforestation and Forest Degradation in Developing Countries.

USGS (2015). *Landsat 8 Data Users Handbook*. (version 1 ed.). U.S. Geological Survey.

Vasques, G. M., Grunwald, S., & Sickman, J. O. (2008). Comparison of Multivariate Methods for Inferential Modeling of Soil Carbon Using Visible/near-Infrared Spectra. *Geoderma*, *146*(1-2), 14–25. doi:10.1016/j.geoderma.2008.04.007

Vohland, M., Besold, J., Hill, J., & Fründ, H. -C. (2011). Comparing different multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. *Geoderma*, *166*(1), 198–205. doi:10.1016/j.geoderma .2011.08.001

van der Werf, G. R., Morton, D. C., DeFries, R. S., Olivier, J. G. J., Kasibhatla, P. S., Jackson, R. B., … Randerson, J. T. (2009). CO2 emissions from forest loss. *Nature Geosci*, *2*(11), 737–738. Available from http://dx.doi.org/10.1038/ngeo671 http://www .nature.com/ngeo/journal/v2/n11/suppinfo/ngeo671_S1.html

Wenjun, J., Zhou, S., Jingyi, H., & Shuo, L. (2014). In Situ Measurement of Some Soil Properties in Paddy Soil Using Visible and Near-Infrared Spectroscopy. *PLOS ONE*, *9*(8), e105708. doi:10.1371/journal.pone.0105708

Werban, U., Bartholomeus, H., Dietrich, P., Grandjean, G., & Zacharias, S. (2013). Digital Soil Mapping: Approaches to Integrate Sensing Techniques to the Prediction of Key Soil Properties. *Vadose Zone Journal*, *12*(4). doi:10.2136/vzj2013.10.0178

Woodhouse, I. H. (2013). *Thirteen shorts chapters on Remote Sensing*. Speckled Press.

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., … Ojima, D. (2004). Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data. *Remote Sensing of Environment*, *91*(2), 256-270. doi:10.1016/j.rse.2004.03.010

Zagal, E., Muñoz, C., Espinoza, S., & Campos, J. (2012). Soil profile distribution of total C content and natural abundance of 13C in two volcanic soils subjected to crop residue burning versus crop residue retention. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science*, *62*(3), 263-272. doi:10.1080/09064710.2011.608707

Zheng, G. & Moskal, L. M. (2009). Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors*, *9*(4), 2719. doi:10.3390/s90402719

Zheng, G., RYU, D., JIAO, C., & HONG, C. (2016). Estimation of Organic Matter Content in Coastal Soil Using Reflectance Spectroscopy. *Pedosphere*, *26*(1), 130–136. doi :10.1016/S1002-0160(15)60029-7

CAPÍTULO 2
Potential Model Overfitting in Predicting Soil Carbon Content by Visible and Near-Infrared Spectroscopy

Publicado en **www.mdpi.com/journal/applsci** *Appl. Sci.* **2017**, 7, 708; doi:10.3390/app7070708

Potential Model Overfitting in Predicting Soil Carbon Content by Visible and Near-Infrared Spectroscopy

Lizardo Reyna[1,†] , Francis Dube[2] , Juan A. Barrera[1] and Erick Zagal[1,*]

1 Department of Soils and Natural Resources, Faculty of Agronomy, Universidad de Concepción, Vicente Méndez 595, Casilla 537, Chillán 3812120, Chile; lreyna@udec.cl, jbarrera@udec.cl

2 Department of Silviculture, Faculty of Forest Sciences, Universidad de Concepción, Victoria 631, Casilla 160-C, Concepción 4030000, Chile; fdube@udec.cl

* Correspondence: ezagal@udec.cl; Tel.: +56-42-2208853

† Current address: Facultad de Ingeniería Agrícola, Universidad Técnica de Manabí, Casilla 82, Lodana, Manabí, Ecuador; lreyna@utm.edu.ec

Abstract. *Soil spectroscopy is known as a rapid and cost-effective method for predicting soil properties from spectral data. The objective of this work was to build a statistical model to predict soil carbon content from spectral data by partial least squares regression using a limited number of soil samples. Soil samples were collected from two soil orders (Andisol and Ultisol), where the dominant land cover is native Nothofagus forest. Total carbon was analyzed in the laboratory and samples were scanned using a spectroradiometer. We found evidence that the reflectance was influenced by soil carbon content, which is consistent with the literature. However, the reflectance was not useful for building an appropriate regression model. Thus, we report here intriguing results obtained in the calibration process that can be confusing and misinterpreted. For instance, using the Savitzky–Golay filter for pre-processing spectral data, we obtained* $R^2 = 0.82$ *and root-mean-squared error (RMSE)* $= 0.61\%$ *in model calibration. However, despite these values being comparable with those of other similar studies, in the cross-validation procedure, the data showed an unusual behavior that leads to the conclusion that the model overfits the data. This indicates that the model should not be used on unobserved data.*

Keywords: chemometrics; SOC; spectral diffuse reflectance; partial least squares regression; cross-validation

2.1 Introduction

Soil total carbon (TC) is composed of organic (all organic components mainly derived from the decomposition of plants and animals; and including living organisms) and inorganic (non-living C, typically as carbonates) carbon forms. Due to the short-term cycle of soil organic

carbon (SOC) and its key role for soil functions, the quantitative evaluation of SOC is essential for determining a suitable management practice to conserve or increase soil carbon stock [1–4]. Monitoring SOC over large areas or long periods of time requires analysis of substantial numbers of samples which can be labor-intensive and expensive. Under those circumstances, the soil spectroscopy technique is an effective method to predict SOC rapidly at minimal cost [5 and 6]. Soil spectroscopy uses the visible and near-infrared (VIS-NIR, 400–2500 nm) and mid-infrared (2500–25,000 nm) spectral reflectance to infer soil properties from a scanned sample [7]. This technique has been used mainly under laboratory conditions, but it can also be applied in the field for a specific site or in an instrument setup for ongoing scanning [6].

Spectral reflectance of soil in the VIS-NIR has been used to predict soil C in different soil types. Sarkhot et al. [8] reported high correlation values for total and organic C (R^2 = 0.85 and R^2 = 0.86, respectively) with an error of 5.33 g·kg^{-1} for total C and 2.88 g·kg^{-1} -15for organic C in an entisol -10i n the first 50-cm layer of soil. Fontán et al. [9] found higher correlation values for inorganic C (R^2 = 0.76) than organic C (R^2 = 0.67) in a vertisol to a depth of 90 cm. In addition, successful predictions of soil carbon fractions have been made by VIS-NIR spectroscopy [10 and 11], suggesting this technique as a reliable alternative for assessing the impact of land use change on soil carbon pools. Other related studies show the performance of different multivariate methods in calibrating models [12–14]. Classical methods for analyzing soil C require a rigorous sample preparation and expensive chemical supplies (e.g., dry combustion). As a complement to these methods, soil spectroscopy can be used to build empirical models to predict soil properties from spectral data. The potential of this technique to predict soil C has been extensively reported [15–19], and important considerations for the effective application of soil spectroscopy have been also reviewed by Reeves III [7]. However, to our knowledge, there are no studies that describe potential issues such as overfitting in model calibration procedures in a realistic field example.

In general, better correlations and fewer errors have been reported for sieved (<2 mm) and dry (air/oven) samples [9 and 20]. Brunet et al. [16] concluded that grinded samples notably improve the predictions of total soil carbon. In contrast, Fystro [15] found that the quality of prediction was not benefited by grinding samples. Soil darkness is the most evident effect of water on spectral variability, but strong effects occur in the infrared spectral reflectance

where the wavelengths of 1450 nm and 1950 nm are absorption bands [21 and 22]. Because water molecules in soil are dispersed, broad bands are usually seen at these wavelengths [21]. Andisols have special properties such as low density and relatively high amounts of organic carbon within the soil profile [23], which have an important effect on soil reflectance.

Before model calibration, a pre-processing step is needed to reduce random noise and dimensionality of the spectral data [12]. The purpose of pre-processing techniques in spectroscopy is to eliminate the portion of reflectance that does not come from the desired properties of the target. In soil spectroscopy, the spectra variability includes incident light reflected in multiple directions (diffuse reflectance) due to soil roughness, soil aggregates, soil structure, and particle size. These physical properties scatter the incident energy in many directions. In a complex soil sample, the variation of reflectance is not strongly wavelength-dependent (Lorentz–Mie scattering) and can be observed as a multiplicative effect [24]. For this reason, filtering data is an essential process before analysis [25 and 26]. Spectral correction and spectral derivatives are two categories of pre-processing techniques. For the first category, the most used methods are multiplicative scatter correction (MSC) and the standard normal variate (SNV). For the second category, Norris–Williams (NW) and Savitzky–Golay (SG) derivatives filters are widely applied to spectral data. In this study, we performed the pre-processing step using the SG filter.

The aim of the present work was to evaluate the potential of soil NIR spectroscopy to predict total organic soil carbon using a limited number of soil samples collected from the native temperate forest. We used the partial least square regression (PLSR) method to find the statistical relation between the analyzed samples and spectral data.

2.2 Materials and Methods

2.2.1 Site Description

The study took place in the Biobio and Araucanía Regions of Chile(**Figure 2.1**). Two soil orders were selected for the experiment: (1) Andisol, located in the Andean range (medial, amorphic, mesic, Typic Hapludands) [23]; and (2) Ultisol, which is located in the coastal range (very fine, mixed, semiactive, mesic Typic Paleudults) [27]. In both experimental sites the land cover is dominated by *Nothofagus obliqua* and *Nothofagus nervosa*. Andisol is

located at latitude 36°48"S, longitude 71°38"W, and Ultisol is located at latitude 37°46" S, longitude 72°58" W. These soils are derived from volcanic ash due to the intense volcanic activity in the Quaternary [28] and are rich in organic carbon [23]. This parent material does not contain carbonates. Furthermore, in the field, these soils do not react to the hydrochloric acid test. Therefore, TC and organic carbon (OC) were equivalents.

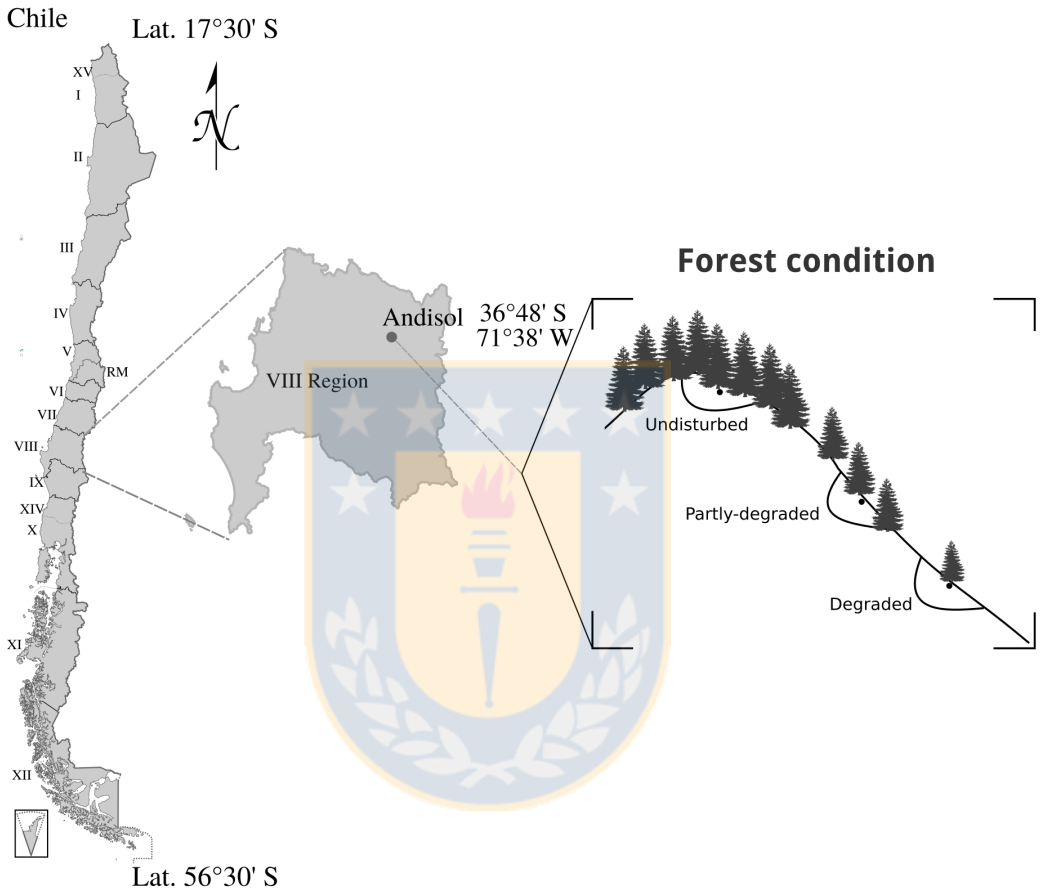

Figure 2.1 Location of the study sites. Two regions were selected; Region VIII (Biobio) and Region IX (Araucanía) which are located in the Andean and coastal ranges of Chile, respectively. In the box, a representation of the different *Nothofagus obliqua* forest conditions is shown for where the soil samples were extracted.

2.2.2 Soil Sampling, Total Carbon Analysis and Spectral Measurement

A total of 70 samples were taken from the sites and three soil layers were sampled at depths of 0–5, 5–20, and 20–40 cm. The sampling strategy was selected to represent different conditions of the native forest, from undisturbed to degraded (**Figure 2.1**, for details see [27]).

The samples were kept in plastic bags and stored at −4 °C for later analysis. Later, the samples were air-dried and put through a 2-mm sieve in order to minimize spectral variation due to fresh organic matter, soil moisture and soil aggregates. One portion of the sample was reserved for analysis and the other for spectral scanning. Total carbon (TC) was obtained by dry combustion method [29] using an elemental analyzer (model Leco CN-2000, macro-analyzer, LECO Corporation, Saint Joseph, Michigan, USA). All TC values were combined in a unique data set that contained the entire range of all depths and both soil types. Samples for scanning were oven-dried at 60 °C for 48 h to standardize the moisture level, then were set in petri dishes (60 × 15 mm) and flattened with a spatula. Thirty-six Andisol samples were taken and 34 Ultisol. Each depth was represented by twelve samples, except for the depths of 5–20 and 20–40 cm for Ultisol, where 11 samples were selected for each one.

Spectral reflectance was measured in the VIS-NIR range (350–1075 nm) at 1-nm intervals using a spectroradiometer (HandHeld 2: Hand-held VNIR, ASD-FieldSpec®, ASD, Boulder, Colorado, USA). We used this interval to detect possible sharp peaks in the spectral. The sensor was located vertically at a distance of 5 cm from the soil sample and it was fixed in a tripod, then a Spectralon panel was used as white reference before sample scans. Outdoor scanning was performed using the natural source of light at 3:00 p.m. on a sunny day. A dark background was used to minimize the influence of ambient light, and to record the soil reflectance provided by natural illumination. Ten consecutive scans were averaged and recorded for each sample.

2.2.3 Spectral Pre-Processing and Reflectance Analysis

Before model calibration we performed several pre-processing configurations using the Savitzky–Golay filter in order to remove both additive and multiplicative effects in the spectral data [24]. This digital filter smooths the data while the original characteristics are minimally affected [30 and 31]. Input of three parameters are needed: window size, polynomial order, and derivative order e.g., (5, 1, 1). Window size must be set in the form 2M + 1, where M is half of the window size [30]. This value is the number of consecutive points that will be used by least-squares smoothing, and for the end points of the curve this window can be asymmetric. If derivative order is given, the data is transformed. In the original publication, a detailed explanation of this filter can be found [25]. We also evaluated two additional functions, the Log(1/R) (where R is reflectance) to transform reflectance (R) into apparent

absorbance [5], and mean-center function (centering) making the spectral curves to fluctuate around zero [32]. We carried out fourteen pretreatments to the spectral data including smooth and transformation by first and second derivative orders in the Savitzky–Golay algorithm.

Spectral data can be decomposed into a small number of explanatory variables containing huge amount of variance [33]. Thus a principal component analysis (PCA) was performed on the spectral data to examine its structure and identify outliers.

The relationship between soil carbon content and reflectance is essentially inverse, i.e., when soil carbon increases the reflectance decreases across the spectra [19, 34 and 35]. To prove this postulation, we conducted an analysis for each soil type in order to observe the influence of soil carbon content on reflectance. Thus, the spectral curves corresponding to each sampled depth were averaged and plotted for visual analysis. In the same way, TC values were also averaged to be related with the spectral reflectance (**Table 2.1**).

Table 2.1 Average of soil total carbon (TC) of each depth sampled for the two soil orders (Andisol and Ultisols. In brackets the numbers of samples averaged

Soil order	Depth (cm)	Average	Standard Deviation
Andisol	0-5	6.4 (12)	2.42
	0-5	4.4 (12)	1.65
	0-5	3.0 (12)	1.29
Ultisol	0-5	5.3 (12)	1.55
	0-5	4.5 (12)	1.26
	0-5	2.7 (12)	1.78

2.2.4 Cross-Validation and Partial Least Squares Regression

Partial least squares regression or PLSR is a multivariate technique based on the combination of dimensionality reduction similar to PCA and multiple linear regression (MLR) [36]. Unlike PLSR, the latent variables or factors (equivalent to principal components in PCA) are calculated taking into account the response variable [33]. With these latent variables a predictor matrix **X** is built. Then, this matrix is used to build a regression model to explain

the variance of the response variable **Y**. When **Y** contains only one variable, the method is referred to as PLSR1, and when it contains more than one variable it is referred to as PLSR2; in both cases we refer to Y as a matrix of response variables. Some exhaustive reviews of this method have been published [36 and 37]. PLSR1 was used to relate TC with spectral data.

A common practice in soil spectroscopy is to use cross-validation techniques to select the optimal number of latent variables for PLSR. We performed leave-*k*-out cross-validation [38 and 39] to select the number of latent variables for regression (also referred to as *k*-fold CV). In this technique, the data set is randomly split into *k* equal-sized groups, where *k* is defined by the user. We used $k = 5$ as recommended by Li et al. [39] and latent variables ranged from 1 to 20. One group is left out, and the model is calibrated with the remaining $(k - 1)$ groups. Then, the prediction accuracy of the model is evaluated using the group that was left out by comparing the predicted and measured values. This process is repeated until each group is used to validate the model. When the value of *k* is equal to the number of samples, this technique is referred to as leave-one-out cross-validation (LOOCV). The mean of statistical indicators such as root-mean-squared error (*RMSE*) (**Eq. 2.1**) and coefficient of determination R^2 (**Eq. 2.2**) were used as a reference to evaluate the model performance.

$$RMSE = \sqrt{\sum_{i=1}^{i} \frac{(y_i - \hat{y}_i)^2}{n}} \tag{2.1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{i} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{i} (y_i - \overline{y}_i)^2} \tag{2.2}$$

where y_i is the measured value of sample i, $\hat{y}_i$ is the predicted value of sample i and $\overline{y}$ is the average of measured samples.

After the optimal number of latent variables were determined, the data set was split into two groups. The group for calibration contained 80% of the data set and the remaining 20% was used as validation group. Some strategies to split data have been applied by researchers [18, 33 and 40]. Nonetheless, we performed an iterative procedure to find the best data split (80/20%) in terms of model performance i.e., several possible data splits of 80/20%

were evaluated by PLSR. This was made using the scikit-learn library [41] of Python programming language (Python Software Foundation, **https://www.python.org/**), which achieves the same data split every time (pseudo-random) in order to reproduce the results. The best calibration/validation sets were selected by the higher R^2 in PLSR.

For the final model selection, all pre-processing configurations listed in **Table 2.2** were applied to the spectral data before splitting. Then, PLSR was performed for each one and the quality of model prediction was evaluated using the most-used quality estimators for regression such as R^2 and *RMSE*. All analyses were performed using a Python-based ecosystem for scientific computing [42] (**https://www.scipy.org/**). All calculations ware made under the Jupyter Notebook environment (formerly Ipython Notebook) to facilitate the reproduction of the results [43] (Supplementary Files: S1-reflectance-analysis and S2-calculations).

2.3 Results and Discussion

2.3.1 Soil Total Carbon

Total carbon measured in the 70 samples ranged from 0.77 to 10.7% (**Figure 2.2**A). The highest and lowest concentration of C in soil were found in the Andisol at the depths of 0–5 and 20–40 cm, respectively. These values were justified by environmental conditions and soil type. For Andisol, the ranges of TC of the three depths overlap mainly in lower values, i.e., values between 2 and 5% were found in all of the sampled depths. Values greater than 8% were found in the top soil layer (0–5 cm). For Ultisol, the lower values of TC were clearly observed in the depth of 20–40 cm. However, outliers were found for values greater than 4%, probably due to such samples being taken from undisturbed *Nothofagus* forest conditions [27]. The minimum and maximum frequency of the TC values were 1 and 14 (**Figure 2.2**B), respectively.

2.3.2 Spectral Pre-Processing

Before filtering the entire spectral data set, we tested the Savitzky–Golay filter on a noisy spectral curve in order to visually inspect its effect on noise (**Figure 2.3**). These filter settings (no transformation) were applied to the entire spectral data and their means were compared;

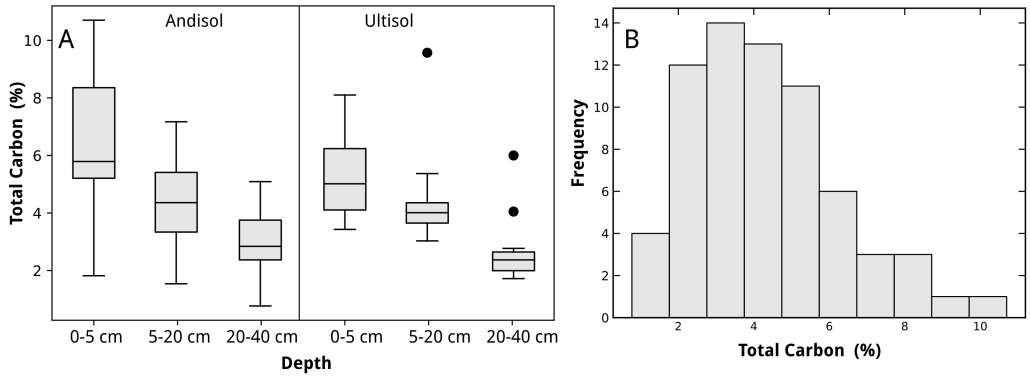

Figure 2.2 Carbon content by depth and soil order. (**A**) The full range variation is represented in the box and whisker diagrams. The top of the gray box is the third quartile and the bottom the first quartile. The horizontal line inside the box is the median of the data. Whiskers above and below the box show the minimum and maximum values and outliers are represented by black circles; (**B**) The distribution of the soil total carbon of the whole data set.

their means showed no significant differences. The extremes of the curves were removed due to noise, and the range of 400–924 nm was selected for the analysis.

After pre-processing, PCA was performed on the filtered data sets to explore their structure. In the most cases, two principal components explained high amount of variability of the spectra (>95%). However, using first derivative transformation, the explained variability reached 70% with five PCs and this value decreased when derivative order increased. Nonetheless, PCA showed potential for classifying samples from different soil orders from spectral data. A sophisticated method for enhancing the performance of classification of soils from spectral data was proposed by Xie et al. [44].

2.3.3 Effect of Soil TC on Reflectance

As we expected, soil layers with higher TC values, tend to have lower reflectance. However, in Andisol samples (**Figure 2.4**A), for 5–20 and 20–40 cm depths, the spectral curves intersected at wavelengths shorter than 400 nm. By contrast, the maximum spectral separation was observed near 650 nm. In the Ultisol samples (**Figure 2.4**B), the curves of the depths of 0–5 and 5–20 cm intersected between 700 and 800 nm. We attribute this curve intersection to the little difference in soil TC between the first two depths in Ultisol soil (**Figure 2.5**). We also calculated the correlation coefficient (r) between every spectral curve and soil TC.

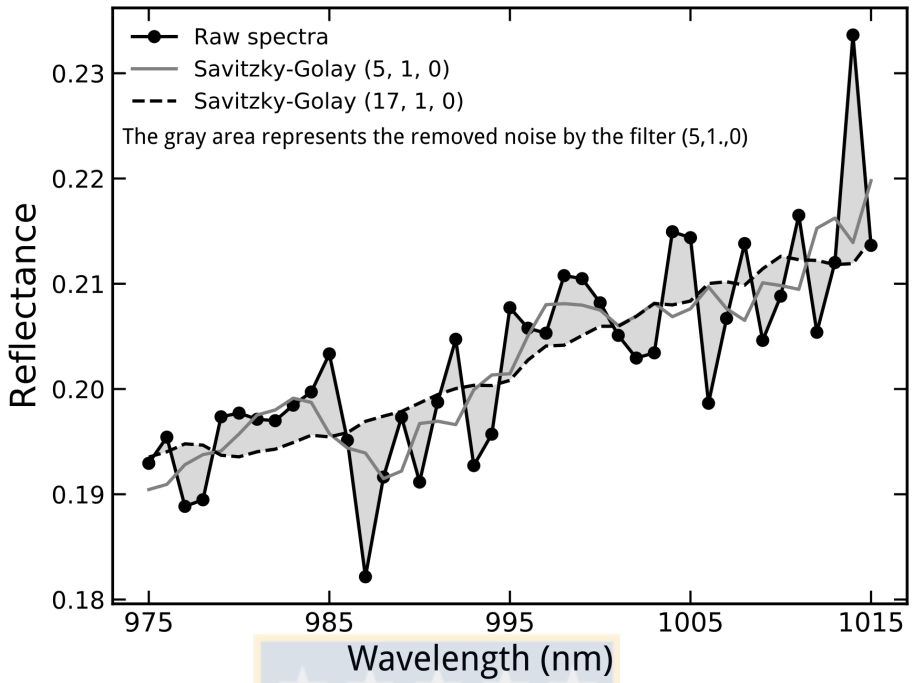

Figure 2.3 Effect of the Savitzky–Golay filter applied to a noisy spectral curve. No derivative transformation was used. Only forty points are displayed in order to appreciate its effect on noise. Inside the parentheses, window size, polynomial order and derivative order are given

In accordance with [19], we found the most negative correlation between TC and spectral wavelengths near the 500–600 nm range (**Figure 2.6**).

By comparing soil reflectance with TC, our results showed the influence of absorption features of soil carbon on reflectance. This is in agreement with previous works published by [19 and 34]. However, the correlation coefficient between spectral bands and TC in the region of 500–700 nm ($r \approx -0.5$), were poorer than those reported by others [19 and 34] ($r = -0.8$ or better). One possible explanation of this is the spectral distortion due to natural source of light used for scanning the samples (outdoor scanning) and low bulk density of the studied soils (≤ 1 g$\cdot$ cm^{-3}) [27] which promote more dispersion of light [45]. This distortion was probably not successfully corrected by the Savitzky–Golay filter. Despite this, near 450 nm (**Figure 2.4**), the absorption features of distinct iron oxides can be observed [34]. For the Ultisol samples, this absorption was lightly observed in the curve. The spectral separability was not clear when the curves were plotted individually, for some samples with higher soil TC the reflectance along the spectra was not necessarily lower than for samples with lower

TC. However, our results showed that averaging the reflectance by depth, the influence of soil TC was evident.

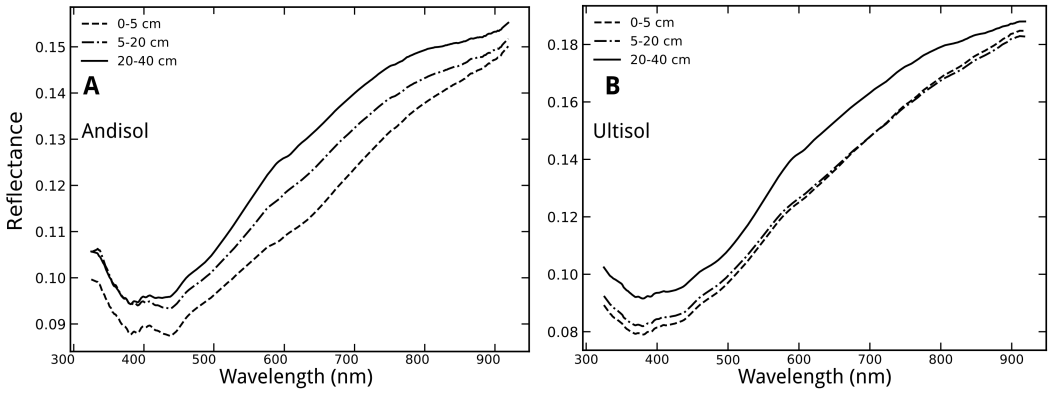

Figure 2.4 The average of soil reflectance of the three soil depths. (**A**) The spectral separability between depths can be observed from 450 nm; (**B**) for the depths of 0–5 and 5–20 cm the spectral separability was not exhibited, and the curves intersected between 650 and 750 nm.

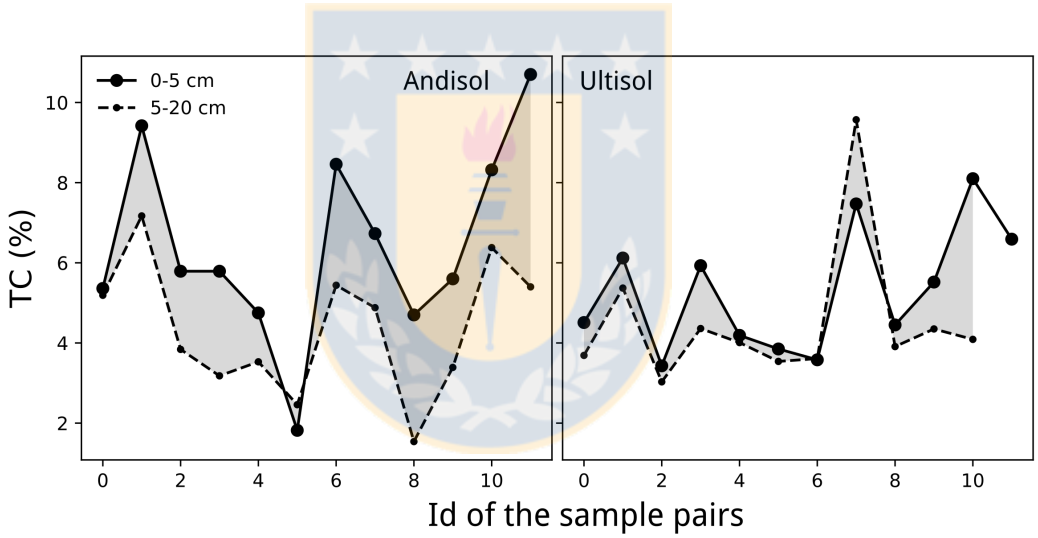

Figure 2.5 Total carbon (TC) in the Andisol and Ultisol soil orders. The gray area represents the difference in soil TC of the two depths.

2.3.4 Cross-Validation (CV) and Partial Least Squares Regression

To better understand the experimental results in this section, we will discuss them separately. Firstly, we will discuss the results obtained in leave-*k*-out CV, secondly we will analyze the results obtained by splitting the data set into calibration/validation subsets (80/20%, respectively), and finally the implications of the potential misinterpretation of the results will be addressed.

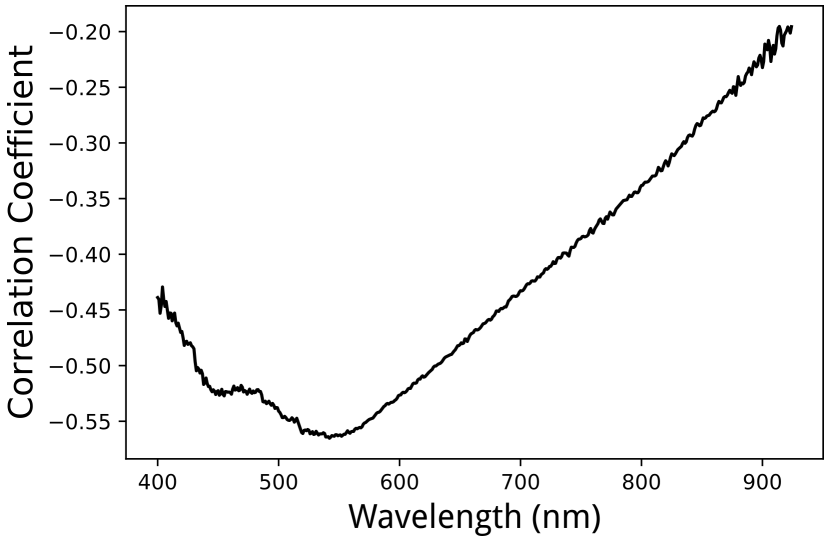

Figure 2.6 Correlation coefficient between individual spectral curves and TC. This graph shows the negative correlation between soil carbon content and soil reflectance. The higher correlation occurs near 550 nm.

2.3.4.1 Cross-Validation

With the parameters used for CV, the smoothed spectra yielded the lowest *RMSE* (1.8%) with 2 latent variables (LVs). The performance of the model was significantly poorer from 6 LVs (**Figure 2.7**) and this did not improve using LOOCV which is considered the most accurate [39]. If this is the case, Abdi [46] manifests that the model is overfitting the data and is not useful for predicting unobserved data. To our knowledge, this unusual behavior of the data in cross-validation has rarely been reported in the literature. According to several authors [39, 47 and 48] the optimal number of LVs is determined when only marginal improvements in model performance are observed, but with our data this marginal improvement was not clear, and the optimal number of LVs was selected based on lowest *RMSE* (**Figure 2.7**A). The R^2 in most cases was negative, indicating a bad model fit (**Figure 2.7**B). Negative R^2 resulted in model calibration, which has been also reported [19].

2.3.4.2 PLSR Calibration

Prior to final model calibration, an iterative process was performed to evaluate 600 possible data splits into calibration and validation data sets, at 80/20%, respectively. The best data split was selected by the highest R^2 resulting from PLSR prediction. The key function for this process was train_test_split of the scikit-learn library, test_size parameter was 0.2 indicating

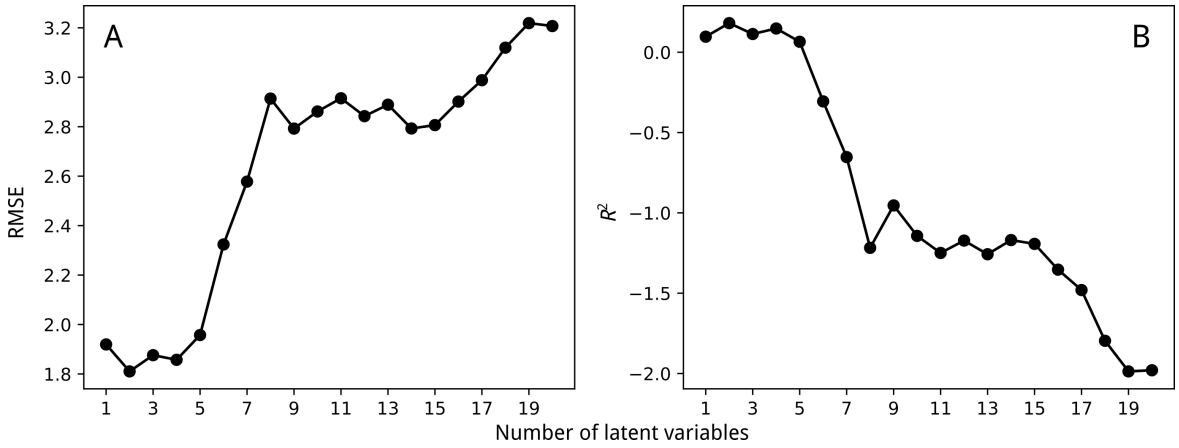

Figure 2.7 Performance of partial least squares regression (PLSR) in cross-validation with $k = 5$. (**A**) Root-mean-squared error (*RMSE*), and (**B**) R^2. These graphs could be used to indicate the consistency of the data for modeling. Both curves indicate that the data is not useful for modeling.

the proportion of the validation set. The optimal value of random_state parameter was 280 in most cases. Using the split data set of the smoothed spectra, the performance of the model was significantly better in terms of R^2 and *RMSE*. Varying the number of LVs to 1 and 5, the values of R^2 were 0.82 and 0.74 and *RMSE* were 0.64% and 0.75%, respectively. Using more than 5 LVs, the R^2 decreased drastically. With some pre-processing settings the model fitted better than others (**Table 2.2**). This has been also found by several authors [16, 19 and 26]. An ordinary least squares regression was performed to inspect the best model (**Figure 2.8**). Our best model (R^2 = 0.82 and *RMSE* = 0.64%) indicates a good prediction capacity for TC in accordance with the standard used by Sarkhot et al. [8], and the worst model was obtained using a second-order derivative transformation (R^2 = 0.23 and *RSME* = 1.51%). These results were congruent with those reported by [49] who used a similar pre-processing configuration, and the second derivative transformation produced the worst models for the most of the predicted variables. In contrast, Vasques, et al. [12] found subtle differences in PLSR performance using derivative transformations in pre-processing the data. On the other hand, Knadel, et al. [50] indicated that non-preprocessed spectral data generated the best model for soil organic carbon. Window size and polynomial order in the Savitzky–Golay algorithm had no important influence on model performance (**Table 2.2**).

Our results demonstrate that with the same data, it was possible to obtain different results in predicting soil TC. This discordance between the two procedures (with and without cross-validation), may lead to an ambiguous interpretation of the predictive capacity of the

built model. For example, avoiding cross-validation, the model resulted in a good capacity for predicting soil TC and could be used to predict TC from soil samples. Before final model calibration, the cross-validation technique was useful to inspect the potential of the spectral data to predict soil TC, and to select the optimal number of the latent variables for regression. PLSR is subject to overprediction when the number of samples for calibration is small (<58, in the soil spectroscopy context) [22]. However, some studies applied soil spectroscopy with fewer samples [12]. We used 70 soil samples from two study sites. After splitting the data and removing outliers, the calibration set resulted with 54 samples in most cases. The Ultisol samples were re-scanned several days later and the results were similar despite a different intensity of reflectance. More research is needed to better understand soil reflectance variability in these soil types.

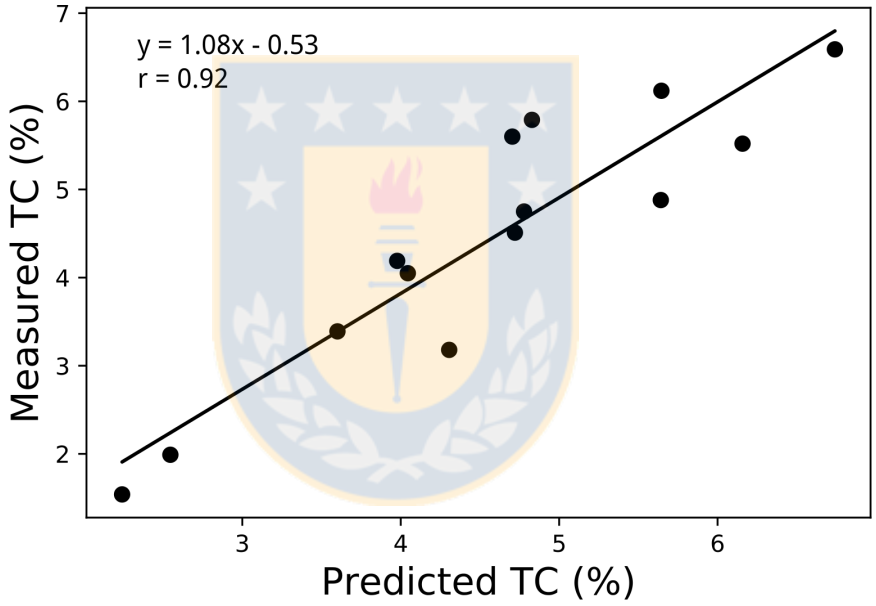

Figure 2.8 Linear regression applied to predicted and measured TC.

In summary we found that firstly, the average of reflectance in the VIS-NIR recorded from soil samples was useful as a descriptive type of information about the carbon content, showing an appreciable relationship with soil TC, especially in the andisol samples. However, it was not appropriate for building a robust model for predicting soil TC. Secondly, the Savitzky–Golay filter was effective in eliminating the most visible noise in the spectral data. Thirdly, using the random_state parameter in the iterative process, we rapidly found the best calibration/validation subsets for model calibration.

Table 2.2 PLSR performance for different Savitzky–Golay filter configurations applied to spectral data. The outlier ID column contains the index of the samples in the data set that were considered as outliers.

SG Filter	Number of LVs	R^2	*RMSE*	Outliers ID
(5, 1, 0)	2	0.82	0.61	21, 60
(5, 2, 0)	2	0.82	0.61	21, 60
(5, 1, 1)	2	0.58	1.22	20, 57
(5, 2, 2)	1	0.23	1.51	–
(11, 1, 0)	2	0.82	0.61	21, 60
(11, 2, 0)	2	0.82	0.61	21, 60
(11, 1, 1)	1	0.54	1.08	–
(11, 2, 2)	1	0.36	1.29	20
(17, 1, 0)	2	0.82	0.61	21, 60
(17, 2, 0)	2	0.82	0.61	21, 60
(17, 1, 1)	2	0.58	1.48	–
(17, 2, 2)	1	0.26	1.59	20
(17, 1, 0) + Log(1/R)	2	0.79	0.66	21, 60
(17, 1, 0) + centering	2	0.79	0.66	21, 60

We attempt with this study to offer additional support for an effective application of soil spectroscopy. In contrast to most published works, we report this negative case of the soil spectroscopy technique to show the potential model overfitting and misinterpretation of the results by soil scientists with little experience in data analysis. The results presented here need to be interpreted with caution because of the unusual behavior observed in reflectance for predicting soil carbon content.

2.4 Conclusions

We conducted the standard procedure to build a statistical model to predict soil TC from spectral data. The results of the experiment warn of possible model overfitting when the sources of

variability of the spectra (particle size and illumination) have not been effectively controlled and the amount and distribution of the soil samples are inadequate. However, we demonstrate that if cross-validation (CV) is avoided, it is possible to obtain a good PLSR model, which may, in turn, be inappropriately applied to unobserved data. To identify this issue, the cross-validation technique was useful for plotting the performance of the model versus the number of latent variables. Notwithstanding, our results confirm that the reflectance was influenced by soil carbon content, although it was only useful at the description level. We concluded that the potential of soil spectroscopy may be minimized when the spectral distortion exceeds the capacity of the filter to correct it. The cause of the spectral variability in these soil samples remains unclear. More sophisticated instruments and more rigorous scanning procedures may help to understand why in this case soil spectroscopy was not feasible.

2.5 **Supplementary Material**

The supplementary materials are available online at:

https://github.com/farliz/supp_material/blob/master/S1-reflectance-analysis.ipynb.

https://github.com/farliz/supp_material/blob/master/S2-calculations.ipynb

2.6 Acknowledgments

The present work benefited from the FONDECYT Project No. 11121279 granted by the National Commission for Scientific and Technological Research of the Chilean Government. L.R. received financial support from the National Secretary of Higher Education, Science, and Technology (SENESCYT) of Ecuador. Two anonymous reviewers are acknowledged for their suggestions that improved the manuscript.

2.7 Author contributions

Lizardo Reyna, Juan A. Barrera and Erick Zagal, conceived, designed and performed the experiments; Francis Dube facilitated access to the experimental sites and soil samples. All authors contributed with valuable discussions and scientific advices in order to improve the quality of the work.

The authors declare no conflict of interest.

2.8 References

1 Stevenson, F. and Cole, M. (1999). *Cycles of Soils*. 2 edition New York: Wiley.

2 Dube, F., Zagal, E., Stolpe, N. and Espinosa, M. (2009). The influence of land-use change on the organic carbon distribution and microbial respiration in a volcanic soil of the Chilean Patagonia. *Forest Ecology and Management, 257*(8), 1695–1704.

3 Zagal, E., Muñoz, C., Espinoza, S. and Campos, J. (2012). Soil profile distribution of total c content and natural abundance of 13c in two volcanic soils subjected to crop residue burning versus crop residue retention. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 62*(3), 263-272.

4 Powlson, D., Smith, P. and Nobili, M. D. (2013). *Soil organic matter*, pages 86–131. Blackwell Publishing Ltd.

5 Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M. and Wetterlind, J. (2010). Chapter Five - Visible and Near Infrared Spectroscopy in Soil Science. In Sparks, D. L., editor, *Advances in Agronomy*, pages 163–215. Academic Press.

6 Viscarra Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J. and Lobsey, C. (2011). Chapter Five - Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time. In Sparks, D. L., editor, *Advances in Agronomy*, number 113 in Advances in Agronomy, pages 243–291. Academic Press.

7 Reeves III, J. B. (2010). Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done?. *Geoderma, 158*(1-2), 3–14.

8 Sarkhot, D. V., Grunwald, S., Ge, Y. and Morgan, C. L. S. (2011). Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. *Geoderma, 164*(1-2), 22–32.

9 Fontán, J. M., Calvache, S., López-Bellido, R. J. and López-Bellido, L. (2010). Soil carbon measurement in clods and sieved samples in a Mediterranean Vertisol by Visible and Near-Infrared Reflectance Spectroscopy. *Geoderma, 156*(3-4), 93–98.

10 Reeves III, J. B., Follett, R. F., McCarty, G. W. and Kimble, J. M. (2006). Can Near or Mid-Infrared Diffuse Reflectance Spectroscopy Be Used to Determine Soil Carbon Pools?. *Communications in Soil Science and Plant Analysis, 37*(15-20), 2307–2325.

11 Knox, N. M., Grunwald, S., McDowell, M. L., Bruland, G. L. and Myers, D. B. et al. (2015). Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy. *Geoderma, 239*, 229–239.

12 Vasques, G. M., Grunwald, S. and Sickman, J. O. (2008). Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. *Geoderma, 146*(1-2), 14–25.

13 Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G. and Buttafuoco, G. (2017). Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. *Geoderma, 288*, 175 – 183.

14 Mouazen, A. M., Kuang, B., De Baerdemaeker, J. and Ramon, H. (2010). Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma, 158*(1), 23–31.

15 Fystro, G. (2002). The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis-NIR spectroscopy and comparative methods. *Plant and Soil, 246*(2), 139–149.

16 Brunet, D., Barthès, B. G., Chotte, J.-L. and Feller, C. (2007). Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity. *Geoderma, 139*(1-2), 106–117.

17 Gomez, C., Viscarra Rossel, R. A. and McBratney, A. B. (2008). Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. *Geoderma, 146*(3-4), 403–411.

18 Wenjun, J., Zhou, S., Jingyi, H. and Shuo, L. (2014). In Situ Measurement of Some Soil Properties in Paddy Soil Using Visible and Near-Infrared Spectroscopy. *PLOS ONE, 9*(8), e105708.

19 Zheng, G., RYU, D., JIAO, C. and HONG, C. (2016). Estimation of Organic Matter Content in Coastal Soil Using Reflectance Spectroscopy. *Pedosphere, 26*(1), 130–136.

20 Guillén, C. E., Dávila, M. J., Gilliot, J. M. and Vaoudour, E. (2013). Aporte de la espectroscopia a la estimación de carbono orgánico de los suelos de la planicie de versalles, francia. *Revista Geográfica Venezolana, 54*(1), 85–98.

21 Baumgardner, M. F., Silva, L. F., Biehl, L. L. and Stoner, E. R. (1986). Reflectance Properties of Soils. In Brady, N. C., editor, *Advances in Agronomy*, pages 1–44. Academic Press.

22 Reeves III, J. B., McCarty, G. W., Calderon, F. and Hively, W. D. (2012). Chapter 20 - Advances in Spectroscopic Methods for Quantifying Soil Carbon A2 - Liebig, Mark A.. In Franzluebbers, A. J. and Follett, R. F., editors, *Managing Agricultural Greenhouse Gases*, pages 345–366. San Diego: Academic Press.

23 Stolpe, N. B. (2006). Descripción de los Principales Suelos de la VII Región de Chile. Technical Report Departamento de Suelos y Recursos Naturales - Universidad de Concepción.

24 Rinnan, ., van den Berg, F. and Engelsen, S. B. (2009). Review of the most common preprocessing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry, 28*(10), 1201–1222.

25 Savitzky, A. and Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures.. *Analytical Chemistry, 36*(8), 1627–1639.

26 Nawar, S., Buddenbaum, H., Hill, J., Kozak, J. and Mouazen, A. M. (2016). Estimating the soil clay content and organic matter by means of different calibration methods of vis-NIR diffuse reflectance spectroscopy. *Soil and Tillage Research, 155*, 510–522.

27 Dube, F. and Stolpe, N. B. (2016). SOM and Biomass C Stocks in Degraded and Undisturbed Andean and Coastal Nothofagus Forests of Southwestern South America. *Forests, 7*(12), 320.

28 Casanova, M., Salazar, O., Seguel, O. and Luzio, W. (2013). *Main Features of Chilean Soils*, pages 25–97. Dordrecht: Springer Netherlands.

29 Wright, A. F. and Bailey, J. S. (2001). Organic carbon, total carbon, and total nitrogen determinations in soils of variable calcium carbonate contents using a Leco CN-2000 dry combustion analyzer. *Communications in Soil Science and Plant Analysis, 32*(19&20), 3243–3258.

30 Schafer, R. W. (2011). What Is a Savitzky-Golay Filter? [Lecture Notes]. *IEEE Signal Processing Magazine, 28*(4), 111–117.

31 Kinoshita, R., Roupsard, O., Chevallier, T., Albrecht, A. and Taugourdeau, S. et al. (2016). Large topsoil organic carbon variability is controlled by Andisol properties and effectively assessed by VNIR spectroscopy in a coffee agroforestry system of Costa Rica. *Geoderma, 262*, 254–265.

32 van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. and van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*, 142.

33 Adeline, K. R. M., Gomez, C., Gorretta, N. and Roger, J. M. (2017). Predictive ability of soil properties to spectral degradation from laboratory Vis-NIR spectroscopy data. *Geoderma, 288*, 143–153.

34 Henderson, T. L., Baumgardner, M. F., Franzmeier, D. P., Stott, D. E. and Coster, D. C. (1992). High dimensional reflectance analysis of soil organic matter. *Soil Science Society of America, Journal*.

35 Zhang, P. and 'an Shao, M. (2014). Spatial variability and stocks of soil organic carbon in the gobi desert of northwestern china.. *PloS one, 9*(4), e93584.

36 Geladi, P. and Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta, 185*, 1–17.

37 Haenlein, M. and Kaplan, A. M. (2004). A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics, 3*(4), 283–297.

38 Jonathan, P., Krzanowski, W. J. and McCarthy, W. V. (2000). On the use of cross-validation to assess performance in multivariate prediction. *Statistics and Computing, 10*(3), 209–229.

39 Li, B., Morris, J. and Martin, E. B. (2002). Model selection for partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 64*(1), 79–89.

40 Nocita, M., Stevens, A., Noon, C. and van Wesemael, B. (2013). Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. *Geoderma, 199*, 37–42.

41 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V. and Thirion, B. et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

42 Jones, E., Oliphant, T., Peterson, P. and others (2001). SciPy: Open source scientific tools for Python. .

43 Shen, H. (2014). Interactive notebooks: Sharing the code. *Nature News, 515*(7525), 151.

44 Xie, H., Zhao, J., Wang, Q., Sui, Y. and Wang, J. et al. (2015). Soil type recognition as improved by genetic algorithm-based variable selection using near infrared spectroscopy and partial least squares discriminant analysis. *Scientific Reports, 5*, 10930.

45 Demattê, J. A. M., Nanni, M. R., da Silva, A. P., Filho, J. F. d. M. and Santos, W. C. D. et al. (2010). Soil density evaluated by spectral reflectance as an evidence of compaction effects. *International Journal of Remote Sensing, 31*(2), 403–422.

46 Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(1), 97–106.

47 Brown, D. J., Bricklemyer, R. S. and Miller, P. R. (2005). Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. *Geoderma, 129*(3-4), 251–267.

48 Viscarra Rossel, R. A. (2008). ParLeS: Software for chemometric analysis of spectroscopic data. *Chemometrics and Intelligent Laboratory Systems, 90*(1), 72–83.

49 Askari, M. S., O'Rourke, S. M. and Holden, N. M. (2015). Evaluation of soil quality for agricultural production using visible-near-infrared spectroscopy. *Geoderma, 243-244*, 80–91.

50 Knadel, M., Gislum, R., Hermansen, C., Peng, Y. and Moldrup, P. et al. (2017). Comparing predictive ability of laser-induced breakdown spectroscopy to visible near-infrared spectroscopy for soil property determination. *Biosystems Engineering, 156*, 157–172.

CAPÍTULO 3

Second-growth *Nothofagus obliqua* forest evaluation by vegetation indices derived from multi-temporal Landsat imagery

Enviado a la revista *Environmental Monitoring and Assessment* **ISSN: 0167-6369**

Second-growth *Nothofagus obliqua* forest evaluation by vegetation indices derived from multi-temporal Landsat imagery

Lizardo Reyna[1] , Francis Dube[2] , Juan A. Barrera[1] and Erick Zagal[1]

1 Department of Soils and Natural Resources, Faculty of Agronomy, Universidad de Concepción, Vicente Méndez 595, Casilla 537, Chillán 3812120, Chile; lreyna@udec.cl, jbarrera@udec.cl

2 Department of Silviculture, Faculty of Forest Sciences, Universidad de Concepción, Victoria 631, Casilla 160-C, Concepción 4030000, Chile; fdube@udec.cl

Abstract. *A timely evaluation of forestland allows detecting a negative impact of human activity on the forest. We have successfully applied a non-spatial approach for evaluating the dynamics of the green biomass of the Nothofagus obliqua forests of the Biobio region of Chile. Twelve experimental plots were selected for assessing the potential of a multi-temporal source of Normalized Difference Vegetation Index (NDVI), Nir reflectance, NIR vegetation (NIRv) in order to detect differences between different forest conditions (degraded, partly-degraded and undisturbed). Leaf area index (LAI) was used as a reference of foliage density to differentiate the forest conditions. The undisturbed condition located on the south-facing slope had an average of LAI of 3.5, while the lowest LAI (0.8) resulted from the degraded condition in the north-facing slope. Despite the modeled insolation resulted lower for south-facing slope, major values of LAI were found for the undisturbed and degraded condition on that side of the slope. Our analysis revealed that Nir band of the Landsat OLI sensor clearly describes the phenological development of the green biomass of the forest. This is because it is less sensitive to the thin clouds than the red band. Multiple comparisons of the means method applied to NDVI, Nir, and NIRv ($p < 0.05$) showed significant differences between forest conditions. These results suggest that Nir reflectance and NDVI could be useful to identify subtle changes in the green biomass of the Nothofagus forest. These subtle changes may be related to a degradation or the recovery process of the forest.*

Keywords: remote sensing; NDVI; Landsat OLI; forest degradation; landscape management, multiple comparison, reflectance

3.1 Introduction

The efforts in using remote sensing for assessing forests have been mainly concentrated in the study of gross primary production (GPP), carbon stock, deforestation and forest degradation (Jagger and Kittner, 2017, Reddy et al., 2015 and Eckert et al., 2011). Most of the applications are performed on pixel-based analysis algorithms to map these variables at large scales. As data source, there are several freely available satellite images such as Landsat (Roy et al., 2014), Sentinel (Malenovský et al., 2012), CBERS (Lino et al., 2000), and MODIS (Fan and Liu, 2016). The growing demand for natural resources including wood, fiber, and services, is increasing the pressure on the remnant native forests. This forces us to find more localized methods to study these forests patches. A timely and quantitative diagnostic of the condition of the native forest patches helps to apply a strategy for an adequate management to conserve them. Thus, a rapid and accurate method for assessing forest changes is needed.

Chilean temperate forest is one of the world's 35 biodiversity hotspots and occupy the largest area of this forest type in South America (Miranda et al., 2015). In the Araucanía region the exotic tree plantations expansion is the main cause of native forest loss (Nahuelhual et al., 2012 and Aguayo et al., 2009). The most evident land cover change was observed in the Coastal range where the 47% of the landscape is now dominated by exotic tree plantations. A similar pattern but less intense was reported for Andean range which contains 70% of the native forest with regard to the 1973-1987 period (Miranda et al., 2015). Forests provide an effective pathway for carbon sequestration. Hence, its dynamics must be monitored to prevent abrupt losses.

The Normalized Difference Vegetation Index (NDVI) is the most commonly used index in remote sensing for assessing forests. The NDVI results from the combination in the form of the ratio of red and near-infrared (NIR) reflectance or radiance of the vegetation (Baret and Guyot, 1991). The strong absorption by leaf pigments of the radiation in the visible region (400 - 700 nm) and weak absorption in NIR (700 - 1300 nm) enhances the contrast between these two spectral regions (Huete, 2012). This is the theoretical foundation of the NDVI and other vegetation indices which can be consulted at (http://www.indexdatabase.de/). NDVI is used for calibrating empirical models to predict some forest biophysical properties including leaf area index (LAI), GPP, fraction of photo-synthetically active radiation (fPAR), and carbon stock (Yin et al., 2017, Schubert et al., 2012 and Bayat et al., 2012). Spectral vegetation

indices offer unprecedented capabilities for mapping forest properties at large scales. However, monitoring slight changes in the forest biomass by remotely sensed spectral reflectance is still a challenging task.

The Landsat 8 satellite was launched on February 11, 2013. In conjunction with its predecessors provides the largest satellite images database in the world. This newly satellite incorporate improved technology for image acquisition including higher signal-to-noise and radiometric resolution in the two sensors; the Operational Land Imager (OLI) and the Thermal Infrared (TIRS) (USGS, 2015). NDVI OLI-derived has higher spatial variability than NDVI from Landsat 7 ETM+ (Ke et al., 2015). This indicates that OLI bands record more spectral details of the earth surface that may used for monitoring changes in forest biomass due to degradation or recovery process. NIR reflectance of vegetation has been reported as better proxy for fPAR than NDVI (Sellers et al., 1992 and Sellers, 1987,1985). Based on Sellers works, in a recent publication Badgley et al. (2017) propose a new index; the near-infrared reflectance of vegetation (NIRv), which eliminates the mixed-pixel problem.

Considering the improved features of the Landsat OLI sensor, and the freely available images, our goal was to evaluate the potential of the NDVI, NIR, and NIRv derived from the OLI sensor to detect differences in the aerial biomass of the Chilean *Nothofagus Obliqua* forest under different degradation states. For this purpose, we used a non-spatial analysis for the period 2013 to 2016.

3.2 Methods

3.2.1 Site description

The study site is located in the Andean Foothills at 36°48"S, Long. 71°38"W in the Biobio Region of Chile (**Figure 3.1**). The landscape is dominated by second-growth *Nothofagus obliqua* forest 'Roble', which belongs to the remnant temperate forest in South America (Miranda et al., 2015). This site has a mean annual precipitation and temperature of 3000 mm and 13.5°C, respectively (Dube and Stolpe, 2016). The soil has high amount of organic carbon in the top layer (Andisol) and no carbonates (Stolpe, 2006).

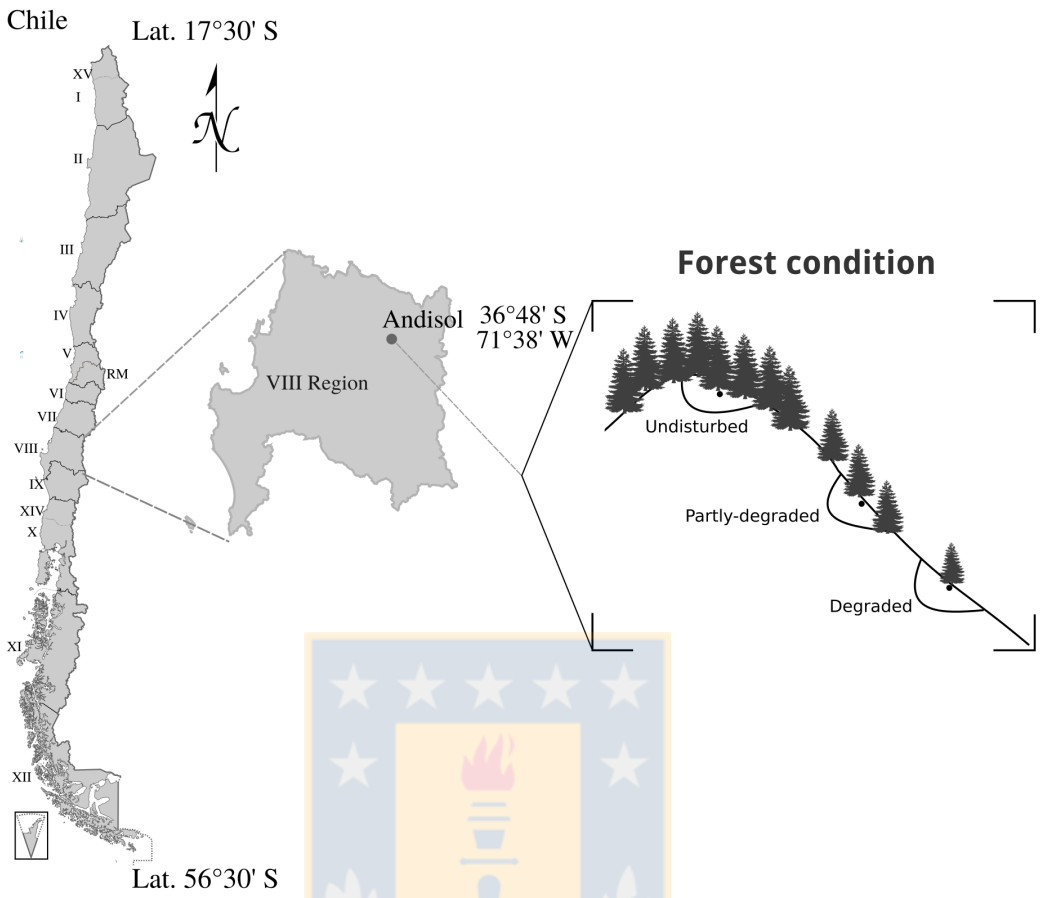

Figure 3.1 The study site is located in the Region VIII (Biobio) in the Andean range of Chile. This figure was adapted from (Reyna et al., 2017).

Twelve points were georeferenced for identifying different forest conditions; degraded (D), partly-degraded (P) and undisturbed (U) (**Figure 3.2**). The geographic location of the points was corrected using a satellite image. These sites correspond to the plots used by (Dube and Stolpe, 2016) whose assessed biomass stock in the selected forest conditions. The plots were labeled according to the conditions and location (**Table 3.1**). For instance; AND-1, the A refers to the Andes location, N/S indicate the north or south exposition of the slope, D/P/U, indicate the forest condition, and the number is the replication.

Figure 3.2 Different forest conditions. **ANU-1**: Undisturbed, **ASP2**: Partly-degraded and **AND-2**: Degraded.

3.2.2 Image acquisition and NDVI

We used landsat-util command line application[1] to search and download the Landsat OLI bands (red and nir, 4 and 5 respectively) for the period 2013 to 2016 (see appendix A). The green band (3) was also downloaded only for false color composite (**Table 3.2**). The site of study is within an overlap zone between two Landsat scenes with approximately one second of difference in acquisition time; path:233 and rows:085 and 086 according to worldwide reference system (WRS-2). Scenes with 20% of cloud were selected to download. For autumn and winter seasons it was difficult to find optimal images because of the cloudy sky. In the two overlapped images (rows 085 and 086) the reflectance of the selected pixels in the bands 4 and 5 (**Table 3.1**) had no significant differences. Row 085 was selected for acquiring the images. Bands 4 and 5 were used to calculate NDVI (Rouse et al., 1974 and Tucker, 1979) which is (**Eq 3.1**) commonly used for evaluating vegetation dynamics (Huete, 2012). In addition, we calculate the NIR vegetation (NIRv) proposed by (Badgley et al., 2017) to minimize the mixed-pixel problem (Chhikara, 1984). The satellite pass over thesite at about 14:30 hrs (Greenwich Mean Time). The images were processed using Grass Gis 7 software (Neteler et al., 2012).

$$NDVI = \frac{Nir - red}{Nir + red} \tag{3.1}$$

[1] **https://github.com/developmentseed/landsat-util**

Table 3.1 Geographic coordinates of selected plots for evaluation. The letter A indicates the Andes location. The N/S indicate the aspect of the slope (north and south, respectively). The third letter indicates the forest condition (Degraded, Partly-degraded and Undisturbed) and the number the replication. All conditions were established in the field by duplicate (1 and 2).

ID	Longitude	Latitude	Elevation (m)
AND-1	71°38'18.811 W	36°48'1.401 S	663
AND-2	71°38'12.138 W	36°48'0.586 S	659
ANP-1	71°38'13.51 W	36°48'0.632 S	662
ANP-2	71°38'16.78 W	36°48'0.864 S	651
ANU-1	71°38'20.623 W	36°48'4.904 S	717
ANU-2	71°38'17.178 W	36°48'3.543 S	700
ASD-1	71°38'30.108 W	36°48'12.464 S	632
ASD-2	71°38'25.741 W	36°48'13.912 S	641
ASP-1	71°38'27.562 W	36°48'13.983 S	638
ASP-2	71°38'23.914 W	36°48'15.061 S	649
ASU-1	71°38'9.151 W	36°47'46.619 S	747
ASU-2	71°38'12.423 W	36°47'49.724 S	714

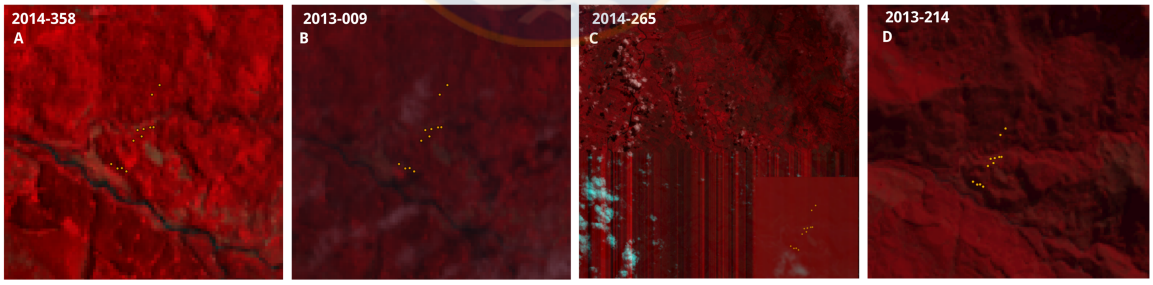

Figure 3.3 False color composition of three Landsat OLI bands (5,4 and 3). The different image conditions are shown in the pictures: **A**) optimal conditions for analysis, **B** cloudy condition, **C** radiometric anomaly and **D** the shadow effect. The points on the images are the locations of the plots.

3.2.3 Slope aspect and insolation

The slope aspect was calculated based on an SRTM 1-arc-second digital elevation model

Table 3.2 Landsat 8 bands designation and spatial resolution.

Bands	Wavelength (μm)	Pixel size (m)
1 - Ultra Blue (coastal/aerosol)	0.43 - 0.45	30
2 - Blue	0.45 - 0.51	30
3 - Green	0.53 - 0.59	30
4 - Red	0.64 - 0.67	30
5 - Near Infrared (NIR)	0.85 - 0.88	30
6 - Shortwave Infrared (SWIR) 1	1.57 - 1.65	30
7 - Shortwave Infrared (SWIR) 2	2.11 - 2.29	30
8 - Panchromatic	0.50 - 0.68	15
9 - Cirrus	1.36 - 1.38	30
10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

(u03_s037_w072) using r.slope.aspect module of the Grass Gis software. The aspect values were used to confirm the exposition of the slopes. Values of the aspect greater than 180 indicates south exposition. To investigate the influence of the insolation on the forest, r.sun module was used to model the number of the hours that the slope receive direct radiation (insolation) for the DOY 214 and 358 which according to (Escobar et al., 2014). Grass Gis r.sun model uses at least four parameters to model solar radiation taking into account topographic angles and shading (Chance et al., 2016) Digital elevation model (DEM), the day of the year, time-step, and Linke Turbidity Index (T_L) all are the needed parameters to run the model. Where $T_L = 1$ indicates an absolutely clear atmosphere. Among the outputs of the model, an insolation time raster map is given in hours.

3.2.4 Nir, NIRv and NDVI

Digital numbers of the images were converted to at-sensor reflectance using the procedure presented in the official manual (USGS, 2015) and implemented in the i.landsat.toar module (GRASS Development Team, 2017). Geographic coordinates of the experimental plots

were used to create a vector point map for extracting reflectance values and then were exported to a comma-separated values (csv) text file for a non-spatial analysis. The text file containing the reflectance values was loaded into a Python (**www.python.org**) session for the calculations and statistical analysis. A Jupyter notebook was used for integrating the Python-based libraries for the analysis and visualization of the data (Jones et al., 2001) and to easily reproduce the results (Shen, 2014). Atmospheric correction was no performed because of atmospheric effects on the NIR band of the OLI sensor are minimal (Ding et al., 2014 and Abdullah et al., 2011). We used a box plot (a.k.a. box and whisker diagram) to statistically describe the NDVI values for the period of the study. This graph was also used to identify rare values of NDVI that can be misinterpreted and for selecting the optimal images to use in the analysis. The mean of the NDVI, NIR, red, and NIRv was plotted to visually appreciate the sensitivity to vegetation changes and to detect extreme values in reflectance. NIRv was calculated as a product of NDVI and NIR in order to extract the portion of the reflectance attributable to the vegetation (Badgley et al., 2017).

3.2.5 Multiple comparison

The variation across time of the NDVI was firstly evaluated by a box plot graph. It displays the statistical description of the NDVI including all experimental plots. Useless data were identified and excluded from the data set due to presence of the clouds, shadows and radiometric anomalies on the image. The second analysis was performed comparing the NDVI, NIR, and NIRv grouped by experimental plots. This allowed identifying differences between forest condition. A multiple comparison was applied to spectral indices using the Tukey's HSD (honest significant difference) method to find significant differences ($\alpha = 0.05$) between forest conditions. Statsmodels[2] library of Python programming language was used for this purpose.

[2] http://www.statsmodels.org

3.3 Results and Discussion

3.3.1 LAI and NDVI

The average of LAI values of each experimental plot were used as reference of the forest green biomass. LAI has been used for several researches for this purpose (Xie et al., 2017, Wu et al., 2015 and Zarate-Valdez et al., 2012). As we note in **Figure 3.4**A, the LAI values decrease according to the level of degradation of the forest increase. For the north-facing slope, the differences among forest condition were significantly different in terms of LAI. However, for the south-facing slope the partly-degraded and degraded conditions have no significant differences of LAI values ($p < 0.05$) (**Figure 3.4**B). The aspect (compass direction) that a slope faces may influence the physical and biotic features of the environment (Helman et al., 2017 and Wang et al., 2015). In the study site, the north-facing slope has more insolation than south-facing slope due to the slope is tilted toward the sun. We found differences in the modeled insolation (direct radiation) between north and south-facing slopes, reaching a maximum of six and two hours for the 214 and 358 doy, respectively by (Niinemets et al., 1998), this difference in sun radiation may influence the leaf properties of deciduous trees. This could explain why the plots located in the south-facing slope have higher values of LAI in degraded and undisturbed conditions. However, more research is required to demonstrate the relationship between amount of incident light and LAI in *Nothofagus* forest.

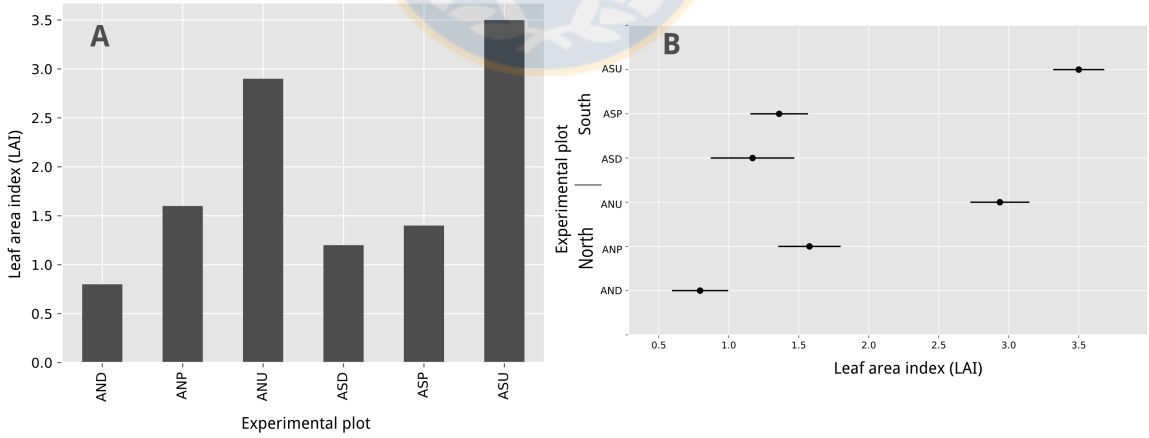

Figure 3.4 **A**) The average of Leaf Area Index (LAI) of each forest condition measured in the experimental plots. **B**) The figure shows the multiple comparison of the means of LAI among the experimental plots. The lack of intersection indicates that both means are different ($p < 0.05$). North and south indicate the direction that the slope faces.

The mean of the NDVI showed the variation of the green biomass density of the *Nothofagus* forest caused by seasonal changes. The maximum values were observed during the summer for the three years (**Figure 3.6** 350-80 DOY). The resulted lower values of NDVI is due to the fact that trees lose their leaves seasonally especially in autumn and winter. As commented previously when the soil surface (including the organic layer) is exposed to the sensor path, the incident radiation is absorbed causing low reflectance values in red and Nir bands. On the other hand, radiometric anomalies in some OLI bands (**Figure 3.3**C) produced extreme values of NDVI which could be interpreted as normal values. For instance, in 2014 for the days 137, 169 and 265 the lowest and highest NDVI values were obtained due to radiometric anomalies. Therefore, such values should be eliminated from the data set or used with caution (see **Figure 3.6**). The NDVI values obtained for the 009 and 060 DOY were affected by clouds (**Figure 3.3**B). The lower values of NDVI in the south-facing slope specially in undisturbed plots, can be explained by the shadow effect (**Figure 3.8**A). As a consequence the sensor capture a weak reflectance. A topographic correction of the images can help to compensate the signal (Chance et al., 2016).

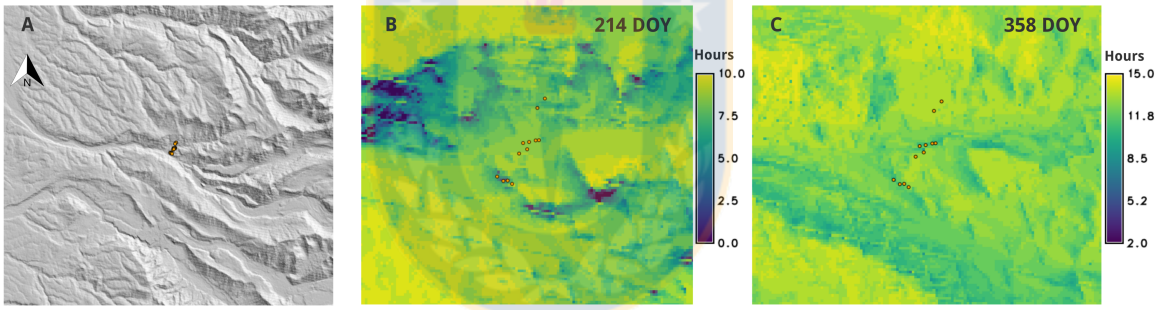

Figure 3.5 **A**) A shaded relief map based on a digital elevation model showing the topography of the study site. **B** and **C**) Solar irradiance was modeled considering the shadow effect caused by local topography. Two contrasting days were selected for modeling 214 and 358 doy, winter and summer respectively. These maps show the time in hours of the irradiation by the sun in a day. The points on the images are the locations of the experimental plots.

3.3.2 Red, Nir and NIRv

The green biomass of the 'Roble' forest is dense at spring and summer seasons. The spectral intensity of the canopy was affected by the soil absorption features and moisture because of spaces among leaves and trees that left exposed soil surface, specially in winter and autumn when the trees lose their leaves. In this forest, the soil contain high amount of carbon (Stolpe,

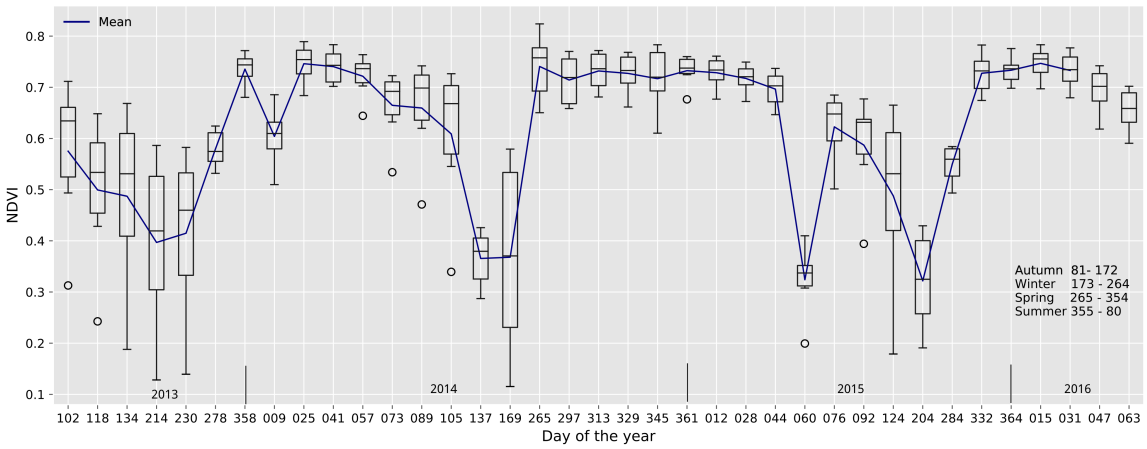

Figure 3.6 NDVI for the period 2013-2016. All boxes describes the NDVI statistics of the 12 sites. First and third quartiles are represented by the bottom and top of the box, respectively. The horizontal line inside the box is the median. Whiskers above and below the box indicate the minimum and maximum values and circles are extreme values. The days of the year corresponding to the seasons are given in the graphic.

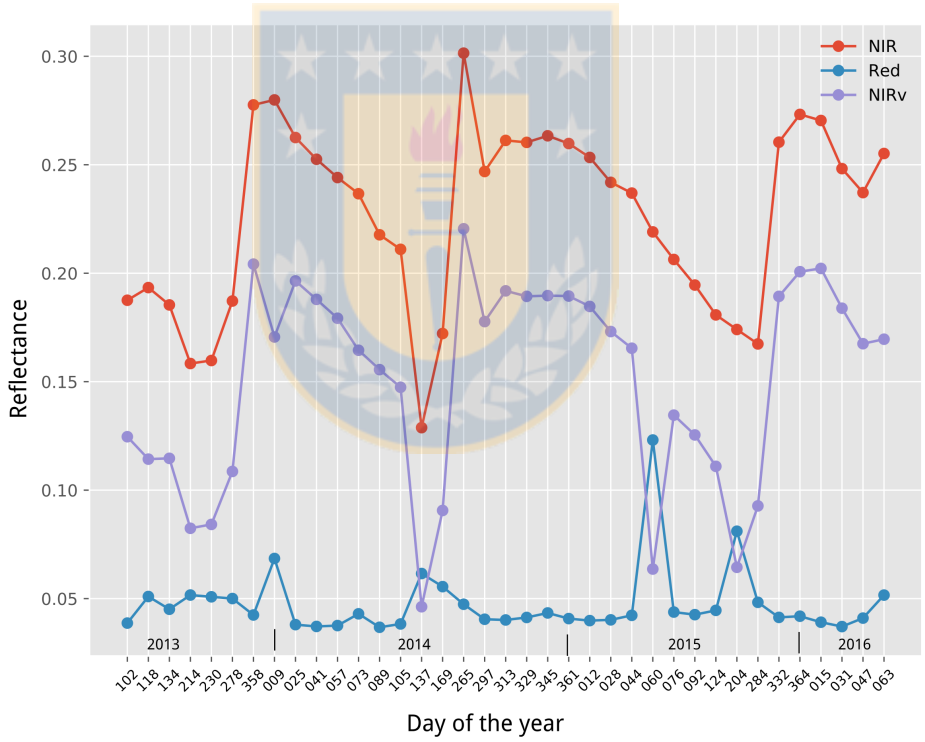

Figure 3.7 Average of the radiance values for the Nir and red bands of the Landsat OLI sensor. The shape of the curves depict the seasonal changes in green biomass of the *Nothofagus obliqua* forest. Every point is the average of the 12 plots.

2006) and usually there is an organic layer over soil surface that significantly reduce the reflected energy. The green leaves have high capacity to absorb radiation in red channel due to

photosynthetic pigments such as chlorophylls and carotenoids (Huete, 2012). For this reason, the spectral intensity of the forest biomass in the red band was lower than Nir band. Due to this fact red channel does not exhibit the phenological changes of the forest. The peaks observed in red band for 2014-009 and 2015-060 dates, were produced by cloud effect (thin cloud) over the plots, see **Figure 3.7**.

The Nir band clearly showed the green biomass dynamics of the forest across time (**Figure 3.7**), due to the linear relationship with fPAR (Badgley et al., 2017). Higher values of Nir reflectance was obtained at summer (from 355 to 80 DOY) when canopy foliage productivity is maximum. Nir band was a minimally effected by the thin cloud compared with the reflectance in the red band e.g. day 9 in **Figure 3.7**. For the other seasons the spectral intensity significantly decreased due to the mixed reflectance provided by soil surface because of the trees lose their leaves each year. The 'M' shape of the reflectance **Figure 3.7** correspond to the phenological development of the *Nothofagus* forest, slightly higher values are observed at summer in 2013-2014 period indicating that canopy foliage was higher at that dates. NIR reflectance better represents fPAR than NDVI (Badgley et al., 2017). Thus, NIR band from Landsat OLI sensor shows the potential to describe the phenological changes of the forest green biomass. This could be useful to quickly identify changes in the forest biomass caused for instance by landslides and deforestation.

NIRv was proposed in order to eliminate the mixed-pixel problem. The strong relationship with gross primary production (Badgley et al., 2017) suggest that it can be used for green biomass analysis. As a product of the Nir and NDVI, NIRv is sensible to the peaks in red reflectance. For example in 009 and 060 doy NIRv decrease because of the influence of red reflectance, these NIRv values indicate low green biomass density (**Figure 3.7**). On the other hand, Nir reflectance showed to be minimally affected by cloud effect (thin clouds) and reflects more accurately the phenological changes in the forest foliage. This implies that Nir reflectance from Landsat OLI sensor better describes the green biomass dynamics compared with red and NIRv. Therefore, it can be useful for describing changes in the foliage development of the *Nothofagus* forest. (Huang et al., 2016) successfully detected forest disturbances from remote sensing technique, but manifest that this condition is very complex due to the presence of both deforestation and forestation features. In addition, a seasonal and spatial variability of the reflectance in the analysis can be expected (Trancoso et al., 2015). This study

demonstrates that using a time series of the single NIR band of the OLI sensor, the forest dynamics can be assessed.

In **Figure 3.8** (left box plot) we appreciate that the differences between forest condition can not be clearly distinguished from spectral data. The values of NDVI, Nir and NIRv overlap among the boxes. South-faced plots resulted in more spread values in all variables and high values were obtained even in the degraded condition. This could be possible due to the influence of background reflectance from the other vegetation type such as grass (see **Figure 3.2**) and NDVI saturation (Wang et al., 2005). This little difference makes difficult to apply some classification algorithm such as maximum likelihood because the groups of pixels are not sufficiently separated (Wernick and Morris, 1988). Nonetheless, the multiple comparison of the means of the spectral variables (**Figure 3.8** right graphs) showed significant differences ($p < 0.05$) between forest conditions in all variables. The NDVI and Nir had significant differences between forest conditions in south-facing slope. For the north-facing slope the difference was only between degraded and undisturbed condition. The horizontal displacement of the lines in the multiple comparison graph can be used as an indicator of the degradation or recovery process in the forest.

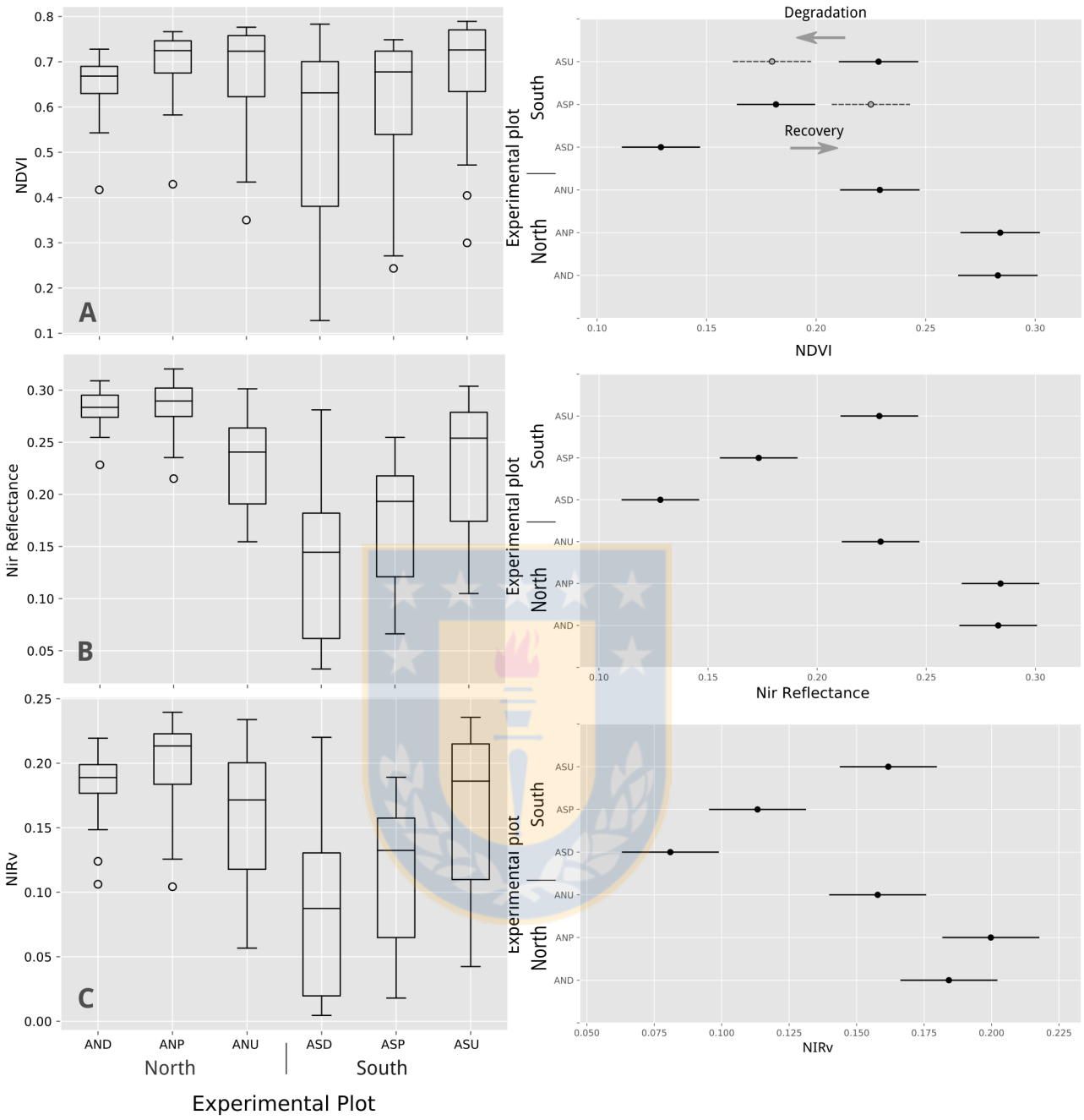

Figure 3.8 Descriptive statistics (left box plot) of some spectral variables grouped by different forest conditions (see Table 1 for details). To the right, the corresponding multiple comparisons of the means is shown. The lack of intersection indicates that both means are different ($p < 0.05$): **A**) NDVI, **B**) NIR reflectance and **C**) NIR vegetation. (NIRv). The segmented lines in the top right graph represent the projected degradation and recovery process of the forest.

Vegetation indices are commonly used in remote sensing for evaluating changes in the forests in the biomass of the forest including its spatial distribution. However, a challenging task comes out when the variation in the forest biomass does not imply land cover change (Sasaki and Putz, 2009). Forest degradation only reduces the ability of the forest to provide ecosystems services but the forest remains. The undisturbed and partly-degraded forest conditions in the study site, had no significant differences in aboveground biomass (Dube and Stolpe, 2016). This means that pixels of these plots can be assigned to the same land cover class in a image classification process. To solve this problem, a non-spatial approach for the reflectance analysis seems to be useful for monitoring these slight changes between forest status. The approach has also revealed the significance of solar radiation (north/south-facing slopes) in detecting changes due to forest degradation which is not depicted in the box plot. The advantage of the method presented here is the simplicity, similar results were found using NDVI and NIR band. These results are consistent with previously reported by (Ding et al., 2014). Therefore, the NIR band of Landsat 8 OLI sensor can be use as a valuable data source for monitoring the evolution of the forests.

3.4 Conclusion

We evaluated the potential of the NDVI, NIR reflectance, and NIR vegetation to find signals of degradation of the second-growth 'Roble' *Nothofagus obliqua* forest that belong to the temperate forest of Chile. While Landsat images are used mainly as a source for land cover classification, here we have adopted a non-spatial analysis to take advantage of Landsat 8 OLI bands to detect a degradation process in the forest. Our results suggest that using multiple comparison of the means of a multi-temporal source of NDVI or NIR reflectance, it is possible to observe the degradation or regeneration process in the *Nothofagus obliqua* forest. This approach can be useful for evaluating patches of forests subjected to either degradation or recovery processes with an easy method. For simplicity, this methodology can be successfully applied using the single NIR band of the Landsat OLI sensor. This approach combined with segmentation image processing promises to be a useful technique for monitoring of a forest patch.

3.5 References

Abdullah, H. M., Akiyama, T., Shibayama, M. and Awaya, Y. (2011). Estimation and validation of biomass of a mountainous agroecosystem by means of sampling, spectral data and QuickBird satellite image. *International Journal of Sustainable Development & World Ecology, 18*(5), 384–392.

Aguayo, M., Pauchard, A., Azócar, G. and Parra, O. (2009). Cambio del uso del suelo en el centro sur de chile a fines del siglo xx: Entendiendo la dinámica espacial y temporal del paisaje. *Revista chilena de historia natural, 82*, 361 - 374.

Badgley, G., Field, C. B. and Berry, J. A. (2017). Canopy near-infrared reflectance and terrestrial photosynthesis. *Science Advances, 3*(3).

Baret, F. and Guyot, G. (1991). Potentials and limits of vegetation indices for lai and apar assessment. *Remote Sensing of Environment, 35*, 161-173.

Bayat, A. T., van Gils, H. and Weir, M. (2012). Carbon stock of european beech forest; a case at m. pizzalto, italy. *APCBEE Procedia, 1*, 159 - 168.

Chance, C. M., Hermosilla, T., Coops, N. C., Wulder, M. A. and White, J. C. (2016). Effect of topographic correction on forest change detection using spectral trend analysis of landsat pixel-based composites. *International Journal of Applied Earth Observation and Geoinformation, 44*, 186–194.

Chhikara, R. S. (1984). Effect of mixed (boundary) pixels on crop proportion estimation. *Remote Sensing of Environment, 14*(1), 207–218.

Ding, Y., Zhao, K., Zheng, X. and Jiang, T. (2014). Temporal dynamics of spatial heterogeneity over cropland quantified by time-series ndvi, near infrared and red reflectance of landsat 8 {OLI} imagery. *International Journal of Applied Earth Observation and Geoinformation, 30*, 139 - 145.

Escobar, R. A., Cortés, C., Pino, A., Pereira, E. B. and Martins, F. R. et al. (2014). Solar energy resource assessment in Chile: Satellite estimation and ground station measurements. *Renewable Energy, 71*(Supplement C), 324–332.

Eckert, S., Ratsimba, H. R., Rakotondrasoa, L. O., Rajoelison, L. G. and Ehrensperger, A. (2011). Deforestation and forest degradation monitoring and assessment of biomass and carbon stock of lowland rainforest in the analanjirofo region, madagascar. *Forest Ecology and Management, 262*(11), 1996 - 2007.

Fan, X. and Liu, Y. (2016). A global study of ndvi difference among moderate-resolution satellite sensors. *ISPRS Journal of Photogrammetry and Remote Sensing, 121*, 177 - 191.

Helman, D., Osem, Y., Yakir, D. and Lensky, I. M. (2017). Relationships between climate, topography, water use and productivity in two key mediterranean forest types with different water-use strategies. *Agricultural and Forest Meteorology, 232*, 319 - 330.

Huang, C., Zhou, Z., Wang, D. and Dian, Y. (2016). Monitoring forest dynamics with multi-scale and time series imagery. *Environmental Monitoring and Assessment, 188*(5), 273.

Huete, A. R. (2012). Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass, 6*(9), 513–532.

Jagger, P. and Kittner, N. (2017). Deforestation and biomass fuel dynamics in uganda. *Biomass and Bioenergy, 105*, 1 - 9.

Ke, Y., Im, J., Lee, J., Gong, H. and Ryu, Y. (2015). Characteristics of landsat 8 oli-derived {NDVI} by comparison with multiple satellite sensors and in-situ observations. *Remote Sensing of Environment, 164*, 298 - 313.

Lino, C. D. O., Lima, M. G. R. and Hubscher, G. L. (2000). Cbers — an international space cooperation program. *Acta Astronautica, 47*(2), 559 - 564.

Miranda, A., Altamirano, A., Cayuela, L., Pincheira, F. and Lara, A. (2015). Different times, same story: Native forest loss and landscape homogenization in three physiographical areas of south-central of Chile. *Applied Geography, 60*(0), 20 – 28.

Malenovský, Z., Rott, H., Cihlar, J., Schaepman, M. E. and García-Santos, G. et al. (2012). Sentinels for science: Potential of sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of Environment, 120*, 91 - 101.

Niinemets, U., Bilger, W., Kull, O. and Tenhunen, J. D. (1998). Acclimation to high irradiance in temperate deciduous trees in the field: changes in xanthophyll cycle pool size and in photosynthetic capacity along a canopy light gradient. *Plant, Cell & Environment, 21*(12), 1205–1218.

Neteler, M., Bowman, M., Landa, M. and Metz, M. (2012). GRASS GIS: a multi-purpose Open Source GIS. *Environmental Modelling & Software, 31*, 124–130.

Nahuelhual, L., Carmona, A., Lara, A., Echeverría, C. and González, M. E. (2012). Land-cover change to forest plantations: Proximate causes and implications for the landscape in south-central chile. *Landscape and Urban Planning, 107*(1), 12 - 20.

Reyna, L., Dube, F., Barrera, J. A. and Zagal, E. (2017). Potential Model Overfitting in Predicting Soil Carbon Content by Visible and Near-Infrared Spectroscopy. *Applied Sciences, 7*(7).

Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. (1974). Monitoring vegetation systems in the great plains with erts. *NASA Goddard Space Flight Center 3d ERTS-1 Symposium*, pp. 309–317.

Reddy, C. S., Rajashekar, G., Krishna, P. H., Jha, C. and Dadhwal, V. (2015). Multi-source and multi-date mapping of deforestation in central india (1935-2010) and its implication on standing phytomass carbon pool. *Ecological Indicators, 57*, 219 - 227.

Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W. and Allen, R. G. et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154–172.

Sellers, P. J., Berry, J. A., Collatz, G. J., Field, C. B. and Hall, F. G. (1992). Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme.. *Remote Sensing of Environment, 42*(3), 187–216.

Schubert, P., Lagergren, F., Aurela, M., Christensen, T. and Grelle, A. et al. (2012). Modeling gpp in the nordic forest landscape with modis time series data—comparison with the modis gpp product. *Remote Sensing of Environment, 126*, 136 - 147.

Sasaki, N. and Putz, F. E. (2009). Critical need for new definitions of "forest" and "forest degradation" in global climate change agreements. *Conservation Letters, 2*(5), 226–232.

Sellers, P. J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing, 6*(8), 1335–1372.

Sellers, P. J. (1987). Canopy reflectance, photosynthesis, and transpiration, II. The role of biophysics in the linearity of their interdependence. *Remote Sensing of Environment, 21*(2), 143–183.

Trancoso, R., Sano, E. E. and Meneses, P. R. (2015). The spectral changes of deforestation in the Brazilian tropical savanna. *Environmental Monitoring and Assessment, 187*(1), 4145.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment, 8*(2), 127 - 150.

USGS (2015). *Landsat 8 Data Users Handbook*. version 1 edition..U.S. Geological Survey

Wang, Q., Adiku, S., Tenhunen, J. and Granier, A. (2005). On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sensing of Environment, 94*(2), 244–255.

Wernick, M. N. and Morris, G. M. (1988). Maximum-likelihood image classification. *Proc. SPIE Digital and Optical Shape Representation and Pattern Recognition, 938*, 317–321.

Wu, M., Wu, C., Huang, W., Niu, Z. and Wang, C. (2015). High-resolution leaf area index estimation from synthetic landsat data generated by a spatial and temporal data fusion model. *Computers and Electronics in Agriculture, 115*, 1 - 11.

Wang, B., Zhang, G. and Duan, J. (2015). Relationship between topography and the distribution of understory vegetation in a pinus massoniana forest in southern china. *International Soil and Water Conservation Research, 3*(4), 291 - 304.

Xie, Y., Wang, P., Bai, X., Khan, J. and Zhang, S. et al. (2017). Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using landsat imagery and the ceres-wheat model. *Agricultural and Forest Meteorology, 246*, 194 - 206.

Yin, G., Li, A., Jin, H., Zhao, W. and Bian, J. et al. (2017). Derivation of temporally continuous lai reference maps through combining the lainet observation system with cacao. *Agricultural and Forest Meteorology, 233*, 209 - 221.

Zarate-Valdez, J. L., Whiting, M. L., Lampinen, B. D., Metcalf, S. and Ustin, S. L. et al. (2012). Prediction of leaf area index in almonds by vegetation indexes. *Computers and Electronics in Agriculture, 85*, 24 - 32.

GRASS Development Team (2017). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.2*. .Open Source Geospatial Foundation

DISCUSIÓN GENERAL y CONCLUSIONES

La espectroscopía aplicada al análisis del COS, es una técnica rápida dado que no requiere una preparación compleja de las muestras. En las condiciones de este estudio, usando el rango espectral visible y parte del infrarrojo cercano (350-900 nm), la reflectancia permitió distinguir muestras de suelo con diferencias en el contenido de C superiores a 1.4%. Diferencias menores pueden causar que las curvas espectrales se intercepten entre si, imposibilitando distinguir las muestras de suelo por su contenido de C.

El método de validación cruzada (cross-validation) junto con la regresión por mínimos cuadrados parciales, resulta útil para dos objetivos: **1.)** para determinar el número óptimo de variables latentes o factores que se utilizarán en el modelo final, y **2.)** para determinar si los datos espectrales son consistentes y pueden ser utilizados para la generación de un modelo estadístico que estime el valor del C en muestras de suelo no analizadas en el laboratorio.

Una densidad aparente menor a 1 $gm \cdot cm^{-3}$ en los suelos (como el orden Andisol) puede dispersar la luz natural incidente a un nivel en que pierde significativamente la correlación con la variable a estudiar. Esto puede ser comprobado a través de la validación cruzada donde según la teoría, a mayor número de variables latentes el modelo se ajusta mejor o la variación del ajuste es mínima. Cuando ocurre lo contrario, es decir que con un mayor número de variables latentes el indicador de ajuste del modelo (R^2 por ejemplo) disminuye; esto significa que los datos no son consistentes y por lo tanto no sirven para generar un modelo predictivo.

Los estados de degradación de un bosque de *Nothofagus obliqua* pueden observarse mediante la respuesta espectral (reflectancia) de su biomasa aérea. Sin embargo, el método de comparación múltiple con un $\alpha = 0.05$ fue el único que pudo revelar las diferencias estadísticas entre los diferentes estados de degradación. Adicionalmente, la orientación de la pendiente influye en la densidad de la biomasa aérea de los diferentes estados de degradación del bosque. Los parches de bosque localizados en una pendiente con exposición sur presentan mayor índice de área foliar (IAF) que los expuestos a la orientación norte.

El NDVI y la banda NIR del sensor OLI de Landsat 8, representan claramente la evolución fenológica del bosque. La banda NIR es menos sensible a los efectos de la atmósfera por lo tanto, puede ser utilizada como única fuente de información para el análisis de la biomasa utilizando el método propuesto en esta investigación. Sin embargo, debido a los efectos topográficos la reflectancia puede no reflejar la condición real de la biomasa bosque.

Los remanentes de bosques nativos son de gran interés por las funciones ecológicas que cumplen como secuestro de C, conservación de fauna y flora, entre otras. El desarrollo de herramientas y métodos que permitan detectar signos de degradación de forma simple y rápida, son esenciales para la gestión de este recurso natural. La metodología evaluada en este estudio confirma que la información espectral es útil para realizar un diagnostico acertado y oportuno de la condición del bosque de *Nothofagus obliqua* de Chile. La disponibilidad gratuita de la imágenes del satélite Landsat 8 brinda la oportunidad de monitorear de forma constante la biomasa forestal. La principal limitación en la obtención de imágenes es la cobertura por nubes, principalmente en la época de invierno y otoño. La aplicación de métodos de corrección de imágenes puede mejorar la exactitud de los análisis. Sin embargo, dicho tema quedó fuera del alcance de este estudio. Los resultados de este trabajo permitieron aceptar la hipótesis planteada sobre la evaluación de la biomasa aérea a través de la reflectancia. Estos resultados quedan como referente del estado del bosque de *Nothofagus obliqua* evaluado en las parcelas experimentales seleccionadas.

El C orgánico es el factor con más relevancia en estudios relacionados con el cambio climático y el suelo. La espectroscopía, ha probado ser una técnica simple, económica y precisa para la estimación de las propiedades del suelo entre ellos, el contenido de C orgánico. Sin embargo, hemos demostrado que en suelos con densidades aparentes menores a 1 $g \cdot cm^{-3}$, que es el caso de los Andisoles, el potencial de esta técnica se ve afectado debido a una mayor dispersión de la energía natural incidente, la cual no puede ser compensada por medio del filtro Savitzky-Golay el cual comúnmente utilizado para este propósito. Si este fuese el caso se recomienda que la muestra sea molida antes de ser escaneada con un espectroradiómetro. Los resultados obtenidos para las condiciones de este estudio, permitieron rechazar la hipótesis planteada respecto a uso de la reflectancia como fuente de datos para la generación de un modelo matemático para predecir el C orgánico del suelo. Los antecedentes expuestos en esta tesis, son de gran ayuda para futuras aplicaciones de la espectroscopia, especialmente en su aplicación directamente en el campo.