



UNIVERSIDAD DE CONCEPCIÓN
Facultad de Ciencias Físicas y Matemáticas
Departamento de Geofísica

Generación de mapas temáticos de uso de suelo pre y post terremoto y tsunami del 27 de febrero del 2010 para localidades afectadas de la región del Biobío a partir de imágenes satelitales, usando algoritmos de clasificación de aprendizaje SVM y Random Forest.

Habilitación Profesional
para optar al título de Geofísico

Matias Ivan Volke Oyarzun

Profesores guías: Dr. Rodrigo Abarca Del Rio (DGEO, Universidad de Concepción)

Dr. Timothy Warner (Department of Geology and Geography, West Virginia University, USA).

Concepción, Chile

Octubre, 2017

UNIVERSIDAD DE CONCEPCIÓN
Facultad de Ciencias Físicas y Matemáticas
Departamento de Geofísica

Generación de mapas temáticos de uso de suelo pre y post terremoto y tsunami del 27 de febrero del 2010 para localidades afectadas de la región del Biobío a partir de imágenes satelitales, usando algoritmos de clasificación de aprendizaje SVM y Random Forest.

Matias Ivan Volke Oyarzun

Habilitación Profesional
para optar al título de Geofísico

Profesores guías: Dr. Rodrigo Abarca Del Rio (DGEO, Universidad de Concepción)
Dr. Timothy Warner (Department of Geology and Geography, West Virginia University, USA).

Comisión: Dra. Ignacia Calisto (DGEO, Universidad de Concepción)
Msc. Oscar Cifuentes (Departamento de Geografía, Universidad de Concepción)
Dr. Jorge Quezada (Departamento de Geología, Universidad de Concepción)



Concepción, Chile

Octubre, 2017

Agradecimientos

Agradezco a mi familia, la cual siempre estuvo entregándome apoyo y ánimo para continuar mis estudios.

Resumen

En el presente trabajo se describen algunas de las localidades urbanas más afectadas por el terremoto y tsunami del 27 de febrero de 2010 en la región del Biobío, Chile, en base a mapas temáticos de uso de suelo, generados usando métodos de clasificación digital de imágenes satelitales obtenidas desde LandSat TM (30m) y ASTER (15m) para periodos de tiempo antes y después ocurrido el terremoto y tsunami. Los métodos de clasificación a utilizar son los cada vez más populares SVM y Random Forest, algoritmos de aprendizaje que vienen a reemplazar a los antiguos métodos requeridos comúnmente para este tipo de estudios como el de Mínima distancia y Máxima Verosimilitud. Para esto se hará una comparación mediante métodos estadísticos como matrices de confusión e índices kappa para demostrar la superioridad en precisión de estos nuevos algoritmos de aprendizaje frente al de Máxima Verosimilitud.

La región de estudio involucrada se enmarca en las costas de la región del Biobío que forman parte de un sistema conjugado de bahías tectónicas abiertas al norte. Los procesos tectónicos y eustáticos del Cuaternario generaron rasgos fisiográficos característicos con extensas planicies de playa y áreas húmedas que hoy en día son ocupadas extensamente por centros urbanos y emplazamientos de infraestructura costera, destacando puertos y caletas pesqueras. En esta costa se encuentra el tercer conglomerado urbano del país (Concepción metropolitano), el cual ha sido históricamente desarrollado entorno al eje costero. Es así como de las 54 comunas que conforman la región, 14 corresponden a comunas costeras, las que agrupan al 54% de la población regional. Esta realidad geográfica hace que las áreas costeras se encuentren permanentemente con riesgo de inundación por tsunami, dado que la mayoría de los terremotos tsumanigenos ocurren como fenómenos locales, es decir generados dentro de la zona de subducción del margen continental chileno, a menos de 200 Km de la costa. LLico, Arauco, Tubul son localidades costeras ubicadas en la provincia de Arauco región del Biobío y Dichato igualmente pueblo costero de la comuna de Tome región del Biobío. Estas localidades fueron de las más afectadas por el terremoto y tsunami del 27/02/2010, siendo estos lugares correspondientes a sectores socio-económicos deprimidos en comparación con el resto del territorio nacional y la mayoría carencia de estudios de riesgos y planes de emergencia locales al momento de generarse el sismo. Por estas razones estas localidades son ideales para estudiar los efectos geomorfológicos y geográficos que tuvieron el terremoto y tsunami del 2010 en estos lugares. La generación de mapas temáticos de uso de suelo nos permiten hacer una buena comparación de la cobertura de suelo para antes y después del 27/F, obteniendo resultados válidos para generar un catastro de los efectos que generó el evento catastrófico.

Índice General

1- Introducción	1
1.1- Terremoto y tsunamis, 27 de febrero del 2010.....	1
1.2- Imágenes satelitales.....	4
2- Hipótesis y Objetivos.....	8
2.1- Hipótesis	8
2.2- Objetivo General	8
2.3- Objetivos específicos	8
3- Metodología	10
3.1- Área de estudio	10
3.2- Marco teórico.	17
3.2.1- Teledetección e imágenes satelitales.	17
3.2.2- Corrección Geométrica.	20
3.2.3- Correcciones Radiométricas.	21
3.2.4- Correcciones topográficas.	22
3.2.5- Imágenes LandSat 5 (sensor TM).....	23
3.2.6- ASTER.....	25
3.2.7- Combinación de las bandas.....	25
3.2.8- Clasificación de la imagen.	26
3.2.9- Validación de los algoritmos: matriz de confusión e índice Kappa	34
3.3- Materiales y metodología.	35
3.3.1- Lectura de datos	36
3.3.2- Correcciones.....	37
3.3.3- Clasificación.....	37
3.3.4- Combinación de bandas	38
3.3.5- Creación de áreas de entrenamiento.....	41
3.4- Preparación de datos.	49
3.5- Procesamiento de datos	50
3.5.1- Entrenando el clasificador.....	51
3.5.2- Optimizando los algoritmos.....	51
4- Resultados	53
5- Discusión	63
6- Conclusión:	80
Bibliografía.....	84

Índice de Figuras

1.1.1. Epicentro terremoto del 27 de febrero del 2010, Chile. (USGS).....	2
1.1.2. Imágenes del nivel de destrucción provocados por el terremoto y tsunami del 2010.....	3
1.2.1. Esquema métodos de clasificación ocupados en Teledetección. Crédito: Elaboración propia.....	6
3.1.1. Localidades estudiadas en este trabajo, afectadas por el terremoto y tsunami, Biobío, Chile.....	10
3.1.2. Mapa geológico sector Arauco-Lebu.....	11
3.1.3. Localidad de Dichato.....	12
3.1.4. Localidad de Llico.....	14
3.1.5. Localidad de Arauco.....	15
3.1.6. Localidad de Tubul.....	16
3.2.1.1. Esquema ilustrativo para el proceso en el que se basa la Teledetección.....	17
3.2.1.2. Espectro electromagnético.....	18
3.2.1.3. Respuesta espectral (Reflectividad) en distintos rangos de longitud de onda para diferentes coberturas de suelo típicas.....	19
3.2.1.4. Cañones RGB- Rojo/Verde/Azul.	20
3.2.8.2.1. Máquinas de vector de soporte lineal (adaptado desde Burges, 1998).....	28
3.2.8.3.1. Esquema del procedimiento general de un clasificador random forest (Adaptado desde Kumaresan et.al., 2014).....	32
3.3.2.1. Esquema sobre el pre-procesamiento de las imágenes satelitales.....	37
3.3.3.1. Esquema sobre el proceso de clasificación de las imágenes.....	38
3.3.4.1. Imágenes de las combinaciones RGB para los periodos antes y después del 27 de febrero del 2010.....	40
3.3.5.1. Áreas de entrenamiento en imágenes LandSat para periodo previo al 27 de febrero del 2010.....	43
3.3.5.2. Áreas de entrenamiento en imágenes LandSat para periodo posterior al 27 de febrero del 2010.....	44

3.3.5.3. Áreas de entrenamiento en imágenes ASTER para periodo previo al 27 de febrero del 2010.....	45
3.3.5.4. Áreas de entrenamiento en imágenes ASTER para periodo posterior al 27 de febrero del 2010.....	46
3.3.5.5. Respuestas espectrales de las diferentes coberturas de interés a lo largo de las 6 bandas para imágenes LandSat.....	48
3.3.5.6. Respuestas espectrales de las diferentes coberturas de interés a lo largo de las 3 bandas para imágenes ASTER.....	49
4.1 Mapas temáticos obtenidos mediante la clasificación del modelo SVM para periodo previo al 27 de febrero del 2010.....	58
4.2 Mapas temáticos obtenidos mediante la clasificación del modelo RF para periodo previo al 27 de febrero del 2010.....	59
4.3 Mapas temáticos obtenidos mediante la clasificación del modelo SVM para periodo posterior al 27 de febrero del 2010.....	60
4.4 Mapas temáticos obtenidos mediante la clasificación del modelo RF para periodo posterior al 27 de febrero del 2010.....	61
4.5 Mapas temáticos obtenidos mediante la clasificación del modelo MV para periodos previo y posterior al 27 de febrero del 2010.....	62
5.1 Mapas temáticos para localidad de Dichato y área de inundación por tsunami del 2010.....	65
5.2 Mapas temáticos para localidad de Llico y área de inundación por tsunami del 2010.....	66
5.3 Mapas temáticos para localidad de Arauco y área de inundación por tsunami del 2010.....	68
5.4 Mapas temáticos para localidad de Tubul y área de inundación por tsunami del 2010.....	70
5.5 Cambios en cobertura de suelo y morfológicos en la localidad de Dichato.....	73
5.6 Cambios en cobertura de suelo y morfológicos en la localidad de Llico.....	75
5.7 Cambios en cobertura de suelo y morfológicos en la localidad de Arauco.....	77
5.8 Cambios en cobertura de suelo y morfológicos en la localidad de Tubul.....	79

Índice de Tablas

3.2.5.1. Descripción para las bandas LandSat 5 (sensor TM).....	24
3.2.6.1. Descripción para las bandas ASTER.....	25
3.2.9.1. Ejemplo Matriz de confusión (se expresan los valores de la matriz en proporción del número total de sitios).....	34
3.3.1. Tabla descriptiva de las imágenes satelitales utilizadas.....	35
3.5.1.1. Evaluación de importancia de las variables (bandas) de las composiciones de bandas LandSat y ASTER en los periodos previo y posterior al 27/F.....	52
4.1 Matrices de confusión e índices kappa para los modelos de clasificación SVM, RF y MV.....	56
5.1 Tabla comparativa de clases temáticas previo y posterior al 27/F para la localidad de Dichato.....	72
5.2 Tabla comparativa de clases temáticas previo y posterior al 27/F para la localidad de Llico.....	74
5.3 Tabla comparativa de clases temáticas previo y posterior al 27/F para la localidad de Arauco.....	76
5.4 Tabla comparativa de clases temáticas previo y posterior al 27/F para la localidad de Tubul.....	78

1- Introducción

Tras el terremoto y tsunami ocurridos el 27 de febrero del 2010 en las costas chilenas, surge la necesidad de investigar sobre los cambios en la cobertura y uso de suelo, y además sobre cambios morfológicos ocurridos en las zonas más afectadas por este desastre. Dicha necesidad se ve apoyada en requerimientos futuros para una mejor implementación de zonas y mapas de riesgo que tomen como base las áreas en donde grandes cambios hayan ocurridos, partiendo de estudios de series de tiempo para imágenes satelitales. Estudios previos de esta índole ya se han realizado. [Rojas \(2013\)](#) identificó la cobertura y uso de suelo pre y post terremoto del 2010 en Concepción metropolitana, basándose en análisis temporales usando algoritmos de Máxima Verosimilitud para clasificación digital de imágenes satelitales LandSat.

Por otra parte [Landereros \(2014\)](#) realizó un análisis de áreas afectadas por el terremoto y tsunami de 2010 en Constitución, por medio de técnicas de generación de mapas de uso de suelo a través de imágenes satelitales de alta resolución (FormoSat), trabajo en el cual logro identificar áreas con claros cambios de cobertura de suelo en Constitución.

Estos trabajos nos dan una base metodológica y una muestra de los resultados que se pueden obtener. Pero para este caso se hará una diferencia importante en los algoritmos de clasificación a utilizar, implementándose los algoritmos de aprendizaje support vector machine (SVM) y random forest (RF), los cuales no han sido muy requeridos en la generación de mapas de cobertura y uso de suelo y los que deberían entregar una clasificación más precisa y confiable, además de conllevar un costo computacional considerablemente menor. Todo esto nos llevaría a respaldar estudios previos que dan muestra de la mejora en precisión de estos nuevos métodos de aprendizaje comparativamente a los ya bien conocidos métodos clásicos ([Dalponte et.al., 2009](#); [Rodriguez-Galiano et.al., 2011](#))

LLico, Arauco, Tubul son localidades costeras ubicadas en la provincia de Arauco región del Biobío y Dichato igualmente pueblo costero de la comuna de Tome región del Biobío. Estas localidades fueron de las más afectadas por el terremoto y tsunami del 27F, siendo estos lugares correspondientes a sectores socio-económicos deprimidos en comparación con el resto del territorio nacional y la mayoría carentes de estudios de riesgos y planes de emergencia locales al momento de generarse el sismo. Por estas razones estas localidades son ideales para estudiar los efectos geomorfológicos y geográficos que tuvieron el terremoto y tsunami del 2010 en estos lugares.

1.1- Terremoto y tsunami, 27 de Febrero del 2010

El territorio chileno, en la extensión de su costa, es complejo de planificar, puesto que se encuentra bajo la amenaza recurrente de desastres naturales. Dentro de todos estos fenómenos naturales, el que con mayor violencia ha dejado huellas en Chile lo constituyen los terremotos, muchos de los cuales han venido acompañados de tsunami. La explicación de este hecho radica en la posición de Chile, ya que el territorio nacional se ubica, de norte a sur, en las inmediaciones de la zona en que colisionan las placas de Nazca y Sudamérica, en una situación en que la primera se hunde bajo la segunda. Este proceso genera una acumulación de energía que se libera cada cierto periodo de tiempo a través de movimientos telúricos de magnitud

diversa y con epicentro en distintas áreas específicas de la zona de contacto entre las placas Nazca y Sudamericana (Fariña L., Opazo C., Vera P., 2012. *“Impactos ambientales del terremoto y tsunami en Chile: Las réplicas ocultas del 27/F”*. Fundación TERRAM.).

El 27 de febrero de 2010 la costa de Chile central fue afectada por un terremoto $M_w = 8,8$ de acuerdo al Servicio Sismológico de Chile y el Servicio Geológico de Estados Unidos. Generó un tsunami de magnitud 4 que afectó principalmente bahías pequeñas en un tramo costero de 800 km, mayormente en las regiones del Maule y Biobío (37°S). Tuvo su epicentro a 50 km. al noreste de la ciudad de Concepción, y a 63 km al Suroeste de la ciudad de Cauquenes de la Región del Maule, frente a las costas entre Curanipe y Cobquecura y a 47,7 km de profundidad bajo tierra (Figura 1.1.1).

El evento en general fue de tal magnitud que generó cambios profundos en diversos aspectos, incluyendo modificaciones en la geografía del territorio. Provocó alzamientos y subsidencias co-sísmicas. El litoral de la región del Biobío tuvo un alzamiento co-sísmico generando el ensanchamiento de playas, la emersión de la plataforma de abrasión marina, desecamiento de ríos y humedales. La elevación del suelo fue de hasta 2.7 metros en la provincia de Arauco (Quezada et. al, 2012). Las ciudades y pueblos de la zona sufrieron grandes daños en vivienda, edificios, obras de infraestructura vial, sanitarias y energéticas.

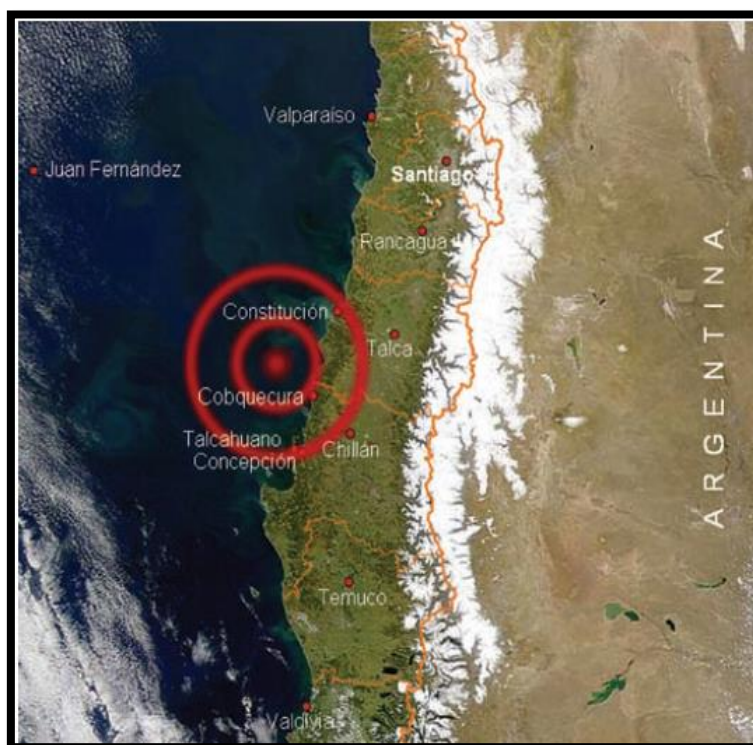


Figura 1.1.1- Epicentro del terremoto del 27 de febrero del 2010 (USGS).

Además, el terremoto trajo asociado un tsunami que azotó las costas de Chile. La alerta de tsunami generada para el océano Pacífico se extendió posteriormente a 53 países ubicados en gran parte de la cuenca. En el litoral chileno este tsunami alcanzó su máximo impacto desde Constitución hasta el golfo de Arauco. La ola más destructiva se produjo aproximadamente a los 90 minutos después del terremoto en

la región del Biobío y su altura vario entre 1 a 4 metros en las costas del centro-sur chileno.

En este sentido, la orientación y la morfología del litoral son determinantes en los efectos del tsunami. Es muy claro el impacto que tuvo en las costas y bahías, en donde se recibieron de manera directa el tren de olas, siendo las bahías muy vulnerables debidas a poseer costas semi-cerradas, donde las olas chocan y rebotan contra sus costas y se produce un efecto de resonancia que amplifica los efectos destructivos de las olas. Este es el caso para la bahía de Dichato, la bahía de Concepción-Talcahuano y del golfo de Arauco, que fueron fuertemente afectadas ([Martínez C., 2011](#)). El mar golpeó fuertemente la costa y penetro por los puntos más vulnerables que corresponden a desembocaduras de ríos y esteros, inundando cientos de metros hacia el interior.

Para visualizar e ilustrar la magnitud y consecuencias de este desastroso evento, solo se tiene que recordar la imagen que se repitió a lo largo de todo el centro-sur de Chile, edificios colapsados, casas destruidas y en definitiva hogares y fuentes de trabajo arrasadas por la fuerza de la naturaleza, como se ilustra en la figura 1.1.2. Todo esto develó una serie de carencias en política pública y gestión, respecto a los cursos de acción frente a una emergencia como la vivida ([Ramírez P. y Sandoval J., 2010.](#) "Tsunami paso a paso: los escandalosos errores y omisiones del SHOA y la ONEMI". Fuente: CIPER.; [Delfin D., 2010.](#) "Cuanto le costó el terremoto a Chile." Fuente: LA TERCERA.). En el mismo sentido los dramas humanos vividos por los habitantes de las zonas afectadas inundaron todos los medios de comunicación y nos conectaron con la magnitud del desastre, encontrándonos con historias de agonía y muerte como las ocurridas en el rio Maule o de las personas que se encontraban en Isla Mocha, que vieron como familias flotaban con sus hijos y se aferraban a lo que fuera ([Guzmán J., 2010.](#) "La tragedia en 2010". Por: Roció Montes. Fuente: EL PAIS.).



Figura 1.1.2- Imágenes del nivel de destrucción provocados por el terremoto y tsunami del 2010 en diferentes ciudades. ([Crédito: Muñoz R. Revista Urbano, octubre 2010.](#))

Posterior al terremoto y tsunami se han realizado varios estudios para ver los efectos geomorfológicos en la zona, los cuales se asocian a cambios en el nivel base de ríos como en la línea de costa, a erosión en playas, a destrucción y recuperación de barras litorales entre otros (Quezada et.al., 2010; Morton et.al., 2011; Cienfuegos et.al., 2014). Además con el propósito de analizar los efectos territoriales generados por el tsunami en las costas de la región, se iniciaron múltiples trabajos en terreno, para determinar entre otras cosas la altura y run-up de la ola, catastrar efectos sobre la población y localizar áreas de mayor impacto (Quezada et.al., 2010; Vargas G. y Farías M., (2011)). Un catastro sobre los daños y áreas afectadas realizado por el departamento de Desarrollo Urbano e Infraestructura de la Secretaría Ministerial de Vivienda y Urbanismo, dio como resultado que los principales efectos del tsunami son las inundaciones generadas por el mar sobre ciudades y pueblos, destruyendo viviendas, mobiliario urbano, calles y caminos, infraestructura portuaria y vegetación (*“Impactos ambientales del terremoto y tsunami en Chile”*. Fundación Terram (2012)).

1.2- Imágenes satelitales:

Extraer información digital desde imágenes generadas por percepción remota ha sido aplicado satisfactoriamente en las últimas décadas. Los datos usados en este tipo de procesamiento provienen de datos multiespectrales, lo cual ha entregado una mejor herramienta para múltiples campos de investigación relativos a la observación de la tierra (Chuvieco, 2006). Es indudable hoy en día, la percepción remota es una de las fuentes más importantes de obtención de información vinculada al ambiente. Actualmente, el uso de los satélites ofrece un programa continuo de adquisición de datos para el mundo entero, con una periodicidad que va desde horas a varias semanas y lo que es importante es el acceso al formato digital que permite una rápida interpretación e integración de los datos a trabajos como estudios de uso de suelo y coberturas (Dalponte et. al., 2009).

Las técnicas de percepción remota son una fuente para proveer información sobre áreas sometidas a desastres naturales. La amplia cobertura de las imágenes, permite acceder rápidamente a aquellas áreas sometidas a estos eventos. La agilidad y la reducción de costos a través de la utilización de imágenes satelitales vienen acompañada de una calidad cada vez mayor en lo que se refiere a resolución espacial y temporal (Lillesand et.al., 2004).

En la última década, avances en tecnologías espaciales y computacionales han hecho posible tener una larga cantidad de datos sobre la tierra y medio ambiente. Encontrándose en estos datos cada vez en mayor medida no solo información espectral, sino que también se puede tener mapas forestales, mapas de cobertura de suelo, datos de radar e información topográfica como elevación y pendiente. La precisión en la clasificación de imágenes satelitales es una importante tarea para muchas aplicaciones prácticas, como la agricultura de precisión, monitoreo y manejo del medio ambiente, temas relevantes a defensa y para aportar información relevante en cuanto a daños producidos en desastres naturales, lo que vendría a colaborar a una mejor planificación territorial.

La percepción remota para la generación de imágenes satelitales abre un nuevo medio para poder realizar catastros sin la necesidad de un gran despliegue en terreno,

los cuales muchas veces son de difícil acceso incurriendo en costos monetarios importantes. Estudios de cambio de cobertura de suelo, pueden posibilitar la generación de estudios de vulnerabilidad para diferentes localidades pobladas afectadas por estos tipos de desastres naturales, gracias a la comparación de los escenarios pre y post tsunami, generando mapas de riesgo por tsunami para eventos de similares características, que sirvan posteriormente como instrumentos de gestión de riesgo. En este sentido varios estudios dan muestra de lo realizado en el campo de la percepción remota sobre desastres naturales. [Richmond B. \(2012\)](#), informa sobre la erosión, depósito y cambio de suelo asociado al tsunami de Tohoku-oki, mediante el contraste entre imágenes satelitales y estudio de campo. [Tappin R., Evans H. \(2012\)](#), mostró el cambio costero en el área de Sundai producto del impacto del tsunami de Tohoku-oki, a través de la interpretación de series de tiempo de imágenes satelitales y observaciones de campo. [Diao, F., Xiong, X. \(2014\)](#), caracterizaron las áreas afectadas por el tsunami de diciembre del 2004 en el océano indico, usando imágenes de moderada resolución (Terra-Aster), siendo esta caracterización mediante la observación y detección de cambios de cobertura de suelo.

El mapeo de la cobertura terrestre utilizando datos de teledetección requiere de una caracterización preciso de categorías complejas de cobertura y uso de suelo. Mapeo de coberturas de suelo y su monitoreo es una de las mayores aplicaciones de los sensores satelitales que circundan la Tierra. El monitoreo de uso de suelo es importante porque tanto modificaciones humanas y/o naturales afectan geofísica y biofísicas propiedades de la tierra ([Bala et.al., 2007](#); [Betts et.al., 2007](#); [Bonan, 2008](#)). En la figura 1.2.1 se observa un esquema que resume una variedad de métodos de clasificación han sido utilizados para crear mapas de cobertura de suelo, usando datos de sensores remotos. Métodos de clasificación van desde el rango de algoritmos no supervisados, o sea en los que no se conoce previamente la realidad del terreno, como ISODATA o K-means a algoritmos paramétricos supervisados en los que si se conoce previamente la realidad de las coberturas de suelo en terreno, tal como el de Máxima Verosimilitud ([Chuvieco, 2006](#)) y los desarrollados más recientemente los algoritmos de máquinas de aprendizaje como redes neuronales artificiales ([Mas y Flores, 2008](#)), arboles de decisión ([Breiman, 2001](#)), máquinas de vectores de soporte ([Mountrakis, 2011](#)).

SVM es una técnica de aprendizaje estadístico no paramétrica supervisada que se basa en la utilización de una función núcleo para solucionar los problemas de inseparabilidad para clasificar de mejor forma grupos de datos, o como en este caso, categorías de coberturas de suelo. Además no posee ninguna suposición de distribución de los datos subyacentes.

En su forma original ([Vapnik, 1979](#)), el método parte con un conjunto de etiquetados ejemplos de entrenamiento y tiene como objetivo encontrar un hiperplano que separe el conjunto de datos en un número predefinido discreto de clases en una manera coherente a los ejemplos de entrenamiento.

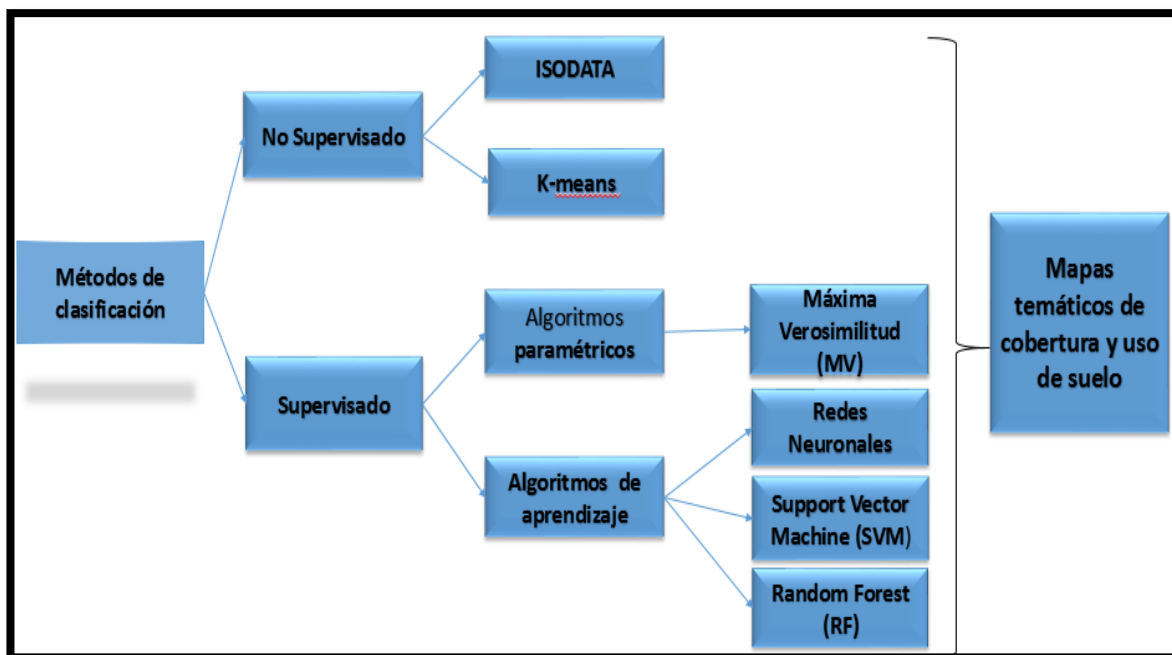


Figura 1.2.1.- Esquema métodos de clasificación ocupados en Teledetección.
Crédito: Elaboración propia.

Los SVMs son particularmente recurridos en el campo de percepción remota debido a su habilidad de clasificar con una buena precisión incluso con un número limitados de áreas de entrenamiento, una limitante común en aplicaciones de la percepción remota. Las aplicaciones SVM han sido probadas en varios trabajos. [Zhu y Blumberg \(2002\)](#), usando imágenes ASTER para el mapeo de áreas urbanas, obtuvo diferentes resultados dependiendo de la resolución de la imagen y la elección de la función núcleo (Ver Marco teórico sección 3.2.8.2.- SVM). [Esch et. al. \(2009\)](#), usando métodos SVM y la combinación de imágenes satelitales para derivar información sobre áreas urbanas, diferenciando las áreas industriales, residenciales y construcciones viales. [Andermann y Gloaguen \(2009\)](#), estudiaron procesos de erosión en zonas orogénicas activas en el Himalaya. Esa investigación implementó SVMs para mejorar la clasificación de uso de suelo, erosión y procesos geomorfológicos. [Candade \(2004\)](#), comparó los clasificadores SVM y redes neuronales usando imágenes LandSat. Concluyó que ha menor número de ejemplos de entrenamiento, SVM entrega una mayor precisión. (Ver Marco teórico sección 3.2.8.2.- SVM)

RF es una técnica emergente de máquina de aprendizaje en el cual se utilizan conjuntos de clasificaciones para asignar un determinado dato a una clase determinada ([Friedl et. al., 2002](#); [Sesnie et. al., 2008](#); [Ghimere et. al., 2010](#)). Conjuntos de algoritmos de aprendizaje usan la misma base clasificadora para producir múltiples clasificaciones de los mismos datos ([Breiman, 2001](#)) o usar una combinación de diferentes clasificadores bases para generar una múltiple clasificación de los mismos datos ([Mountrakis et. al., 2009](#)). La colección de múltiples clasificadores de los mismos datos son combinados usando una regla de aproximación base (como máxima votación, suma o reglas Bayesianas) o basados en una iterativa minimización del error mediante la reducción del peso de los ejemplos clasificados correctamente (es decir boosting) ([Ghimere et. al., 2010](#)).

RF tiene una mayor precisión que otros algoritmos de máquinas de aprendizaje porque el grupo de clasificadores trabajan más precisamente que cualquier clasificador individual ([Ghimere et. al., 2010](#)). Existen múltiples trabajos en los que se ha ocupado RF en clasificación de coberturas de suelo usando imágenes multiespectrales e hiperespectrales ([Pal, 2005](#); [Lawrence et. al., 2006](#); [Sesnie et. al, 2008](#); [Ghimere, et. al., 2010](#)). (Ver Marco teórico sección 3.2.8.3.- RF).

En la última década, algoritmos de máquinas de aprendizaje han emergido como las más precisas y eficientes alternativas a los algoritmos paramétricos convencionales. Esto último queda en evidencia cuando se está enfrente de una alta dimensionalidad, espacios complejos de datos y un mapeo de área extenso ([Huang et.al., 2002](#)). Estos algoritmos son eficientes y efectivos porque no tienen una suposición de distribución de datos (normalidad). En este trabajo se desarrollaran dos métodos de aprendizaje: máquinas de vectores de soporte (SVM) y random forest (RF), las cuales son unas poderosas máquinas que son relativamente poco conocidas y no han sido ampliamente evaluados en comparación a otras técnicas de reconocimiento de patrones más convencionales.

2- Hipótesis y Objetivos

Se describirán algunas de las localidades urbanas más afectadas en base a mapas temáticos de uso de suelo, generados usando métodos de clasificación digital de imágenes satelitales obtenidas desde LandSat TM (30m) y ASTER (15m) para periodos de tiempo pre y post 27 de febrero de 2010. Los métodos de clasificación a utilizar son los cada vez más populares SVM y RF, algoritmos de aprendizaje que vienen a reemplazar a los antiguos métodos requeridos comúnmente para este tipo de estudios como el de Mínima distancia y Máxima Verosimilitud.

2.1- Hipótesis

Tras el terremoto y tsunami ocurrido el 27 de febrero del 2010 en el centro Sur de Chile, diversos catastros de las zonas más afectadas y estudios geomorfológicos se realizaron con el propósito de ver y cuantificar los cambios y daños producidos por este desastre. En este trabajo se propone a los mapas de cambios de cobertura y uso de suelo como una gran herramienta alternativa o complementaria a los trabajos en terreno y otras metodologías tradicionales, dando respuesta a la inquietud sobre si esta nueva metodología nos entrega realmente resultados útiles. Esto en cuanto a la delimitación de áreas afectadas, catastros de zonas afectadas, y en general, como una herramienta para la creación de mapas de riesgo y vulnerabilidad.

La generación de mapas de cobertura de suelo debe realizarse con la mayor precisión posible por lo que se necesita buscar las mejores metodologías para la clasificación de imágenes satelitales, y así, asegurar la mejor precisión y confiabilidad de los resultados. En este sentido creo que la implementación de los recientemente desarrollados algoritmos de máquinas de aprendizaje SVM y RF, deberán mejorar los resultados obtenidos mediante los métodos paramétricos convencionales (Máxima Verosimilitud).

2.2- Objetivo General

El objetivo principal de esta investigación, es de realizar un estudio de clasificación novedoso mediante el uso de imágenes satelitales multiespectrales de moderada resolución espacial, para las localidades mencionadas anteriormente, afectadas en gran medida por el terremoto y tsunami del 27F. El estudio se centrara en dos periodos de tiempo distintos, antes y después del catastrófico evento, por lo que se tendrá cambios notorios en los usos de suelo y recubrimientos de estos para las áreas costeras.

2.3- Objetivos específicos

-Demostrar que los algoritmos de aprendizaje estadístico no paramétricos Supervisados SVM y RF arrojan una clasificación de coberturas de suelo con una precisión mucho mayor que los algoritmos paramétricos supervisados como Máxima Verosimilitud, implementándose estas dos clases de algoritmos de clasificación en imágenes multiespectrales de moderada resolución (LandSat y ASTER).

-Conocer la mecánica e implementación de los métodos de clasificación de imágenes satelitales mediante máquinas de aprendizaje como SVM y Random Forest, los cuales no han sido utilizados mayormente en estudios previos en Chile.

-Trabajar con imágenes satelitales multiespectrales recolectadas desde sus respectivos servidores, para luego ser procesadas y corregidas tanto radiométrica como topográficamente.

-Establecer áreas en donde se puede apreciar un cambio en la cobertura de suelo para las localidades a estudiar y así poder contrastar esta información con estudios previos de delimitación para áreas de inundación y catastros en terreno de áreas afectadas.

3- Metodología

3.1- Área de estudio

En esta investigación se analizaron 4 localidades de la región del Biobío. Estas tienen en común que son localidades costeras que fueron gravemente afectadas por el terremoto y tsunami del 27 de febrero. Las localidades están indicadas mediante diamantes azules en la *figura 3.1.1.*, siendo estas:

- 1) Dichato (36.55°S ; 72.933°O), provincia de Concepción, Región del Biobío.
- 2) LLico (37.2°S ; 73.5667°O), provincia de Arauco, Región del Biobío;
- 3) Arauco (37.25°S ; 73.3167°O), provincia de Arauco, Región del Biobío.
- 4) Tubul (37.2167°S ; 73.433°O), provincia de Arauco, Región del Biobío;

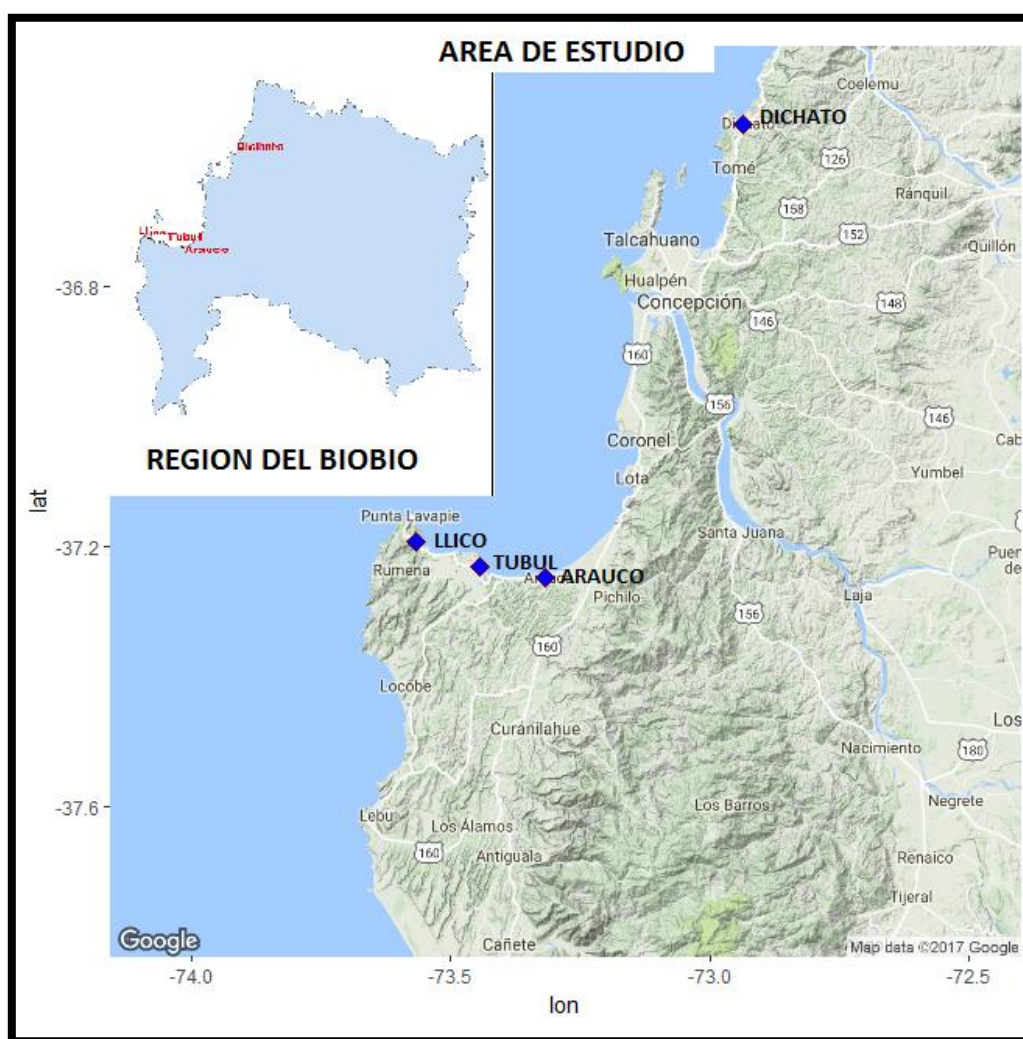
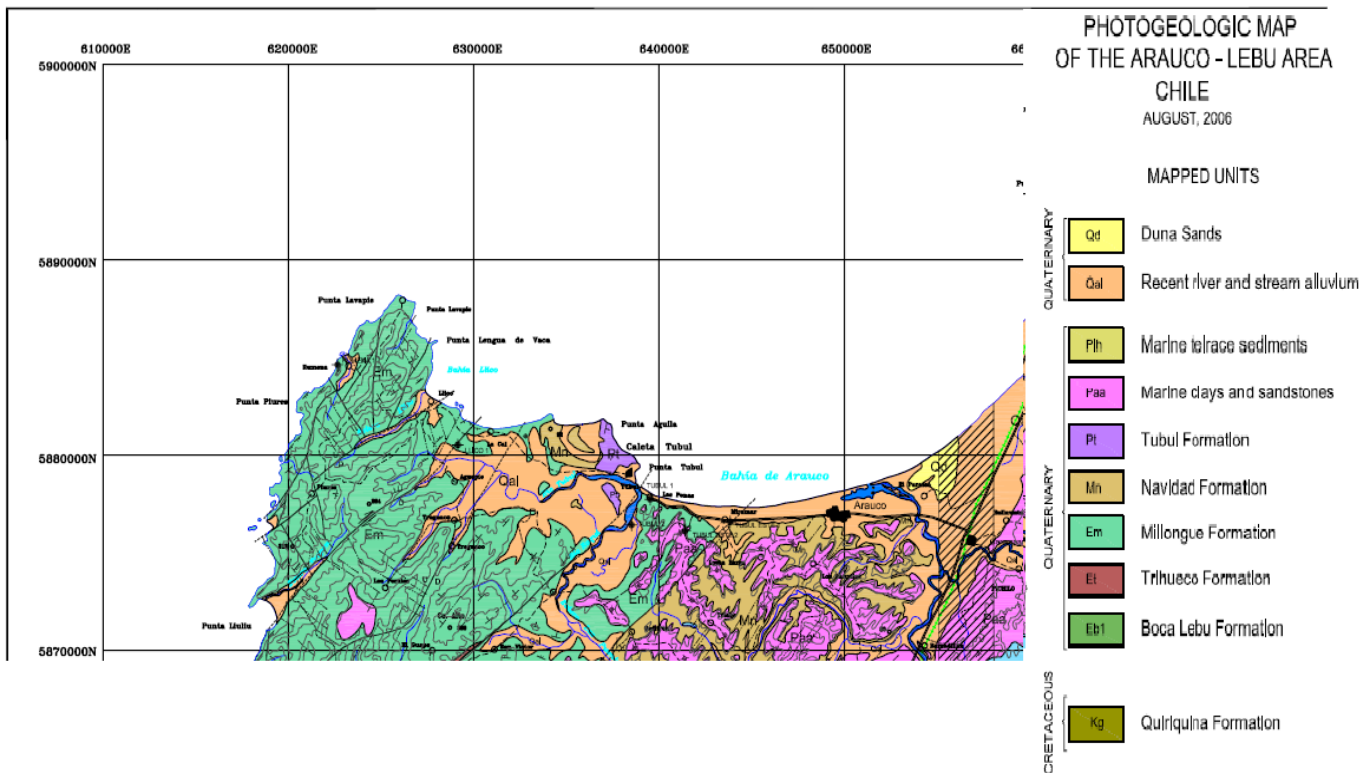


Figura 3.1.1- Ubicación localidades estudiadas en este trabajo, afectadas por el terremoto y tsunami.

MAPA GEOLOGICO PARA EL AREA ARAUCO-LEBU



*Figura 3.1.2- Mapa geológico sector Arauco-Lebu.
Fuente, SERNAGIOMIN.*

-Dichato es un pueblo costero ubicado en la comuna de Tome en la región del Biobío. El pueblo está ubicado a 37 km. al norte de la ciudad de Concepción. Su geografía es una bahía muy cerrada. El terremoto y tsunami de febrero de 2010 dejó al 80% del pueblo de Dichato destruido, inundando 80 hectáreas del poblado. Parte de esta destrucción puede visualizarse en la figura 3.1.3., donde se aprecia el gran nivel de destrucción en viviendas costeras y la gran cantidad de material que arrastró el tsunami.

La inundación se ve facilitada por la presencia del estero de Dichato, dada su baja altitud. Las características de la bahía, relacionada con su forma confinada y la posibilidad de resonancia y amplificación de ondas multiplicaron los efectos del tsunami (Morales Muñoz, Roberto (2010). *“Terremoto y Tsunami del 27 de febrero de 2010. Efectos Urbanos en Localidades de la Provincia de Arauco”*).

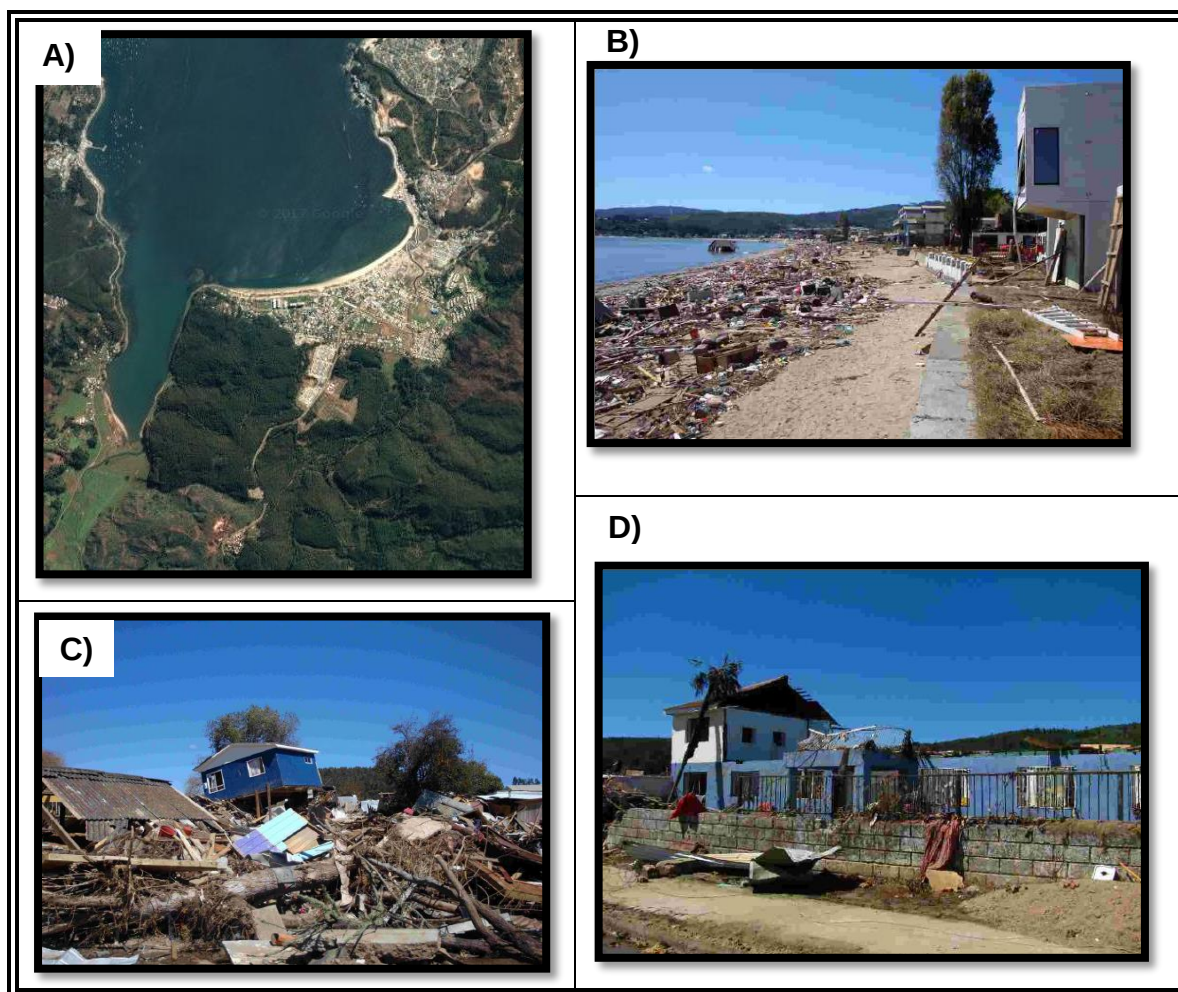


Figura 3.1.3.- Localidad de Dichato. A) Localidad de Dichato, provincia de Concepción ([Martínez C. et. al., 2011](#)); B) Extremo occidental de Dichato. Depósitos del tsunami, que incluyen una casa, entre otros materiales ([SERNAGEOMIN](#)); C) Vista sureste que muestra uno de los sectores más impactados con viviendas livianas ([SERNAGEOMIN](#)); D) Construcción de albañilería ubicada próxima al cauce del estero Dichato ([SERNAGEOMIN](#)).

-Llico: La localidad costera de Llico se encuentra ubicada en la costa del Golfo de Arauco a 12 Km. al poniente de Tubul. Producto del tsunami hubo daño total en las instalaciones pesqueras, en la costanera y en la calle principal del pueblo. También produjo un área de inundación que comprende en gran medida el sector construido, en el cual se produjo una destrucción total de las viviendas, y se enmarca dentro del área bajo la cota 5 m.s.n.m. El área de inundación total fue de 45 ha aproximadamente.

A diferencia de Tubul, Llico es un asentamiento con carácter de balneario, con abundantes viviendas de veraneo, algunas emplazadas frente a la playa, las que fueron destruidas en su totalidad. La totalidad de viviendas destruidas fue de 320 viviendas (Fariña L., Opazo C., Vera P., 2012. *Impactos ambientales del terremoto y tsunami en Chile: Las réplicas ocultas del 27/F*. Fundación TERRAM).

La destrucción ocurrida en Llico puede apreciarse en la figura 3.1.4., donde la imagen A) muestra la imagen google earth de Llico, B) tiene diferentes zonas en donde se registró un gran daño en infraestructura. Por último C) grafica el antes y después del borde costero de la localidad de Llico, visualizándose los cambios en el terreno debido al paso del tsunami.

El sitio natural de Llico corresponde a una pequeña llanura litoral de sedimentación fluvio-marina reciente, en color café claro en la figura 3.1.2., conformando una bahía relativamente angosta, y que está delimitada por una playa de borde y una duna muy intervenida por la acción humana. En ambos costados de esta llanura se desarrollan dos cursos de agua que canalizan el drenaje superficial desde las quebradas interiores hasta alcanzar el mar. El nivel del terreno de la llanura de Llico va subiendo con una suave pendiente hacia el interior, factor que impidió la penetración del mar más allá de la cota 5 m.s.n.m., evitando la destrucción de por lo menos el 50% de las viviendas emplazadas en esta localidad. Sin embargo, la orientación que presenta esta angosta bahía totalmente abierta hacia el norte y la presencia de un gran bloque o punta de cerro saliente de la línea de costa que forma parte de la plataforma sedimentaria marina del Eoceno, ubicada hacia el poniente de la playa, fueron los factores de la morfología litoral que agravaron los efectos del tsunami, al producirse un efecto de resonancia.

Debido al levantamiento del nivel del suelo generado por el sismo, la línea de costa se recogió aproximadamente 10 metros mar adentro, con respecto al nivel medio del mar que tenía antes del terremoto, quedando las instalaciones portuarias en tierra (Morales Muñoz, Roberto (2010). *“Terremoto y Tsunami del 27 de febrero de 2010. Efectos Urbanos en Localidades de la Provincia de Arauco”*).

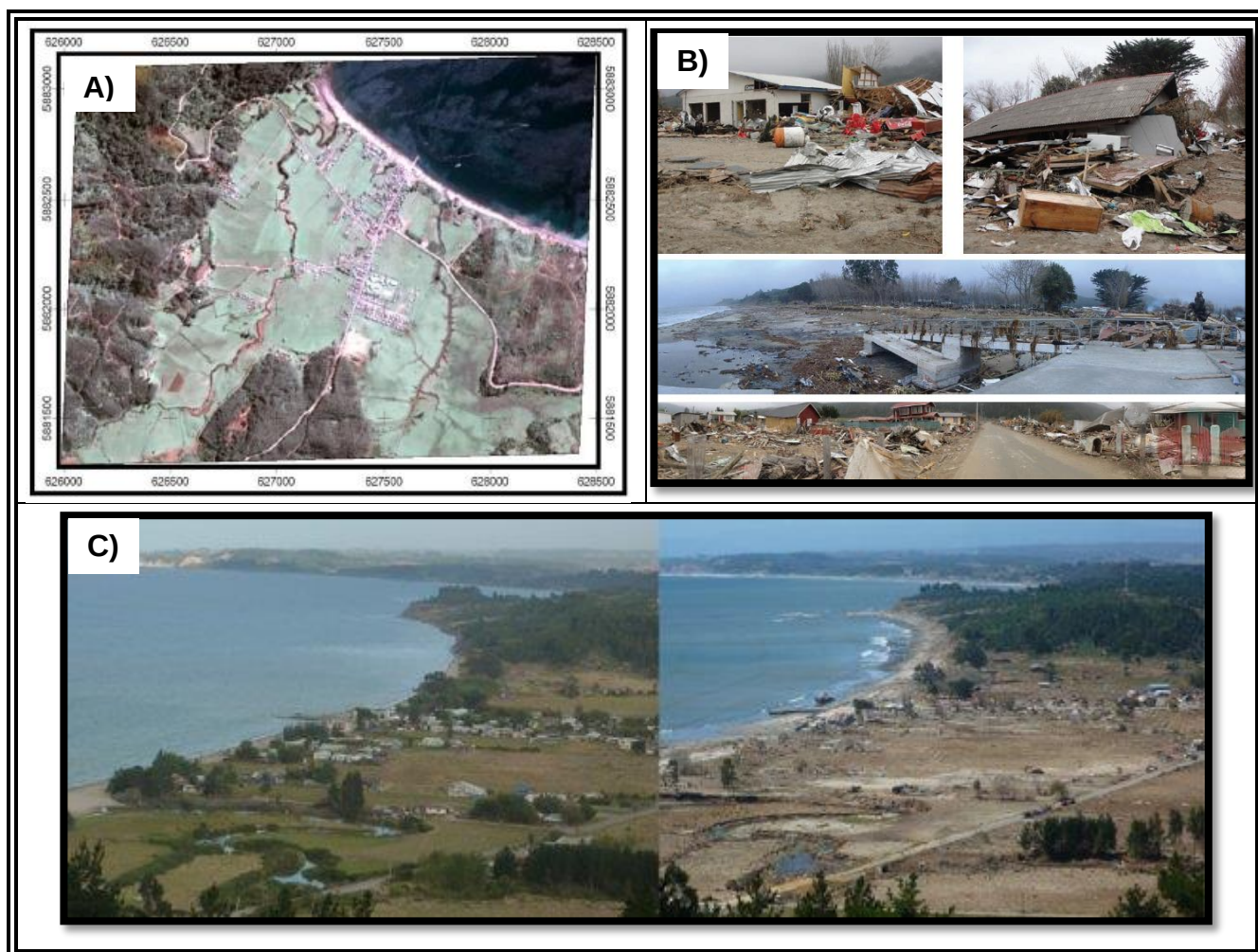


Figura 3.1.4.- Localidad de Llico. A) Localidad de Llico, golfo de Arauco ([Martínez C. et al., 2011](#)); B) Imágenes del nivel de destrucción en caleta Llico ([Foto: Roberto Morales](#)); C) Antes y después del terremoto y tsunami ([Foto: Héctor Lisboa](#)).

-Arauco es una ciudad ubicada en la provincia de Arauco en la región del Biobío. La playa de Arauco es de aproximadamente 10 km. de extensión; caracterizándose por sus arenas finas, en su mayoría de origen volcánico (ver figura 3.1.2.). La ciudad de Arauco limita al noreste con la laguna litoral del río Carampangue, extensa zona de unas 400 hectáreas de superficie que conforman el humedal, como se aprecia en la figura 3.1.5-A. Cuenta con zonas de playa, dunas, lagunas, estuarios sujetas a las variaciones de la marea. En el borde urbano de la ciudad hay una caleta de pescadores, quienes deben navegar por la laguna para acceder al mar a través de la barra del río Carampangue. En Arauco no se presentaron mayores daños debido al efecto mitigador de los humedales asociados al estuario del río Carampangue y la protección del cordón dunario que antecede el casco urbano. Solamente se evidenció ingresos del flujo en algunos sectores de la ciudad ([Fariña L., Opazo C., Vera P., 2012. Impactos ambientales del terremoto y tsunami en Chile: Las réplicas ocultas del 27/F. Fundación TERRAM](#)). En la figura de abajo se representan las consecuencias de estos flujos, que lograron alcanzar a sectores construidos y arrastrar escombros tierra adentro (ver figura 3.1.5-B).

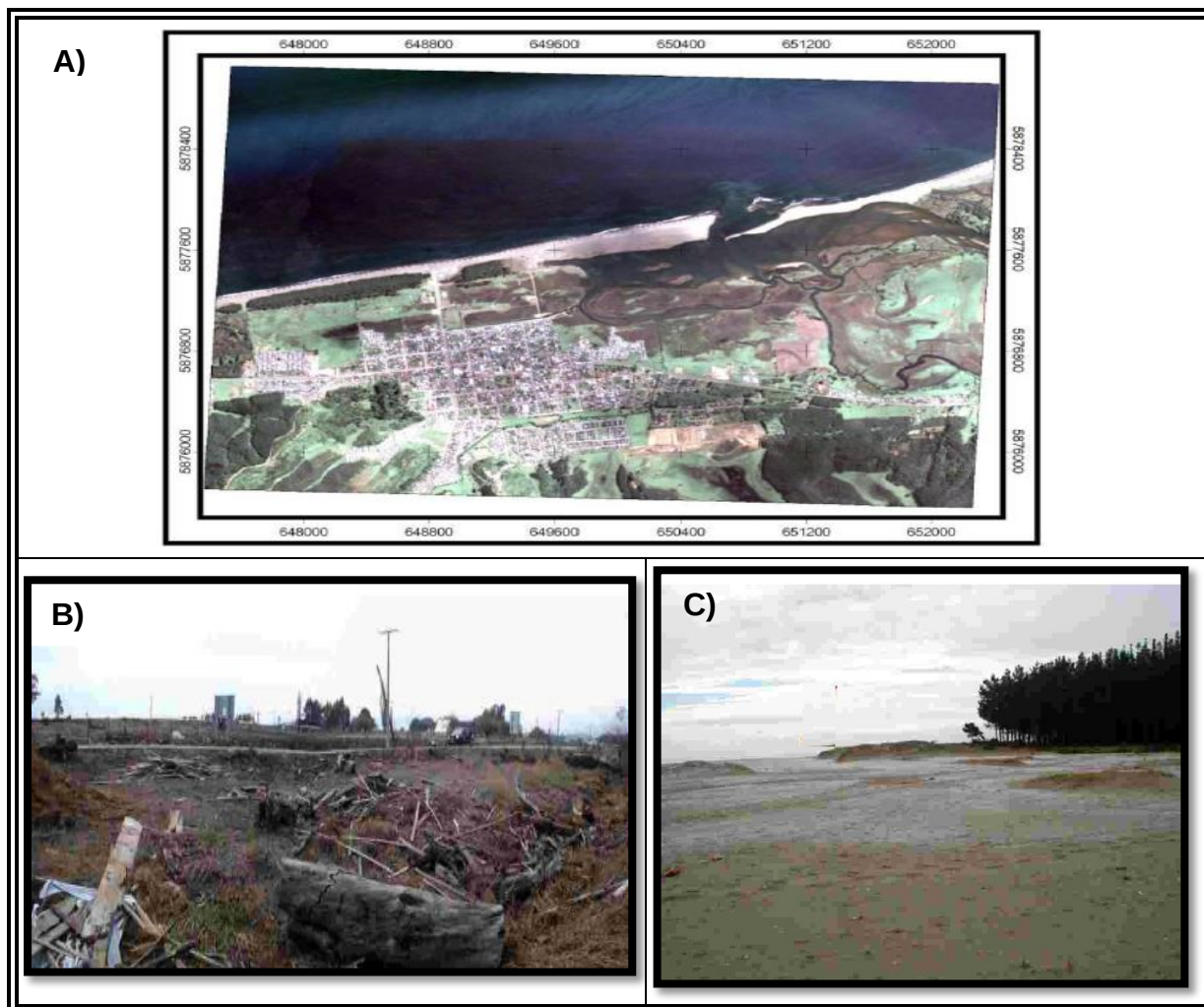


Figura 3.1.5.- Localidad de Arauco. A) Localidad de Arauco, golfo de Arauco (Martínez C. et. al., 2011); B) Vista panorámica hacia el ENE, que exhiben materiales depositados por el tsunami (SERNAGEOMIN); C) Vista al norte del humedal ubicado junto al pueblo de Arauco (SERNAGEOMIN).

-Tubul: Se encuentra ubicado en la costa del golfo de Arauco a 11 km de la ciudad de Arauco. La ciudad se encuentra próxima al humedal Tubul-Raqui, el cual tiene dos cursos de agua principales el río Tubul y el río Raqui. El tsunami produjo un área de inundación que llegó a los pies de los paleo-acantilados, abarcando una superficie aproximada de 75 hact., siendo el área de destrucción total el sector constituido por el borde de playa y de manera parcial, en la ribera del río. El número de viviendas destruidas con daño total fue de 300 viviendas y con daño parcial 400 (en el sector ribereño) (Fariña L., Opazo C., Vera P., 2012. *Impactos ambientales del terremoto y tsunami en Chile: Las réplicas ocultas del 27/F*. Fundación TERRAM). El nivel de destrucción se puede visualizar en la figura 3.1.6, donde se muestra el gran efecto en infraestructura que tuvo el terremoto y tsunami del 27 de febrero.

En cuanto a la geología de la zona, Tubul se encuentra asentado en una barra de arena de sedimentación fluvial marina reciente (cuaternario), indicada en la figura 3.1.2. con color café claro, formada por una llanura natural muy baja, ubicada entre el

pie del paleo acantilado (color morado en la figura 3.1.2) y el sistema fluvial-humedal Tubul-Raqui. Los sectores altos sobre el acantilado marino corresponden a terrazas marinas formadas por areniscas del plioceno.

Previo al tsunami, la dinámica litoral fue sedimentando arenas sobre la llanura baja, conformando una barra y ocasionando la obstrucción del drenaje superficial, dando origen a lagunas ocasionales, pajonales y vegas, en general terrenos bajos, mal drenados en su interior. Estas características explican la disposición del asentamiento humano, el cual se localiza sobre las dunas y arenas de la barra, debido a que son terrenos más altos y sin riesgo de anegamiento por precipitaciones invernales. Este es un terreno de gran fragilidad y vulnerabilidad frente a fenómenos marinos, como tsunami, ya que no poseen ninguna protección natural que los separe de la costa. Con el tsunami del 27 de febrero Tubul, si bien se vio muy afectada, no fue totalmente destruida ya que fue favorecida por su orientación, la que no está abierta ni expuesta directamente hacia el norte (ver figura 3.1.2), desde donde venía el frente de olas. También favoreció la gran boca que forma el humedal en su desembocadura, la que al ser un área baja el agua privilegia su penetración por ahí ([Morales Muñoz, Roberto \(2010\). Terremoto y Tsunami del 27 de febrero de 2010. Efectos Urbanos en Localidades de la Provincia de Arauco](#)).

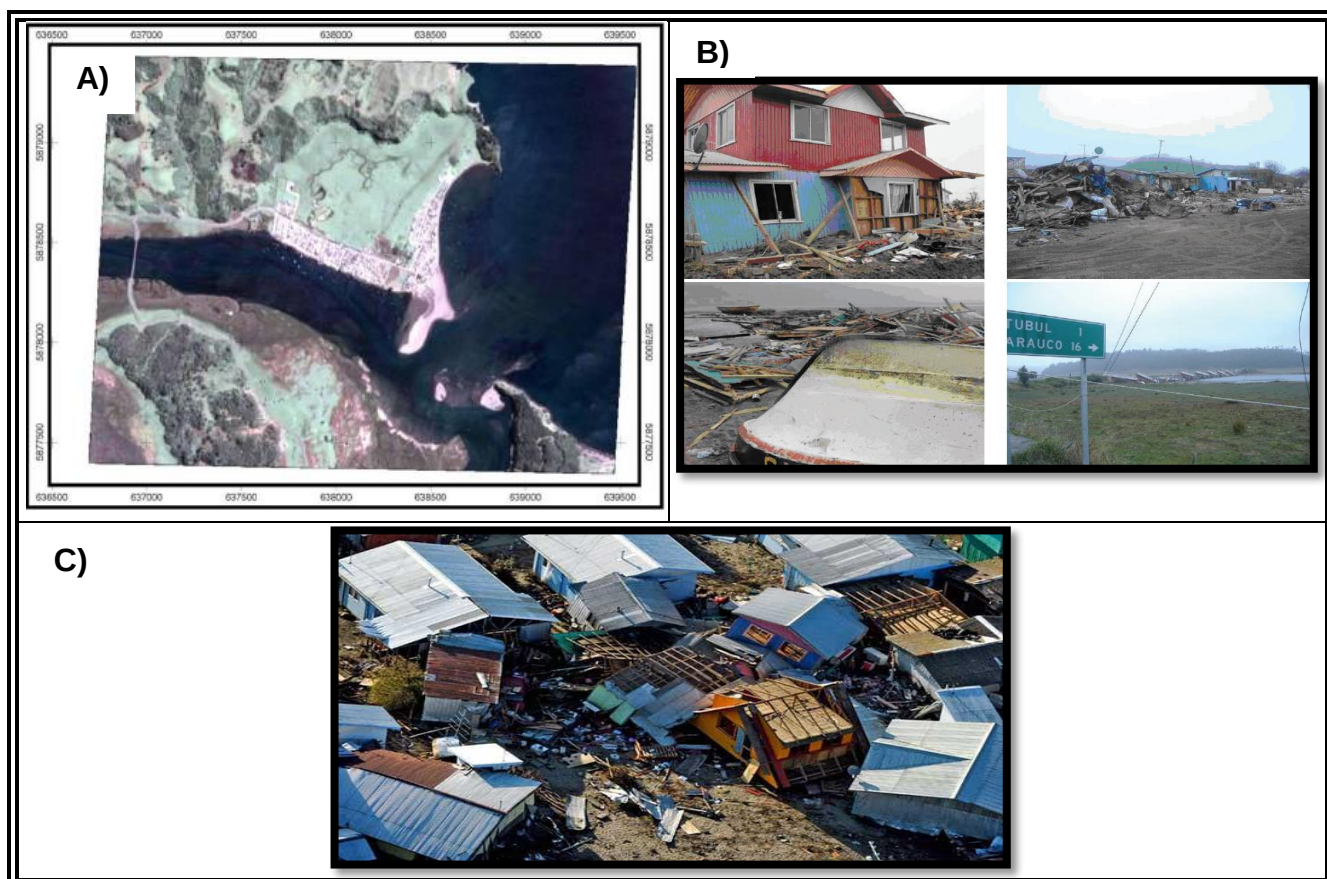


Figura 3.1.6.- Localidad de Tubul. A) Tubul ([Martínez C. et. al., 2011](#)); B) Imágenes del nivel de destrucción en caleta Tubul ([Foto: Roberto Morales](#)); C) Imagen casas arrasadas en Tubul ([Foto: Nicolas Piwonka; National Geographic](#)).

3.2- Marco teórico:

3.2.1- Teledetección e imágenes satelitales:

La teledetección se puede definir como el proceso de obtener información de un objeto analizando datos adquiridos mediante algún dispositivo que no está en contacto físico con dicho objeto (Chuvieco E., 2006). Se basa en la técnica de observación remota de la superficie terrestre y de la atmósfera que la envuelve utilizando sensores capaces de captar la radiación electromagnética, reflejada o emitida, como se ilustra en la Figura 3.2.1.1. Incluye el proceso de análisis de los datos para su posterior transformación en magnitudes físicas.

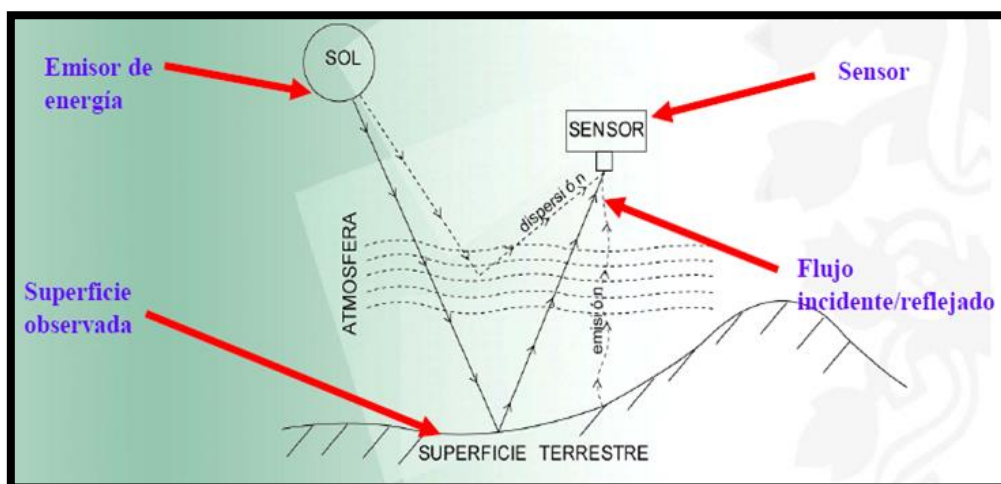


Figura 3.2.1.1- Esquema ilustrativo para el proceso en el que se basa el mecanismo de Teledetección. (Vogeler, J.C.; Cohen, W.B., 2016)

La teledetección respecto a otras formas de investigación medio ambiental nos estrega varias ventajas como: 1) Cobertura global de la superficie terrestre; 2) Observación a distintas escalas; 3) Frecuencia de adquisición; 4) Homogeneidad en la adquisición; 5) Información en las regiones del espectro no visible; 6) Formato digital. (Chuvieco E., 2006)

La teledetección es el sistema integral de captura de información territorial, a partir de la radiación electromagnética captada por el sensor. La radiación electromagnética es una de las tres formas de transmisión de energía en la naturaleza (conducción, convección y radiación) siendo la única que se transmite sin contacto material entre el emisor y receptor.

El conjunto de longitudes de onda que puede adoptar la radiación se denomina espectro electromagnético, que se muestra en la figura 3.2.1.2, graficándose en unidades de frecuencia y de longitud de onda. El espectro incluye desde las longitudes de ondas ínfimas (rayos X y Gamma), hasta aquellas usadas en telecomunicaciones, con longitudes de onda muchos mayores. Entre las distintas bandas del espectro electromagnético, sólo un restringido rango puede usarse en teledetección. Esto se debe principalmente al efecto absorbente de la atmósfera, que filtra cierto tipo de radiaciones.

Las regiones del espectro con alta transmisividad atmosférica son: 1) espectro visible (0,4-0,7 μ), dentro de la cual se distinguen tres bandas espectrales correspondientes a los tres colores básicos: azul, verde y rojo; 2) infrarrojo Próximo (0,7-1,3 μ), se usa mayormente para la observación del estado de la vegetación y de concentraciones de humedad; 3) infrarrojo lejano o térmico (10-12,5 μ), puede detectar procesos de transferencia de calor, ya que en esta banda se sitúa la máxima emitancia para las temperaturas habituales de la superficie terrestre; 4) microondas (longitudes de onda mayores de 1mm), esta banda resulta prácticamente transparente a la cubierta nubosa. (Vogeler, J.C.; Cohen, W.B., 2016)

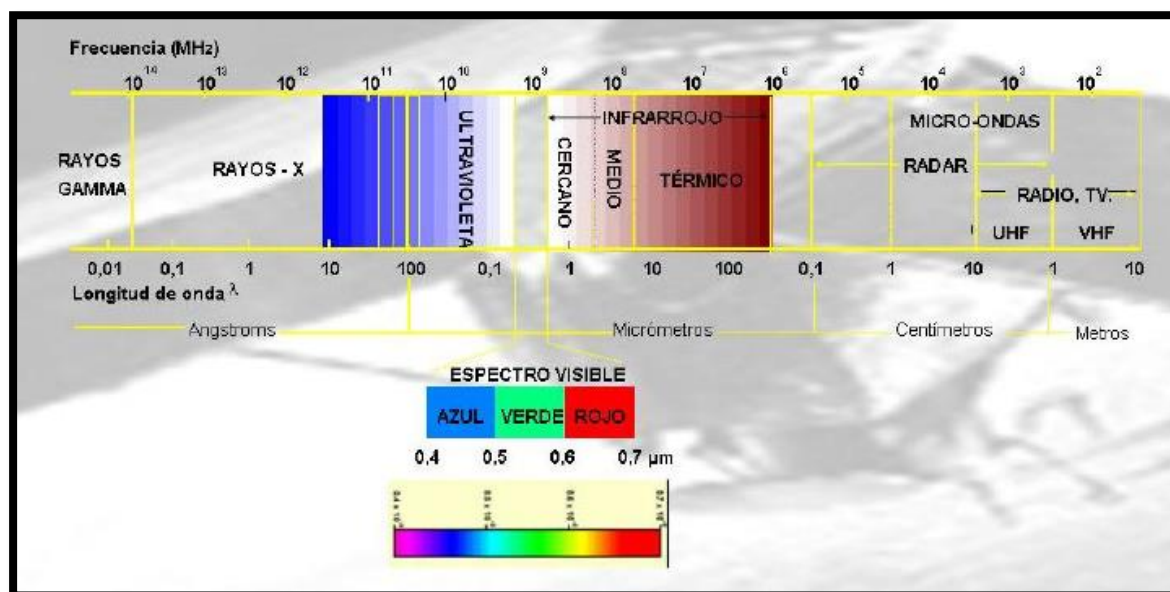


Figura 3.2.1.2- Espectro Electromagnético. (Ferrero S. et. al., 2002)

Cada elemento de la superficie terrestre ofrece una forma particular de reflejar o emitir la radiación electromagnética, dependiendo de sus propiedades físico-químicas y variables según la zona del espectro que analicemos. Esa variabilidad se manifiesta en la denominada “firma espectral” o “signatura”, propia de cada objeto y que en la Figura 3.2.1.3 se muestran para las coberturas más reconocibles.

De cara a la identificación de procesos y coberturas en la superficie terrestre, lo que nos interesa es la reflectividad de estos respecto a las diferentes longitudes de onda. La vegetación variará su firma espectral según la propia reflectividad de la hoja (estado fenológico, forma y contenido de humedad), características morfológicas de la planta (altura, concentración), entre otras. Pese a lo anterior, existen características generales: reducida reflectividad en el visible y alta en el infrarrojo cercano, reducción drástica de la reflectividad en el infrarrojo medio (debido al contenido de agua de la planta) permitiendo conocer el estado de vigor de la planta.

Los suelos tienen una curva espectral ascendente, desde el espectro visible hasta el infrarrojo y se debe a factores como la composición química, la textura, la estructura y el contenido de humedad. En general, tienen mayor reflectividad, cuanto más secos y menor contenido de materia orgánica posean. En los suelos, es determinante para las longitudes de onda larga, el contenido de humedad.

El agua tiene su mayor reflectividad en las longitudes de onda más cortas del espectro (azul y verde). De todas formas la reflectividad es baja en cualquier longitud de onda, ya que el agua absorbe y transmite la mayor parte de la radiación que recibe. En general: aguas poco profundas, tienen gran influencia de la composición de los sedimentos del fondo, aumentando la reflectividad general; aguas profundas son muy poco reflectivas.

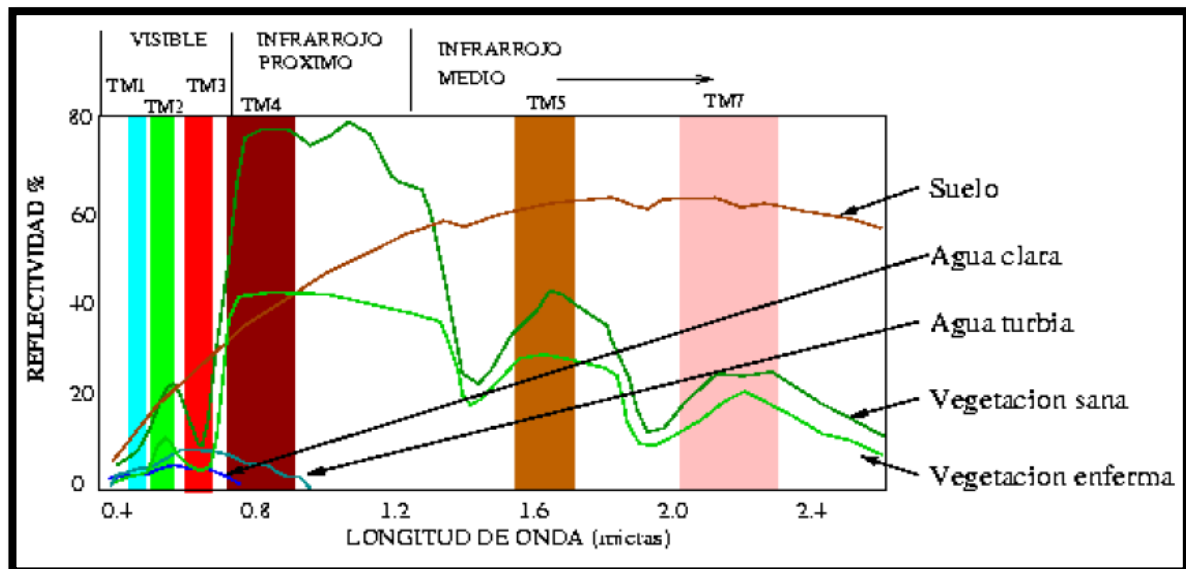


Figura 3.2.1.3- Respuestas espectrales (reflectividad) en distintos rangos de longitud de onda para diferentes coberturas de suelo típicas. (Ferrero S. et. al., 2002)

El flujo de energía recibido por el sensor no solo depende de la reflectividad de la cubierta sino también de otros factores como: condiciones atmosféricas, pendiente y orientación del terreno y geometría de la observación.

Para captar y medir la reflectividad se utilizan sensores a bordo de los satélites. Estos sensores pertenecen a un sistema óptico-electrónico, en la cual la radiancia recibida por los componentes ópticos se descomponen en varias longitudes de onda, cada una de ellas se envía a un conjunto de detectores sensibles a esa región del espectro que la amplifica y la convierte en una señal eléctrica y finalmente a un valor numérico conocido como nivel digital (ND). Estos valores numéricos pueden convertirse otra vez a valores de radiancia conociendo los coeficientes de calibrado

Tradicionalmente las imágenes satelitales, al menos cuando se representan en una sola banda, se visualizan con niveles de gris. Pero para una mejor visualización y discriminación al ojo humano se le asigna a cada ND 3 valores en una paleta de colores (rojo, verde y azul). Ya que se disponen de varias bandas y por otra parte los monitores y tarjetas de video disponen de tres canales: rojo, verde, azul (RGB). Por lo tanto se utiliza cada canal para representar el ND de una banda y así poder obtener

una composición de colores, mejorando la visualización de la imagen. En la figura 3.2.1.4., se muestra los tres canales principales, y a partir de estos se crean el resto de colores para la mejor diferenciación visual de las imágenes.

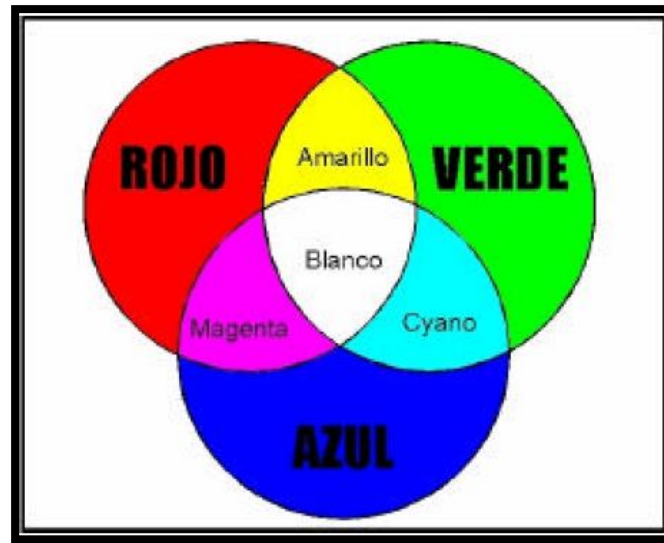


Figura 3.2.1.4- Cañones RGB- Rojo/Verde/Azul. Por combinación de ellos se construyen el resto de los colores (Rodríguez J., 2005).

La imagen captada por el sensor, tiene los siguientes componentes:

Espacial: la posición específica de cada pixel en la imagen a una resolución y un sistema de captura dados.

Radiométrica: según su codificación y estructura nos definirán un rango de niveles de grises por pixel.

Espectral: la energía total se dividirá en el momento de la captura por el sensor, en zonas del espectro electromagnético, registrándose cada zona en una banda.

Temporal: la información de cada pixel por banda y codificada en un valor de ND, corresponde a un instante dado.

Una vez definida la imagen como una matriz numérica bidimensional (valores discretos-ND-), luego deberá de realizarse los tratamientos de imágenes. Conocidas como correcciones.

3.2.2- Corrección Geométrica:

Las correcciones geométricas son necesarias debido a dos motivos: a las variaciones espaciales ocurridas en el proceso de captura e inherentes al movimiento del sensor y a la necesidad de ajustar, dicha información a un sistema de referencia determinado (sistema geodésico, proyección cartográfica, etc.). Son correcciones del tipo: rotación, cambio de escala y la traslación de los diferentes pixeles de la imagen en 3 dimensiones.

Las correcciones geométricas tienen como objetivo modificar la geometría de los pixeles, ajustándolos normalmente a un sistema geométrico dado. Un paso para la corrección geométrica es la georreferenciación, la cual se realiza usando parámetros orbitales, generalmente incluidos en la cabecera de la imagen, las cuales son: posición

exacta del satélite, altitud, velocidad del barrido, velocidad de la plataforma, meridiano de cruce con el Ecuador. Además se usan puntos de control, cálculo de la función de transformación, transferencia de ND (vecino más próximo, interpolación bilineal, convolución cubica), evaluación de ajuste (RMSE).

3.2.3- Correcciones Radiométricas:

Las correcciones radiométricas son aquellas que afectaran directamente a la radiometría de la imagen. Dichas correcciones se hacen necesarias debido a los valores (ND) erróneos registrados en la fase de captura y/o transmisión de la información. Dichos errores pueden ser debidos a perturbaciones atmosféricas que lo modifican (fenómenos de absorción y dispersión atmosférica) o a la deficiencia en la construcción del sensor. La corrección atmosférica se efectuará según modelos matemáticos que ligaran el valor registrado de ND con los valores reales existentes del terreno. Los errores producidos por el sensor se solventaran conociendo los datos de calibración radiométrica del mismo (Roy et.al., 2002). Para lograr a cabalidad la corrección radiométrica, se recurren a dos etapas claramente diferenciables:

1- Conversión a radiancia espectral en el sensor (Q_{cal} -to- L_λ) (Chander et.al., 2009): cálculo de la radiación espectral en el sensor es un paso fundamental en convertir datos de la imagen desde múltiples plataformas a una escala radiométrica común con significancia física. Calibración radiométrica envuelve un re-escalamiento de las filas de números digitales (Q) transmitidas desde el sensor a números digitales calibrados (Q_{cal}), las cuales tienen la misma escala radiométrica para todas las escenas procesadas en la tierra para un periodo específico. Durante la calibración radiométrica, valores de los pixeles (Q) desde las filas, y datos de la imagen sin procesar, son convertidos a unidades de radiancia espectral absoluta usando cálculo 32-bit floating-points. Valores de radiancia absoluta son luego escalados a 8 bits ($Q_{cal-max} = 255$). Por lo que la ecuación (3.1) de conversión a radiancia espectral en una primera etapa es:

$$L_\lambda = A * Q_{cal} + B. \quad (3.1)$$

Dónde:

L_λ es la radiancia espectral

A y B son coeficientes de calibración diferentes para cada sensor.

La conversión a radiancia espectral en el sensor requiere del conocimiento del límite superior e inferior de los factores de re-escalamiento original. La ecuación (3.2) es usada para realizar la conversión Q_{cal} -a- L_λ para productos nivel 1.

$$L_\lambda = \left(\frac{L_{max_\lambda} - L_{min_\lambda}}{Q_{cal-max} - Q_{cal-min}} \right) (Q_{cal} - Q_{cal-min}) + L_{min_\lambda} \quad (3.2)$$

L_λ = Radiancia espectral en la apertura del sensor [$W/(m^2 sr \mu m)$].

Q_{cal} = Valores de pixeles calibrados [ND].

$Q_{cal-min}$ = Valores de pixeles calibrados mínimos.

$Q_{cal-max}$ = Valores de los pixeles calibrados máximos.

L_{min_λ} = Radiancia espectral en el sensor, el cual es escalado de $Q_{cal-min}$.

L_{max_λ} = Radiancia espectral en el sensor, el cual es escalado de $Q_{cal-max}$.

2-Conversion a reflectancia al tope de la atmósfera (TOA) ([Chander et.al., 2009](#)): para ver la variabilidad escena-a-escena ésta debe ser lograda mediante la conversión de la radiancia espectral en el sensor a reflectancia en el tope de la atmósfera. Cuando se comparan imágenes desde diferentes sensores, hay 3 ventajas para ocupar TOA en lugar de la radiancia. Primero, remueve el efecto coseno de la diferencia de los ángulos solar cenital debido a la diferencia de tiempo entre la adquisición de datos. Segundo, reflectancia TOA compensa diferentes valores de la irradiancia solar exo-atmosferica surgiendo desde las diferencias de bandas espectrales. Tercero, la reflectancia TOA corrige para la variación de la distancia Tierra-sol entre diferentes adquisiciones de datos. La reflectancia TOA de la Tierra es calculada de acuerdo a:

$$\rho_\lambda = \frac{\pi \cdot L_\lambda \cdot d^2}{E_{sun_\lambda} \cdot \cos \theta_s} \quad (3.3)$$

Donde

ρ_λ = Reflectancia TOA.

d = distancia Tierra-sol. ($d = 1 - 0.0167 \cdot \cos(0.9856(dia - 4))$)

E_{sun_λ} = Irradiancia Solar [$W/(sr m^2)$]

θ_s = Ángulo de elevación solar.

3.2.4- Correcciones Topográficas:

La capacidad de proveer información detallada y confiable de las imágenes también es restringida por el efecto que las variaciones de las pendientes y ángulo de orientación del terreno en combinación con los ángulos cenital y azimutal del sol en el momento de la toma. Todo esto genera una fuerte irradiancia espectral difusa, particularmente en terrenos montañosos, en donde las superficies de las laderas orientadas directamente hacia los rayos del sol reciben más luz, que las con orientación opuesta. Esta diversificación de la geometría sol-objetivo-sensor, la cual depende de las características topográficas de la zona ([Teillet, et. al., 2001](#)), impone una variación adicional en los datos radiométricos en pixeles con cobertura de suelo y características biofísicas estructurales muy parecidas ([Soenen et. al., 2005](#)), variaciones que pueden inducir a errores en procesos posteriores disminuyendo la calidad en los resultados ([Moreira E. et. al., 2014](#)).

En el estudio se aplica corrección C ([Moreira E. et. al., 2014](#)), el cual consiste en un parámetro de corrección C para cada banda de la imagen y se incorpora a la ecuación original del método del coseno como un término aditivo de corrección topográfica. Este parámetro ejerce una influencia moderada sobre la corrección del coseno y reduce la sobre-corrección de pixeles débilmente iluminados, basado en que la iluminación se determina en función de la pendiente y ángulo de orientación del terreno, además de los ángulos de elevación y orientación del sol en el momento de toma de imagen.

Entonces partiendo del hecho de que existe una relación lineal entre los datos de iluminación y reflectancia de la imagen de la forma:

$$L = b + m(\cos i) \quad (3.4)$$

Donde

L = reflectancia.

b y m = parámetros de regresión lineal entre iluminación y reflectancia.

$\cos i = IL$ = Iluminación, definida por:

$$IL = \cos \theta_p \cdot \cos \theta_i + \sin \theta_p \cdot \sin \theta_i \cdot \cos(\theta_\alpha - \theta_0)$$

θ_p : Pendiente del terreno.

θ_i : Angulo cenital solar al momento de captura.

θ_α : Angulo azimutal solar.

θ_0 : Angulo de orientación del terreno.

Los parámetros de intersección (b) y pendiente (m) son requeridos para determinar el parámetro $C=b/m$. Este se obtienen a partir de una regresión lineal entre los valores del coseno del ángulo de incidencia solar ($\cos i = IL$) como variable independiente y la información espectral (reflectancia) como variable dependiente para cada una de las bandas de la imagen:

$$\rho_n = \frac{\rho(\cos \theta_p \cdot \cos \theta_i + C)}{(IL + C)} \quad (3.5)$$

ρ_n : Reflectancia normalizada

ρ : reflectancia original.

3.2.5- Imágenes LandSat 5 (sensor TM)

La serie de satélites LANDSAT, son satélites enviados por los Estados Unidos para el monitoreo de los recursos terrestres. La constelación LANDSAT está formado por 7 satélites que provienen, tanto conceptual como estructuralmente, de los satélites para fines meteorológicos Nimbus. Llevan a bordo diferentes instrumentos, siempre con la filosofía de captar la mayor información de la superficie terrestre, con mayor precisión y a mayor tamaño, de ahí sus mejoras radiométricas, geométricas y espaciales. Actualmente son administrados por la NASA y en tanto la producción y comercialización de las imágenes depende del Servicio Geológico de los Estados Unidos (USGS). Las imágenes LANDSAT están compuestas por 7 u 8 bandas espectrales, que fueron elegidas especialmente para el monitoreo de la vegetación, para aplicaciones geológicas y para el estudio de los recursos naturales. Estas bandas pueden combinarse produciendo una gama de imágenes de color que incrementan notablemente sus aplicaciones. Para el caso de LANDSAT 5, la altitud en que orbitan es de 705 km; el tiempo de revista es de 16 días y el periodo orbital es de 98,9 minutos.

El sensor abordo es el TM, el cual es un sensor de barrido multiespectral avanzado. Opera simultáneamente a siete bandas espectrales, siendo tres en el visible, una en el infrarrojo cercano, dos en el infrarrojo medio y una en el infrarrojo termal. Una resolución espacial de 30 metros en las bandas del visible e infrarrojo medio y 120 metros en la banda del infrarrojo termal. Información resumida en la tabla 3.2.5.1. La escena terrestre registrada por el sensor es de 185 km.

Nº Banda	Rango Espectral (µm)	Zona del espectro	Resolución en el terreno (m)
1	0.45-0.515	Visible-Azul	30
2	0.525-0.605	Visible-Verde	30
3	0.63-0.69	Visible-Rojo	30
4	0.75-0.9	Infrarrojo cercano	30
5	1.55-1.75	Infrarrojo lejano	30
6	10.4-12.5	Térmico lejano	120
7	2.09-2.35	Térmico cercano	30

Tabla 3.2.5.1- Descripción para las bandas LandSat 5 (sensor TM).

Las utilidades de las diferentes bandas en la extracción de información:

-Banda 1: (0,45-0,52µ –Azul-) Diseñada para penetración de cuerpos de agua, es útil para el mapeo de costas, para diferenciar entre suelo y vegetación y para clasificar distintos recubrimientos boscosos. También es útil para diferenciar los diferentes tipos de roca en la superficie terrestre.

-Banda 2: (0,54-0,6µ –Verde-) Especialmente diseñada para evaluar el vigor de la vegetación sana, midiendo su pico de reflectancia verde.

-Banda 3: (0,63-0,69µ) Es una banda de absorción de clorofila, muy útil para la clasificación de la cubierta vegetal.

-Banda 4: (0,76-0,9µ) Es útil para determinar el contenido de biomasa, para la delimitación de cuerpos de agua.

-Banda 5: (1,55-1,75µ) Indicativa del contenido de humedad de la vegetación y del suelo. También sirve para discriminar entre nieve y nubes.

-Banda 6: (10,4-12,5µ) el infrarrojo termal es útil en el análisis del estrés de la vegetación, en la determinación de la humedad del suelo y el mapeo termal.

-Banda 7: (2,08-2,35µ) Especialmente seleccionada por su potencial para discriminar rocas y para el mapeo hidrotermal.

3.2.6- ASTER

ASTER (The Advanced Spaceborne Thermal Emission and Reflection Radiometer) es un instrumento que presenta una órbita heliosincrónica a una distancia de 705 km, con un ciclo de repetición de 16 días, un ancho de barrido de 60 km y una distancia entre orbitas de 172 km. ASTER está compuesto por 3 subsistemas, VNIR, SWIR y TIR; cada uno de los cuales presenta características particulares tales como 3 bandas en la región espectral del visible e infrarrojo cercano (VNIR) con una resolución espacial de 15 metros; 6 bandas en la región espectral del infrarrojo de onda corta (SWIR) con una resolución espacial de 30 metros y 5 bandas en el infrarrojo térmico con una resolución espacial de 90 metros (ERSDAC, 2001). Esta información es resumida en la tabla 3.2.6.1.

ASTER también presenta un telescopio con visión hacia atrás que escanea en la región espectral de la banda 3B, lo que permite realizar modelos digitales de elevación (DEM) por pares estereoscópicos.

Sensor	Subsistema	Nº banda	Rango espectral	Resolución Espacial	Resolución Radiométrica
ASTER	VNIR	1	0,52-0,60 μ m	15 m	8 bits
		2	0,63-0,69 μ m		
		3N	0,78-0,86 μ m		
		3B	0,78-0,86 μ m		
	SWIR	4	1,60-1,70 μ m	30 m	8 bits
		5	2,145-2,185 μ m		
		6	2,185-2,225 μ m		
		7	2,233-2,285 μ m		
		8	2,295-2,365 μ m		
		9	2,360-2,430 μ m		
	TIR	10	8,125-8,475 μ m	90 m	12 bits
		11	8,475-8,825 μ m		
		12	8,925-9,275 μ m		
		13	10,25-10,95 μ m		
		14	10,95-11,65 μ m		

Tabla 3.2.6.1- Descripción para las bandas ASTER.

Las imágenes a ocupar son productos L1T, los cuales se encuentran en formato de niveles digitales enteros (ND) con una resolución radiométrica de 16 bits. Estos se pueden convertir a valores de reflectancia TOA con factores de escala previstos en el archivo de metadatos incluidos en el producto. Los datos de una imagen L1T se encuentran radiométrica y geoméricamente corregidos y están disponibles en formato geoTIFF.

3.2.7- Combinación de las bandas

Al combinar distintas bandas de una imagen satelital multi-espectral para obtener una imagen RGB, obtenemos distintos resultados que pueden ayudar a determinar mejor los objetivos que tratamos de identificar. Estas combinaciones de bandas se obtienen de la unión de tres bandas como RGB. Con distintas bandas para cada componente de la combinación obtenemos distintos resultados, como el color natural (432).

3.2.8- Clasificación de la imagen:

Luego de realizar las correcciones de la imagen satelital, se hace necesario vincular cada valor registrado en los píxeles a una clase temática dada por el usuario, entendiéndose por clase temática a las categorías de coberturas de suelo en las que se quiere clasificar o dividir los datos de la imagen.

Para este caso se usará un tipo de clasificación supervisada. En este método se definen a priori las clases de información que existen en la imagen (no dónde se encuentran). Para ello se seleccionarán unas muestras significativas en la imagen (muestras de entrenamiento) y en terreno, con el objetivo de la interpretación real de dichas muestras, etiquetándose posteriormente todos los píxeles con criterios estadísticos, finalmente se agruparan formando las diversas clases definitivas. Para esto, las fases a seguir son las siguientes: 1-definición de clases finales a obtener. 2- Selección de muestras significativas en el terreno y su correspondiente identificación en la imagen para cada clase. 3-Definición del clasificador en función de dichos valores muestrales. Para este trabajo se usaran los algoritmos SVM y Random Forest y se compararán con el método de clasificación más usado, como lo es el método de máxima verosimilitud. 4-Obtención de la imagen raster clasificada.

3.2.8.1- Máxima Verosimilitud (MV)

El método de clasificación por máxima verosimilitud (MV) o máxima probabilidad ([Chuvieco E., 2006](#); [Tso B. y Mather P., 2009](#)) parte del supuesto de que los valores de reflectividad en cada una de las clases definidas por el usuario, siguen una distribución de probabilidad normal multivariante. Se utiliza entonces el vector de medias y matrices de varianza-covarianza para estimar la probabilidad de que un píxel dado pertenezca a cada una de las clases. El píxel se asignará finalmente a la clase cuya probabilidad de pertenencia sea mayor.

Una vez terminada la asignación de píxeles a las distintas clases, se pueden establecer unos umbrales de probabilidad para cada categoría, rechazando los píxeles con una probabilidad muy baja, evitando de esta manera que se clasifiquen píxeles en clases en las que realmente no pertenecen. De esta manera tendremos un número de píxeles sin clasificar, y en el caso de ser muy grande se debe retornar la fase de entrenamiento para deducir así nuevas categorías o perfeccionar las ya existentes. En todo caso el valor de probabilidad máxima puede utilizarse como indicador de la certeza en la clasificación.

Frecuentemente se utiliza un enfoque bayesiano y se incluye también la probabilidad a priori de que un píxel pertenezca a una clase determinada, debido a que algunas categorías cuentan con una superficie mayor, y por lo tanto la probabilidad de que una celdilla pertenezca a ella es, a priori, mayor. El método de Máxima Verosimilitud es el más utilizado en teledetección ([Yonezawa C., 2007](#)). Sin embargo, el supuesto básico de normalidad no siempre se cumple, por lo que debería verificarse antes de aplicarlo. No obstante, algunos autores han planteado que, aun en el caso de que las bandas de entrada no cumplan este criterio, los resultados siguen siendo bastantes consistentes ([Swain P. y Davis S., 1978](#)).

Según [Richards \(1999\)](#) la función que se utiliza para la clasificación de píxeles en una de las clases en el método de máxima verosimilitud está dada por la ecuación (3.6).

$$g_i(x) = \ln p(w_i) - \frac{1}{2} x \ln[\Sigma_i] - \frac{1}{2} x(x - m_i)^t \Sigma_i^{-1}(x - m_i) \quad (3.6)$$

Donde:

$g_i(x)$ es la probabilidad de que un píxel pertenezca a una clase determinada.

i = clases.

x = N° de bandas.

$p(w_i)$ = probabilidad de que clase w_i ocurra en la imagen.

$|\Sigma_i|$ = Determinante de la matriz covarianza en los datos de la clase w_i .

Σ_i^{-1} = Matriz inversa.

m_i = Vector promedio.

3.2.8.2- SVM

Algoritmos de aprendizaje que usan vectores de soporte, como SVM, son basados en la simple idea originada en la teoría de aprendizaje estadístico ([Vapnick V., 1995](#)). La simplicidad proviene del hecho que las máquinas de vector de soporte aplican un simple método lineal a los datos, pero en un espacio característico de alta dimensionalidad, no linealmente relacionado al espacio de entrada. En este nuevo espacio de alta dimensionalidad, SVM busca el hiperplano que separe y maximice el margen entre las clases en este espacio. Además a la hora de definir el hiperplano se consideran solo los ejemplos de entrenamiento de cada clase que caen justo en las fronteras de dicho margen, siendo estos los llamados vectores de soporte, como se observan en la Figura 3.2.8.2.1.

En clasificación, SVMs separa las diferentes clases de datos por medio de un hiperplano:

$$\langle w, \Phi(x) \rangle + b = 0 \quad (3.7)$$

Correspondiendo a la función de decisión

$$f(x) = \text{sign}(\langle w, \Phi(x) \rangle + b) \quad (3.8)$$

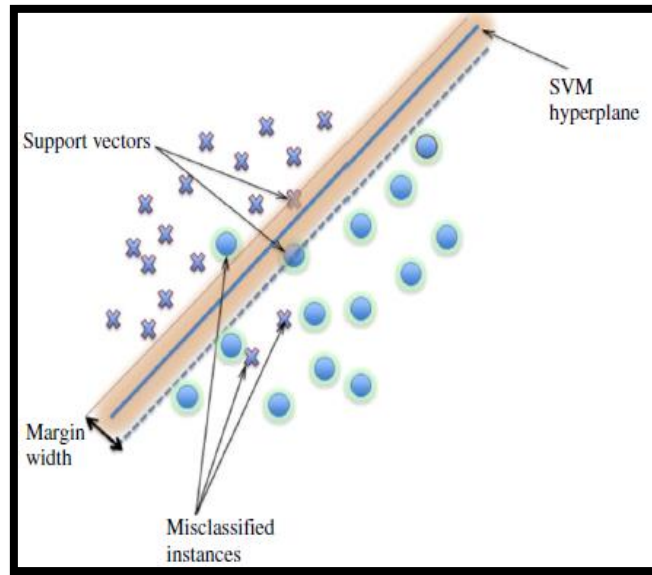


Figura 3.2.8.2.1- Maquinas de vectores de soporte lineal.
(Adaptado desde Burges B. et. al., 1999).

Esto puede ser mostrado como el hiperplano óptimo (Vapnik, 1995), el cual es el margen máximo de separación entre dos clases, como se ilustra en la figura 3.2.8.2.1., donde trata de separar dos clases de la manera más lineal posible.

Para entender la teoría que está detrás, supongamos que se tiene un conjunto $\mathbf{S}$ de puntos etiquetados para entrenamiento. Cada punto de entrenamiento $\mathbf{x}_i \in \mathbb{R}^n$ pertenece a alguna clase y se le asigna una etiqueta y_i , ($i = 1, \dots, l$).

Se mapea el espacio de entrada en un espacio de características de una dimensión mayor y buscar el hiperplano óptimo allí. Sea $\mathbf{z} = \Phi(\mathbf{x})$ la notación del correspondiente vector en el espacio de características, o sea en el espacio de mayor dimensionalidad, con un mapeo Φ de $\mathbb{R}^n$ a un espacio de características $\mathbf{Z}$.

Se desea encontrar el hiperplano $\mathbf{w} \cdot \mathbf{z} + b = 0$, definido por el par $(\mathbf{w}, b)$, tal que podamos separar el punto $\mathbf{x}_i$, suponiendo que estos satisfacen:

$$\begin{aligned} \mathbf{w} \cdot \mathbf{z}_i + b &\geq 1 & \text{para } y_i = 1 \\ \mathbf{w} \cdot \mathbf{z}_i + b &\leq -1 & \text{para } y_i = -1 \end{aligned} \quad (3.9)$$

O sea que los datos de entrenamiento estén sobre o por debajo de los límites de los hiperplanos, lo que se puede denotar como:

$$y_i(\mathbf{w} \cdot \mathbf{z}_i + b) - 1 \geq 0 \quad (3.10)$$

Si el conjunto $\mathbf{S}$ no es linealmente separable, como el mostrado en la figura 3.2.8.2.1, el análisis previo debe ser generalizado introduciendo algunas variables no negativas $\epsilon_i \geq 0$, donde los datos correctamente clasificados $\epsilon_i = 0$, quedando la ecuación (3.10) definida como:

$$y_i(\mathbf{w} \cdot \mathbf{z}_i + b) \geq 1 - \epsilon_i, \quad i = 1, \dots, l \quad (3.11)$$

($l = \text{n}^\circ \text{ de muestras de entrenamiento}$)

Consideremos los puntos en los que se cumple la igualdad en la ecuación (3.11). Estos definen dos hiperplanos, donde el margen entre los dos se puede demostrar que es $\frac{2}{\|w\|}$

Entonces se puede encontrar el par de hiperplanos que den el máximo margen minimizando la siguiente ecuación:

$$\begin{aligned} &\text{Minimizar} && \left\{ \frac{1}{2} \cdot \|w\|^2 + C \sum_{i=1}^l \varepsilon_i \right\} \\ &\text{Sujeto a} && y_i(w \cdot z_i + b) \geq 1 - \varepsilon_i, \quad i = 1, \dots, l \\ &&& \varepsilon_i \geq 0 \end{aligned} \quad (3.12)$$

El primer término hace referencia al inverso del margen y el segundo al error de clasificación, donde el parámetro costo C controla el costo pagado por el SVM para clasificaciones erróneas de un punto de entrenamiento. El ajuste de este parámetro puede hacer un balance entre la maximización del margen y la violación de la clasificación (Weingessel A., 2004). $\sum_{i=1}^l \varepsilon_i$ es una cota superior sobre el número de errores en el entrenamiento.

Posteriormente se pasa a una formulación Langrangiana del problema para que las restricciones del problema sean sustituidas por restricciones sobre los multiplicadores de lagrange, que son más fáciles de manejar y para que los datos de entrenamiento solo aparezcan en forma de producto escalares entre vectores.

Entonces la solución w puede ser mostrada como una expansión:

$$w = \sum_{i=1}^l \alpha_i y_i z_i \quad (3.13)$$

Donde α_i denota los multiplicadores de lagrange.

Obteniendo el siguiente lagrangiano:

$$L_p = \frac{1}{2} \cdot \|w\|^2 - \sum_{i=1}^l \alpha_i y_i (w \cdot z_i + b) + \sum_{i=1}^l \alpha_i \quad (3.14)$$

Ahora esto se resuelve al maximizar L_p sujeto a las restricciones de que el gradiente de L_p se anulen con respecto a w y b , y $\alpha_i \geq 0$.

$$\frac{\partial L_p}{\partial w} = 0 \rightarrow \sum_{i=1}^l \alpha_i y_i z_i$$

$$\frac{\partial L_p}{\partial b} = 0 \rightarrow \sum_{i=1}^l \alpha_i y_i = 0$$

Sustituyendo las restricciones en el lagrangiano (3.14), se obtiene que el problema de buscar el hiperplano óptimo en (3.12), queda como un problema cuadrático para buscar los α_i :

$$\begin{aligned} &\text{Maximizar} && w(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j z_i z_j \\ &\text{Sujeto a} && \sum_{i=1}^l y_i \alpha_i = 0 \\ &&& 0 \leq \alpha_i \leq C \end{aligned} \quad (3.15)$$

Por el teorema de Khun-Tucker, la solución $\bar{\alpha}_i$ satisface:

$$\begin{aligned} &\bar{\alpha}_i (y_i (\bar{w} \cdot z_i + b) - 1 + \bar{\varepsilon}_i) = 0 \\ &(C - \bar{\alpha}_i) \varepsilon_i = 0 \end{aligned} \quad (3.16)$$

Donde los coeficientes α_i están acotados superiormente por el parámetro C (costo). Notese que hay un multiplicador de Lagrange para cada muestra de entrenamiento. Tras obtener una solución aquellos puntos para los que $\alpha_i \geq 0$ se les denomina vectores de soporte y yacen sobre el hiperplano.

La función decisión generalizada de (3.12) y (3.15)

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i z_i z + b\right) \quad (3.17)$$

La extensión a un problema multiclase, se utiliza el enfoque propuesto por Cramer y Signer (2000). Este algoritmo trabaja mediante la resolución del problema de optimización individual incluyendo los datos desde todas las clases.

Por otra parte, se tiene que SVM usa un mapeo implícito Φ de los datos de entrada en un espacio característico de alta dimensionalidad, definida por una función núcleo, es decir una función que trabaja solo el producto interno $\langle \Phi(x), \Phi(x') \rangle$, entre las imágenes de dos puntos de datos x, x' en el espacio de características. El proceso de aprendizaje luego toma lugar en el espacio de características, y los puntos de datos solo aparecen dentro del producto punto con otros puntos.

Como no se tiene conocimiento de Φ , el cálculo de (3.14) y (3.16) es imposible, por eso se usa la función $K(\cdot, \cdot)$, llamado núcleo que calcula el producto punto entre los puntos de entrada en el espacio de características Z , esto es:

$$z_i \cdot z_j = \Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j) \quad (3.18)$$

El cual es computacionalmente más simple que explícitamente proyectar x_i, x_j en el espacio característico Z . Una vez seleccionada una función núcleo válida, se puede trabajar en espacios de cualquier dimensión sin un costo computacional adicional. Por lo que la solución al hiperplano óptimo queda dada por:

$$\text{Maximizar } w(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.19)$$

$$\text{Sujeto a } \sum_{i=1}^l y_i \alpha_i = 0, \quad 0 \leq \alpha_i \leq c$$

Y la función de decisión es:

$$f(x) = \text{sign}(w \cdot z + b) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x_j) + b\right) \quad (3.20)$$

El problema puede ser fácilmente resuelto usando un solucionador estándar QP como quadprog() existente en el paquete quadprog (Weingessel A., 2004) o ipop() en el paquete kernlab (Karatzoglou y Smola, 2005) ambos disponibles en R (R Development Core Team 2005).

3.2.8.3- Random Forest:

Algoritmos de aprendizaje conjunto (Random Forest, bagging y boosting) tienen un creciente interés porque demuestran una precisión y robustez a no ser influenciado mayormente por datos anómalos, comparado con otros clasificadores. La filosofía detrás de los clasificadores de conjunto, o sea de RF, es que se basa sobre la premisa básica que un conjunto de clasificadores realizan una mejor clasificación que de manera individual. Breiman L. (2001) sugiere el algoritmo RF, el cual presenta varias ventajas para aplicaciones percepción remota: 1) trabaja eficientemente para largas bases de datos; 2) puede manejar miles de variables de entrada sin eliminación de variables; 3) da estimaciones de que variables son importantes en la clasificación. 4) genera una estimación interna imparcial de la generalización de error; 5) es computacionalmente más ligero que otros métodos de conjunto de árboles (boosting).

RF es una combinación de clasificadores donde cada clasificador contribuye con un voto para la asignación de las clases más frecuente al vector de entrada (x),

$$\widehat{C}_{rf}^B = \text{majority vote}\{\widehat{C}_b(x)\}_1^B$$

Donde $\widehat{C}_b(x)$ es la predicción de clase de los b árboles.

El proceso de clasificación es básicamente el que se muestra en la figura 3.2.1.1.

Un RF aumenta la diversidad de los árboles haciendo que crezcan de diferentes subconjuntos de datos de entrenamiento creado mediante “*bagging*” o “*bootstrap aggregating*” (Breiman L., 1996). “*Bootstrap aggregating*” (empaquetamiento) es una técnica para la creación de datos de entrenamiento mediante el remuestreo aleatorio del conjunto de datos original con reemplazo, es decir sin la eliminación de los datos seleccionados desde la muestra de entrada para la generación del siguiente subconjunto. RF usa “árboles” como clasificadores bases $\{h(x, \theta_k), k = 1, \dots, \}$, donde x es el vector de entrada y $\{\theta_k\}$ son las muestras de entrenamiento aleatoria e independientemente distribuidas (Breiman L., 2001; Hastie et al., 2009).

Por esta razón algunos datos pueden ser usados más que otros en los clasificadores de entrenamiento, mientras que otros pueden no ser usados. Así, se logra una mayor estabilidad del clasificador, lo hace más robusto cuando se enfrentan a pequeñas variaciones en los datos de entrada y, al mismo tiempo, aumenta la precisión de la clasificación (Breiman L., 2001). Cada subconjunto seleccionado usando el proceso de empaquetamiento para hacer cada árbol individual, generalmente contiene 2/3 del conjunto de datos de calibración (ver esquema en figura 3.2.1.1).

Las muestras que no están presentes en el subconjunto de calibración se incluyen como parte de otro subconjunto llamado out-of-bag (oob). Se debe tener en cuenta que un diferente subconjunto oob se forma para cada árbol, a partir de los elementos no seleccionados por el proceso de empaquetamiento. Estos elementos oob, que son no considerados para el entrenamiento del árbol, se pueden clasificar por el árbol para evaluar el rendimiento.

Al final de la ejecución, cada elemento de entrada del subconjunto oob ha sido clasificado en promedio por un tercio del número total de árboles generados en el conjunto (Peters et al., 2007). La proporción entre las clasificaciones erróneas y el número total de elementos oob aporta una estimación no sesgada del error de generalización (Breiman L., 2001). El error de generalización converge a medida que el número de árboles aumenta, por lo tanto, los RF no sobre-ajustan los datos.

Además, cuando el RF hace crecer un árbol, utiliza la mejor división de un subconjunto aleatorio de características de entrada o variables predictivas en la división de cada nodo, en lugar de utilizar las mejores variables de división. Por lo tanto, esto puede disminuir la fuerza de cada árbol, pero reduce la correlación entre los árboles, lo que reduce el error de generalización (Breiman L., 2001).

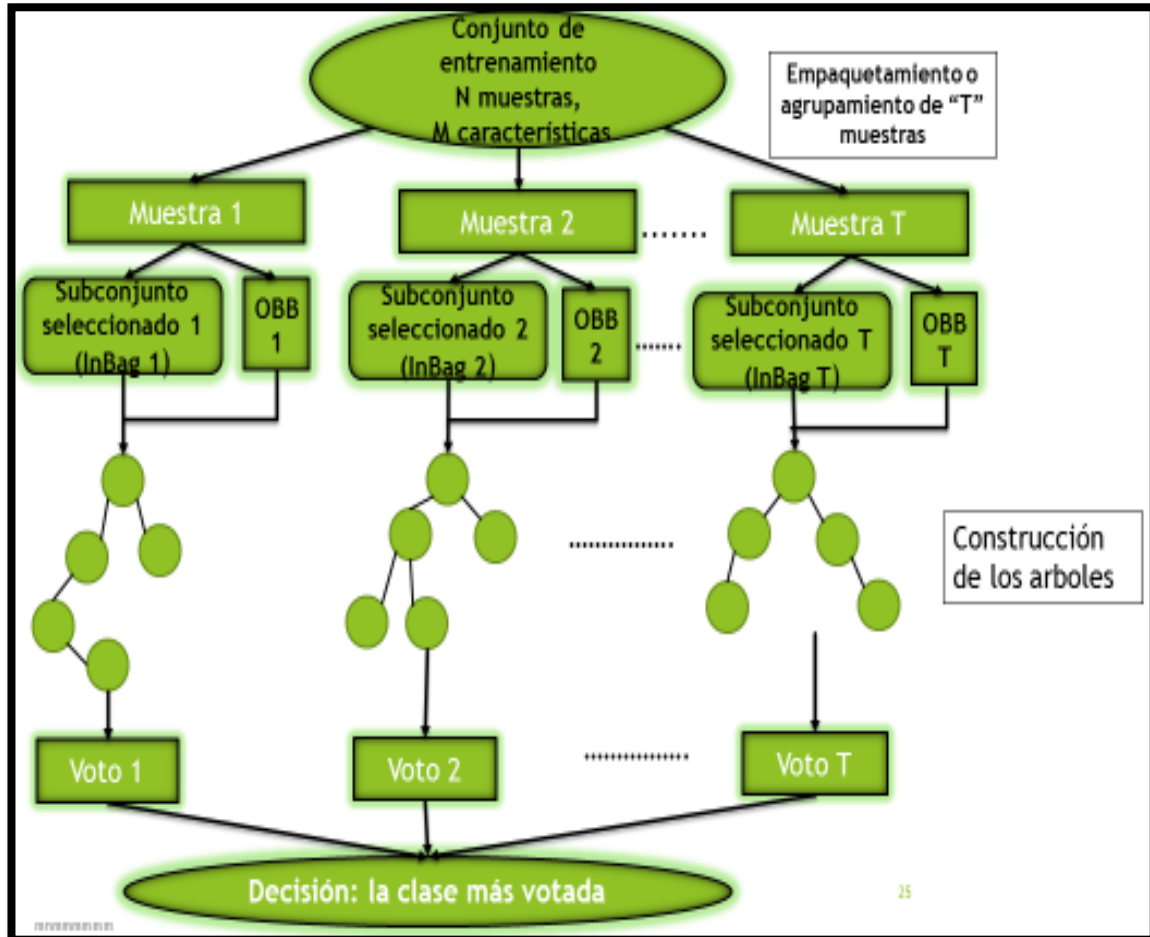


Figura 3.2.8.3.1- Esquema del procedimiento de un clasificador Random Forest
(Crédito: Elaboración propia)

El diseño del árbol requiere la elección de una medida de selección de atributos adecuada que maximice la disimilitud entre las clases. Existen muchas aproximaciones para seleccionar los atributos que se pueden utilizar para la inducción en árboles de decisión. Algunos de los más frecuentes son la relación de ganancia (Quinlan J., 1993), el índice de Gini (Breiman L., 2001) y Chi-cuadrado (Mingers L., 1989b). Un RF utiliza generalmente el índice de Gini como una medida para la mejor selección de división, midiendo la impureza de un elemento dado con respecto al resto de las clases.

Para un conjunto de datos de formación T dado, el índice de Gini puede expresarse como:

$$\sum \sum_{j \neq i} \left(\frac{f(C_i, T)}{|T|} \right) \left(\frac{f(C_j, T)}{|T|} \right) \quad (3.21)$$

Donde $f(C_i, T)/|T|$ es la probabilidad de que un caso seleccionado pertenezca a la clase C_i .

RF también proporciona una evaluación de la importancia relativa de las diferentes características o variables durante el proceso de clasificación. Este aspecto es útil para estudios multi-fuente, donde la dimensionalidad de los datos es muy alta, y es importante saber cómo influye cada variable predictiva en el modelo de clasificación para poder seleccionar las mejores variables ([Gislason et al., 2004](#); [Pal, 2005](#); [Ghimire et al., 2010](#)). Para evaluar la importancia de cada característica (por ejemplo, la banda de imágenes de satélite), el RF conmuta una de las variables aleatorias de entrada manteniendo constante el resto y mide la disminución de precisión que se ha producido mediante la estimación de error oob y la disminución del índice Gini ([Breiman L, 2001](#)).

El clasificador de RF sólo necesita la definición de dos parámetros para generar un modelo de predicción: el número de árboles de clasificación deseados (k) y el número de variables de predicción, (m), utilizados en cada nodo para hacer crecer el árbol. En otras palabras, para clasificar un nuevo conjunto de datos se utiliza un número constante de m variables predictivas aleatorias, y cada uno de los ejemplos del conjunto de datos se clasifica por un k número de árboles definido por el usuario. De esta manera el valor final de la clase asignada a cada ejemplo será igual al valor más frecuente para el número total de k árboles generados.

[Breiman \(2001\)](#) sugirió que al aumentar el número de árboles el error de generalización siempre converge y el sobre-entrenamiento no es un problema debido a la "Ley Fuerte de Grandes Números" ([Feller W., 1968](#)). Por otra parte, la reducción del número de variables predictivas (m) hace que cada árbol individual del modelo sea menos fuerte, pero también reduce la correlación entre los árboles, lo que aumenta la precisión del modelo. Teniendo esto en cuenta, es necesario optimizar los parámetros k y m para minimizar el error de generalización.

En resumen el esquema del algoritmo Random Forest es:

- 1-Aleatoriamente se crea (seleccionando con reemplazo) el conjunto de entrenamiento de igual tamaño que el conjunto original. Al seleccionarse aleatoriamente con reemplazo no todos los datos del conjunto general estarán en el conjunto de entrenamiento.
- 2-Los datos que no forman parte del conjunto de entrenamiento forman el conjunto de validación o out of bag data (datos OOB).
- 3-En cada punto de división del árbol o nodo, la búsqueda de la mejor variable para dividir los datos no se realiza sobre todas las variables sino sobre un conjunto, m , de las mismas. La elección del subconjunto de variables se realiza de forma aleatoria.
- 4-Se busca la mejor división de los datos de entrenamiento teniendo en cuenta solo la m variable seleccionada. Para esta tarea se debe implementar una función objetivo, la cual busca la mejor forma de agrupar los datos según las probabilidades de pertenencia de cada dato. Habitualmente ésta es la entropía o el índice de Gini.
- 5-Los anteriores procesos son repetidos varias veces, de forma que se tienen un conjunto de árboles de decisión entrenados sobre diferentes subconjuntos de datos y de atributos.

6-Una vez el algoritmo entrenado, la evaluación de cada nueva entrada es realizado con el conjunto de árboles. La categoría final de la clase (clasificación) es realizada por el voto mayoritario del conjunto de árboles.

3.2.9- Validación de los algoritmos: matriz de confusión e índice Kappa

El análisis de datos de confiabilidad se hace generalmente a través de una matriz de confusión, que permite confrontar la información de los sitios de verificación con aquellos de la base cartográfica que se pretende evaluar (ver ejemplo en la tabla 3.2.9.1.). En la matriz de confusión las filas representan las clases de referencia y las columnas las clases del mapa temático. La diagonal de la matriz expresa el número de sitios de verificación para los cuales hay concordancia entre el mapa y los datos de referencia, mientras que los marginales expresan errores de asignación. La proporción de puntos correctamente asignados (diagonal) expresa la confiabilidad de la clasificación. Pueden existir dos tipos de errores: el error de comisión, el que representa la proporción de sitios de verificación dada en una cierta clase, pero que en realidad pertenece a otra categoría. El error de omisión, el que se refiere a la proporción de sitios de verificación correspondientes a una categoría que fue clasificada en otra (Chuvieco E., 2006).

	MAPA				
REFERENCIA	1	2		q	TOTAL
1	p_{11}	p_{12}		p_{1q}	p_{1+}
2	p_{21}	p_{22}		p_{2q}	p_{2+}
.....					
Q	p_{q1}	p_{q2}		p_{qq}	p_{q+}
TOTAL	p_{+1}	p_{+2}		p_{+q}	

Tabla 3.2.9.1- Ejemplo Matriz de Confusión (Se expresan los valores de la matriz en proporción del número total de sitios).

Los índices de confiabilidad generalmente no toman en cuenta los elementos fuera de la diagonal de la matriz, por esta razón se generalizo el uso del coeficiente kappa, que utiliza las sumas marginales de la matriz y da cuenta de la contribución del azar en la confiabilidad de la clasificación.

$$K = \frac{P_0 - P_c}{1 - P_c} \quad (3.22)$$

Donde K es el índice Kappa. P_0 es la proporción del área correctamente clasificada (confiabilidad global) y P_c la confiabilidad resultante del azar.

P_0 se obtiene sumando los elementos de la diagonal:

$$P_0 = \sum_{k=1}^q p_{kk} \quad (3.23)$$

P_c se calcula sumando el producto de las sumas marginales:

$$P_c = \sum_{k=1}^q p_{k+} \cdot p_{+k} \quad (3.24)$$

3.3- Materiales y metodología:

Los datos utilizados corresponden a imágenes satelitales multi-espectrales y modelos digitales de elevación (DEM) obtenidos de sensores remotos. Las imágenes satelitales utilizadas fueron obtenidas por el satélite LandSat 5 y ASTER.

Las imágenes LandSat corresponden a 2 imágenes orto-rectificadas LandSat TM 07/03/2009 (path 001/row 085; path 001/row 086) y 3 imágenes orto-rectificadas LandSat TM 2010 (path 001/row 086-21/01/2010; path 001/row 085-26/03/2010; path 001/row 086- 26/03/2010). Las descargas de estas imágenes se pueden realizar a través de estos sitios web:

<http://earthexplorer.usgs.gov/>

<http://glovis.usgs.gov/>

Las imágenes ASTER para el área de estudio fueron obtenidas para las fechas 27/02/2009 y 11/03/2010. Estas imágenes contienen 14 bandas agrupadas en 3 regiones espectrales, pero sólo se usarán las que se encuentran en la región del visible e infrarrojo cercano (VNIR; bandas 1-3) con 15m de resolución, ya que VNIR son las que contienen una mejor información espectral para distinguir tipos de cobertura de suelo. La descarga de estas imágenes se obtuvo del sitio web:

<https://search.earthdata.nasa.gov/>

Estas imágenes fueron seleccionadas para comprender dos periodos de tiempo distintos, el antes y después del terremoto y tsunami del 27 de febrero de 2010 para el área de estudio y así poder visualizar el cambio en la cobertura de suelo. Además la selección de estas imágenes se debió a que no tienen presencia de nubes en el área de estudio. Las imágenes poseen un nivel de procesamiento “level 1T – Terrain Correction”.

En la tabla 3.3.1 se resume la descripción de las imágenes satelitales a utilizar.

SATELITE	AÑO	DD/MM/YYYY	DESCRIPCIÓN	BANDAS
LANDSAT	2009	07/03/2009	path001/row085	1-5,7
			path001/row086	
	2010	21/01/2010	path001/row086	
		26/03/2010	path001/row085	
			path001/row085	
ASTER	2009	27/02/2009	AST_L1T_0030227	1-3 (VNIR)
	2010	11/03/2010	AST_L1T_0030311	

Tabla 3.3.1- Tabla descriptiva de las imágenes satelitales utilizadas.

El modelo de elevación digital (DEM), puede ser definido como una representación matricial regular de la variación continua del relieve en el espacio (Burrough P., 1998). Se usaron 2 DEM SRTM (Shuttle Radar Topography Mission) de libre disposición (<https://search.earthdata.nasa.gov/>), con una resolución espacial de 30 metros correspondientes al área de estudio (s37_w73_1arc_V3; s38_w74_1arc_V3).

En el procesamiento se realizan 3 pasos fundamentales: Pre-procesamiento, Clasificación y Validación.

Antes de ser utilizadas e interpretadas, las imágenes a clasificar deben pasar por algunos procesos que son realizados para darle un sentido físico a los valores y corregir ciertos errores que se producen y son inherentes al sistema de clasificación. La energía captada por el sensor se almacena como niveles digitales (ND). La conversión de ND a variables físicas es un paso importante en la fase de interpretación de las imágenes, ya que permite trabajar con variables físicas de significado estándar comparables entre un mismo sensor a lo largo del tiempo y entre distintos sensores (Chuvieco, 2006). Por lo anterior es que las imágenes deben ser calibradas y convertidas a reflectancia. Este proceso es realizado para todas las imágenes, junto con las correcciones radiométricas y topográficas, de esa forma tenemos los valores de reflectancia para cada banda en un rango de entre 0 y 1. Para realizar estos procesos se utilizó el software libre y de código abierto QGIS v2.8.1 (Quantum Gis Development Team, 2015), el cual incorpora un complemento denominado “Semi-automatic Classification Plugin” (Congedo L., 2013), que contiene herramientas para ejecutar de manera semi-automática el pre-procesamiento, para las correcciones radiométricas. La información de las imágenes se encuentran en un archivo de cabecera de la imagen y el resto se deriva de ellos. Los datos que debemos localizar y consignar son los relativos al ángulo de elevación solar, la radiancia (máxima y mínima) en cada banda y el factor corrector de la distancia al sol. Los dos primeros datos aparecen explicitados en el archivo de cabecera.

3.3.1- Lectura de datos

Cargamos las bandas correspondientes a las imágenes a utilizar (LandSat (6 bandas por imagen); ASTER (3 bandas por imagen)). Para los procesamiento a realizarse posteriormente se debe asegurar de haber tratado previamente los datos nulos, los cuales corresponden a valores de píxeles que no se encuentran dentro de la escala de la imagen (16-bits), es decir entre los valores de ND 0 a 256.

Las imágenes tienen un formato en el que se encuentra un rombo dentro de un cuadrado, siendo el cuadrado el que contiene la imagen y las zonas del rombo que no son cubiertas por el cuadrado corresponden a píxeles con datos erróneos, los cuales no nos entregan ninguna información e induce a errores, por lo que se le asigna el valor -999.99 y se marcan como valores nulos, dejándolos fuera de todos los procesos que se realicen sobre la imagen.

También es importante verificar la proyección de las imágenes, las que deben estar en “EPSG: 32718-WGS84/UTM18S”. Esto se puede visualizar en la información general de la imagen, dando clic en las propiedades de ésta. Todas nuestras imágenes se encuentran en “EPSG: 32718-WGS84/UTM18N”, por lo que es necesario re-proyectarlas, proceso que es realizado en QGIS mediante la herramienta “WARP (reproject)” de la librería “GDAL”. Además se recorta la imagen para que su dimensión corresponda mayormente a nuestra zona de interés y así se evitará trabajar con una cantidad innecesaria de píxeles, lo cual acarrearía un mayor gasto computacional en todos los procedimientos a realizarse. Para esto se utiliza la herramienta de QGIS “CLIP RASTER BY MASK LAYER” de la librería “GDAL/OGR”, en el cual se pide una máscara o capa vectorial (shape) con las dimensiones del área de interés.

Con todo lo anterior realizamos la etapa de pre-procesamiento, la cual esta resumida en el esquema de la figura 3.3.2.1

3.3.2- Correcciones

Como ya se mencionó, nuestras imágenes satelitales son un producto L1T, consistiendo en una serie cuantificada, calibrada y escalada de niveles digitales (ND), los cuales representan los datos de una imagen multi-espectral. Los datos de las bandas están derivados en 16 bits en formato no cifrado y pueden ser re-escalados a valores de reflectancia TOA empleando los coeficientes radiométricos provisto en el archivo de metadatos 'MTL.txt' para las imágenes LandSat o contenidas dentro de la imagen '.hdf' para ASTER. Entonces usando el complemento de QGIS "Semi-automatic Classification Plugin" se logra la conversión a reflectancia TOA con las correcciones radiométricas incorporadas para cada una de las bandas.

Para realizar las correcciones topográficas se usa la herramienta "SAGA (2.1.2)", donde se utilizan las opciones "Terrain Analysis-Lighting" → "Topographic Correction", en la cual se debe ingresar cada una de las bandas por separado y el algoritmo pide como entrada un modelo de elevación digital con la misma extensión, resolución y proyección que las bandas de las imágenes a utilizar, por lo que se hace necesario recortar el DEM de las misma forma en que fueron recortadas las imágenes. Las celdas para el DEM son de una resolución de 30x30 metros, coincidentemente con el tamaño de celda de las imágenes LandSat, pero no con las ASTER que son de 15x15 metros, debiéndose re-escalar el DEM a esa resolución, para ser utilizado en ese tipo de imágenes. Además el algoritmo nos pide como entradas los valores de Azimut y elevación solar al momento de la captura y el método de corrección topográfica que se utilizara, el cual es C-corrección.

Las fases de las correcciones de las imágenes y la información satelital de las imágenes necesarias para llevar a cabo estas correcciones se encuentran en el esquema de la figura 3.3.2.1.

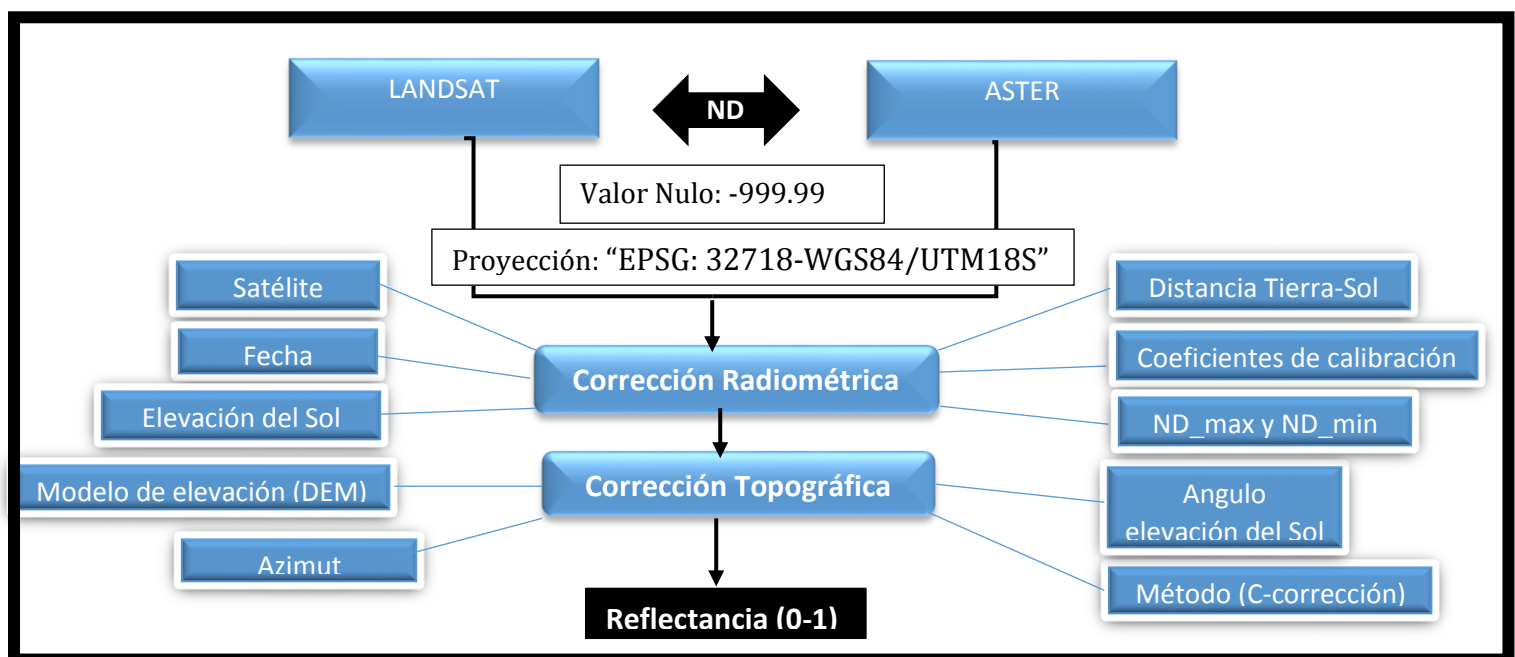


Figura 3.3.2.1- Esquema sobre el pre-procesamiento de las imágenes satelitales.
(Crédito: Elaboración propia).

3.3.3- Clasificación

La cartografía temática a partir de datos generados por teledetección, se basan principalmente en la clasificación de la imagen (Foody, 2005), proceso extensamente estudiado (Maulik U., 2009; Pilipi et.al, 2014). Una manera de extraer información temática de datos satelitales es a través de la interpretación visual. Sin embargo, esta se limita a la experticia del operador y a una subutilización de toda la información contenida en la imagen. Es por esto que la clasificación automática es más adecuada (Rozenstein y Karnieli, 2011).

Existen dos pasos fundamentales en una clasificación supervisada, la que a partir de un conjunto de píxeles conocidos, intenta asignar las clases a aquellos conjuntos desconocidos (Chuvieco E., 2006). Esto corresponde a la selección de áreas de entrenamiento y la aplicación de algoritmos de selección de píxeles a una clase, de acuerdo a las estadísticas de dichas áreas. Para este caso ocuparemos las máquinas de aprendizaje Random Forest y SVM.

Finalmente es necesario evaluar la precisión de la clasificación, para lo cual, se tabulan los datos a través de una matriz de confusión, de donde se derivan índices de concordancia. En el presente caso de estudio, se utilizaran regiones de entrenamiento seleccionadas por el usuario, de los cuales el 25% se usaran para el proceso de asignación, mientras que el resto se empleará en el de validación. El esquema general, que resume el proceso de clasificación se muestra en la figura 3.3.3.1.

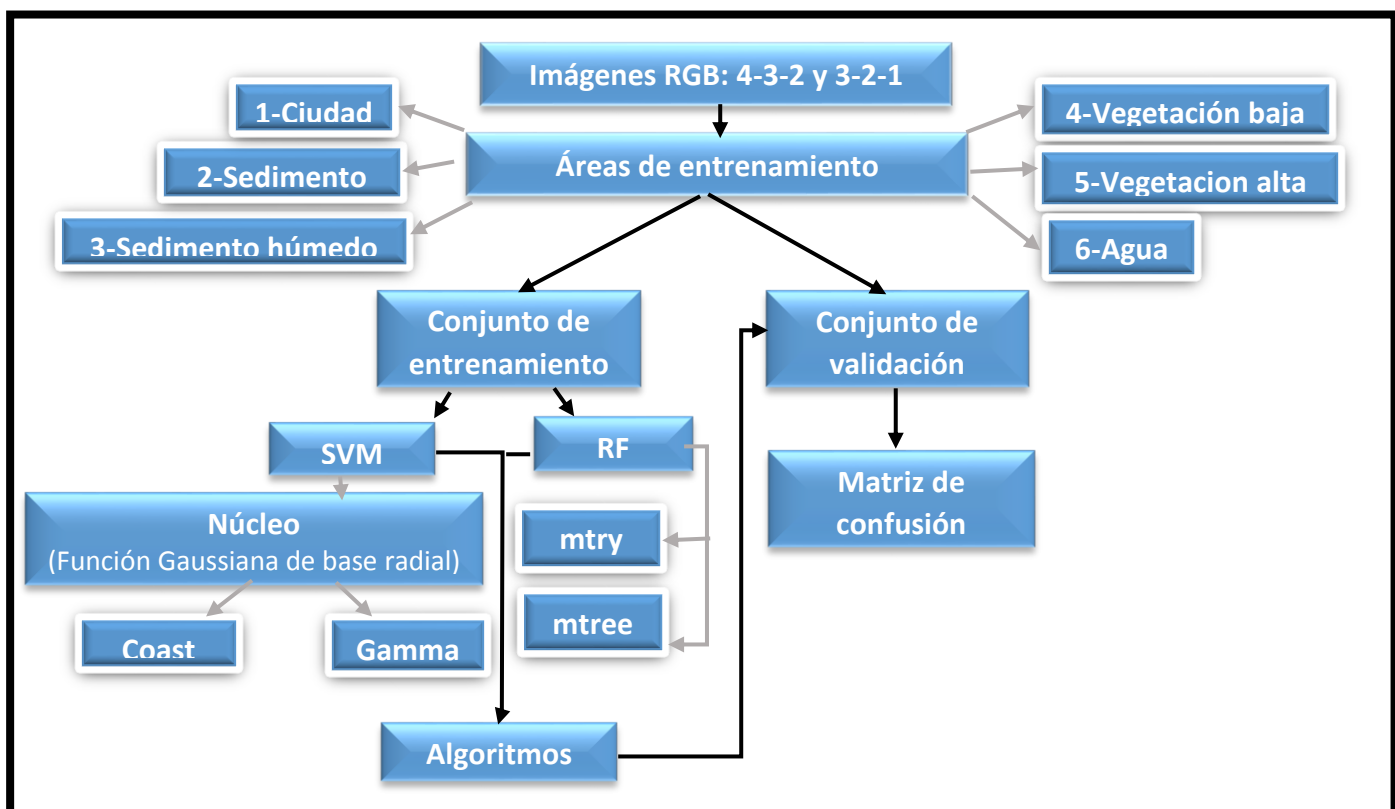


Figura 3.3.3.1- Esquema sobre el proceso de clasificación en imágenes satelitales.
(Crédito: Elaboración propia)

3.3.4- Combinación de bandas

En las bandas corregidas, se realiza una composición para las 6 bandas extraídas de Landsat y las 3 bandas extraídas de ASTER, a través de la herramienta “Semi-Automatic classification Plugin”- “Stack Raster Image” de QGIS, obteniendo las siguientes imágenes RGB de la figura 3.3.4.1. En esta se tienen 3 imágenes RGB para el 2009 (2 Landsat y una ASTER) y 4 imágenes RGB para el 2010 (3 Landsat y 1 ASTER).

Para el 2009, en las imágenes Landsat, se tiene que 3.3.4.1.A. se utiliza para las localidades ubicadas en el golfo de Arauco y 3.3.4.1.D. para la localidad de Dichato. Mientras que la imagen ASTER 3.3.4.1.F, se utiliza para las localidades del golfo de Arauco.

Por otra parte para el 2010, en las imágenes Landsat, 2 se utilizan para localidades del golfo de Arauco (3.3.4.1.B y 3.3.4.1.C) y 3.3.4.1.E. es utilizada para la localidad de Dichato. Mientras que la imagen ASTER 3.3.4.1.G. se utiliza para localidades de Arauco.

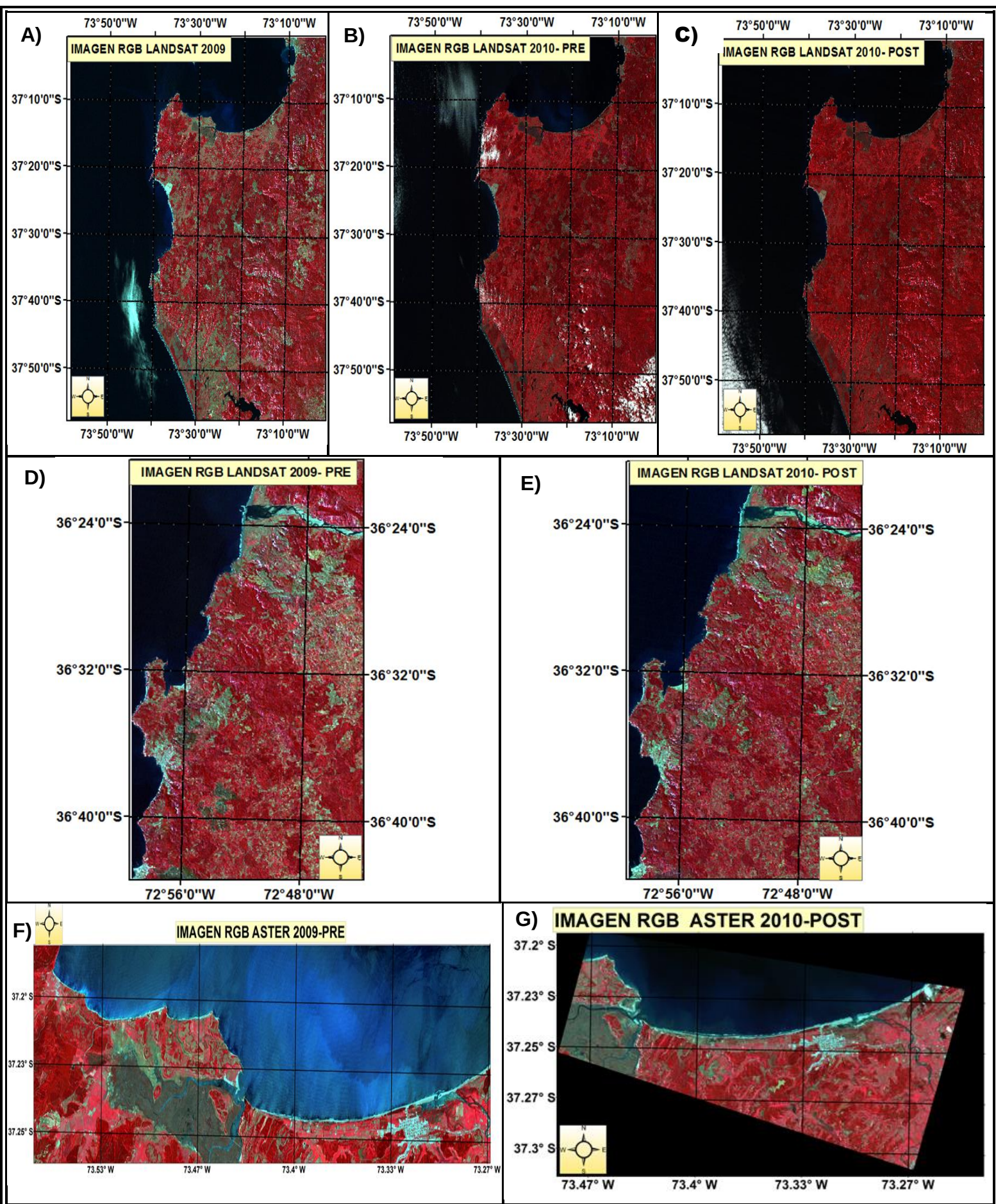


Figura 3.3.4.1- Imágenes de las combinaciones RGB para los periodos previos y posteriores al 27 de febrero del 2010.

Usando las entradas RGB, se creó un falso color 432 y 321 para las imágenes LandSat y ASTER respectivamente. Esto quiere decir que en la entrada del rojo se utilizó la banda 4 en Landsat y la banda 3 para ASTER. Para la entrada del azul se ingresó la banda 3 en Landsat y la banda 2 en ASTER. Finalmente para la entrada del verde se ingresó la banda 2 en Landsat y la banda 1 en ASTER.

En este tipo de combinación se ocupan junto a las bandas del visible, la banda infrarroja cercana (IRC) (banda 4 para LandSat y banda 3 para ASTER), la cual es útil para identificar límites entre suelo y agua. Los cuerpos de agua con sedimentos en suspensión aparecen en tonos azul claro y los que poseen poco o casi nulo material sedimentario aparecen en tonos azules más oscuros. Las áreas urbanas y el suelo expuesto aparecen en tonos celestes-blancos. También la banda IRC es sensible a la clorofila, permitiendo que se observen variaciones de la vegetación, los cuales aparecen en tonos rojos. Todas estas tonalidades son apreciables en las imágenes de la figura 3.3.4.1., lo que nos indica la presencia de varios tipos o clases distintas de cobertura de suelo.

3.3.5- Creación de áreas de entrenamiento:

En esta etapa se ocupó el software ERDAS Imagine, el cual es un software para la detección con capacidad de editor gráfico raster diseñado por ERDAS para aplicaciones geoespaciales. En ERDAS se cargan las composiciones de bandas de cada una de las imágenes y se crean polígonos (shape) para seleccionar cada una de las siguientes clases temáticas de interés:

- 1- Ciudad: representa la cobertura correspondiente al suelo urbano y a rocas expuestas que al igual que el cemento presentan un alto albedo, es decir, una gran cantidad de la energía incidente es reflejada.
- 2- Sedimento seco: representa la cobertura correspondiente a suelo expuesto con bajo contenido de humedad y materia orgánica, como arena, laderas expuestas, áreas no cultivadas, zonas taladas, áreas quemadas, áreas de erosión y áreas con prácticamente nula vegetación.
- 3- Sedimento húmedo: representa la cobertura correspondiente a suelo con un alto contenido de humedad, como lo son los suelos de pantanos. También corresponde a tierras de cultivo y depósitos costeros.
- 4- Vegetación baja: formación vegetal donde la cobertura herbácea es sobre el 40%, esto incluye los terrenos con rotación de cultivos, tipos de árboles y arbustos con un área de extensión menor al 25%. Áreas utilizadas por la agricultura, incluyendo cultivos de cereales, hortalizas y frutas.
- 5- Vegetación alta: cobertura vegetal en la cual el estrato arbóreo es establecido en parte por especies naturales como Coihue, Olivillo, Patagua y Boldo. Por otro lado, también representa bosques en donde el estrato arbóreo es principalmente formado por especies exóticas como el eucaliptus y pino radiata.
- 6- Agua: superficies cubiertas por agua, tanto dulce como salada.

Los polígonos creados para cada una de estas coberturas de interés o áreas de entrenamiento, se visualizan en las siguientes figuras, las cuales están para diferentes periodos de tiempo. Donde la figura 3.3.5.1. corresponden a las áreas de entrenamiento para imágenes Landsat previo al 27/F, la figura 3.3.5.2. a las áreas de entrenamiento para imágenes Landsat posterior al 27/F, la figura 3.3.5.3 a las áreas de entrenamiento para imágenes ASTER previo al 27/F y por último la figura 3.3.5.4. a las áreas de entrenamiento para imágenes ASTER posterior al 27/F.

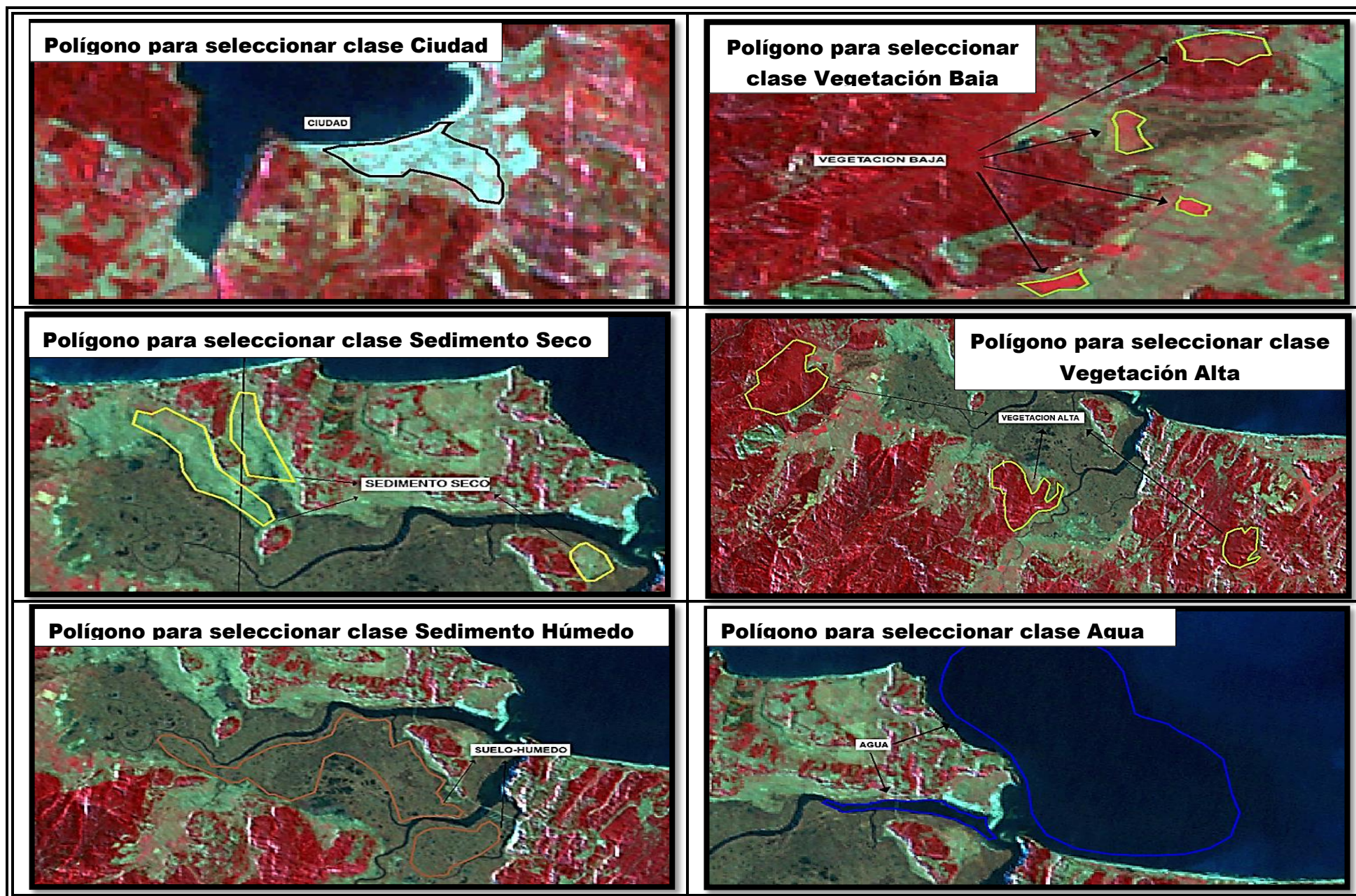


Figura 3.3.5.1- Áreas de entrenamiento en Imágenes LandSat para periodo de tiempo previo al 27 de febrero del 2010.

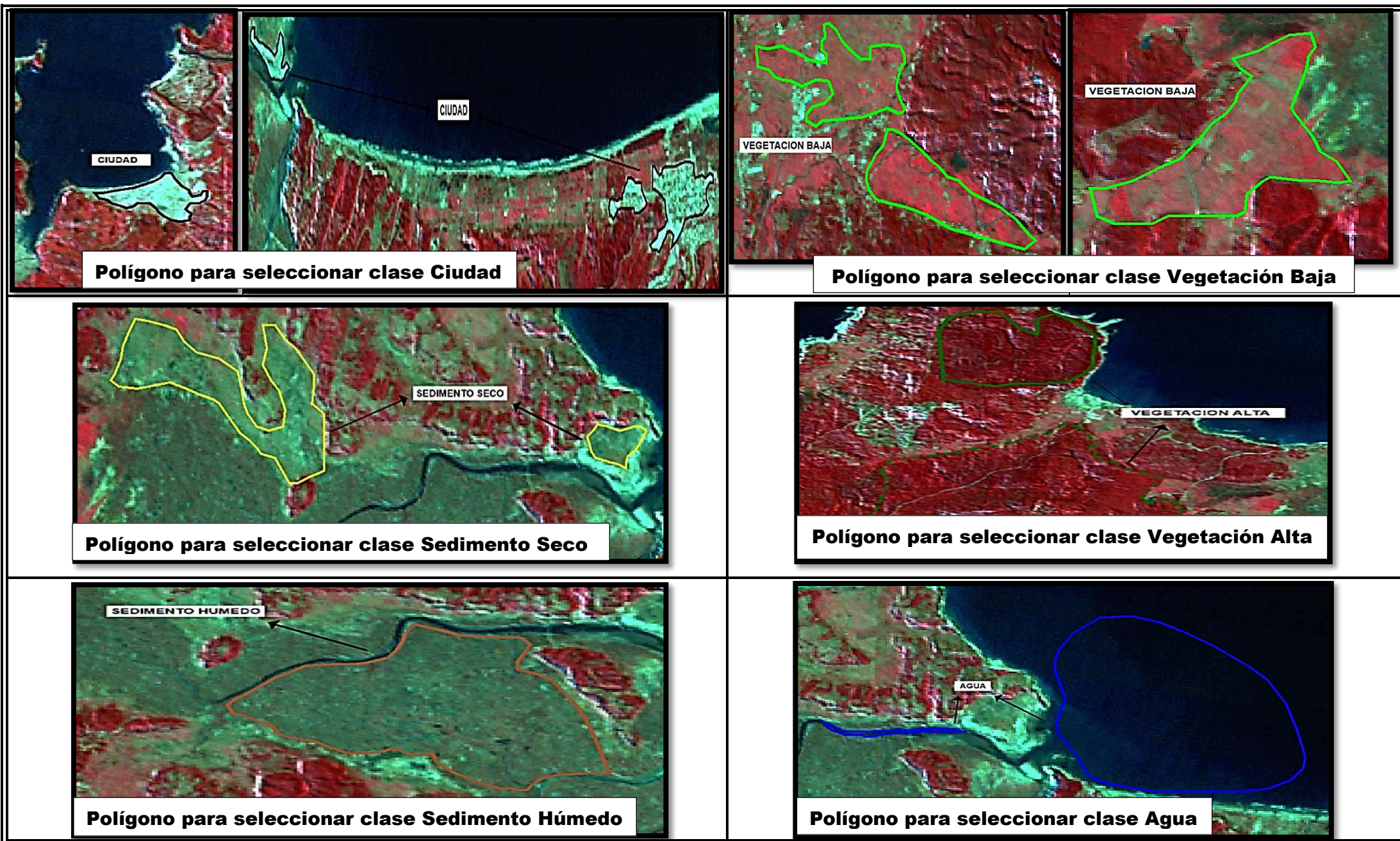


Figura 3.3.5.2- Áreas de entrenamiento en imágenes Landsat para periodo de tiempo posterior al 27 de febrero del 2010.

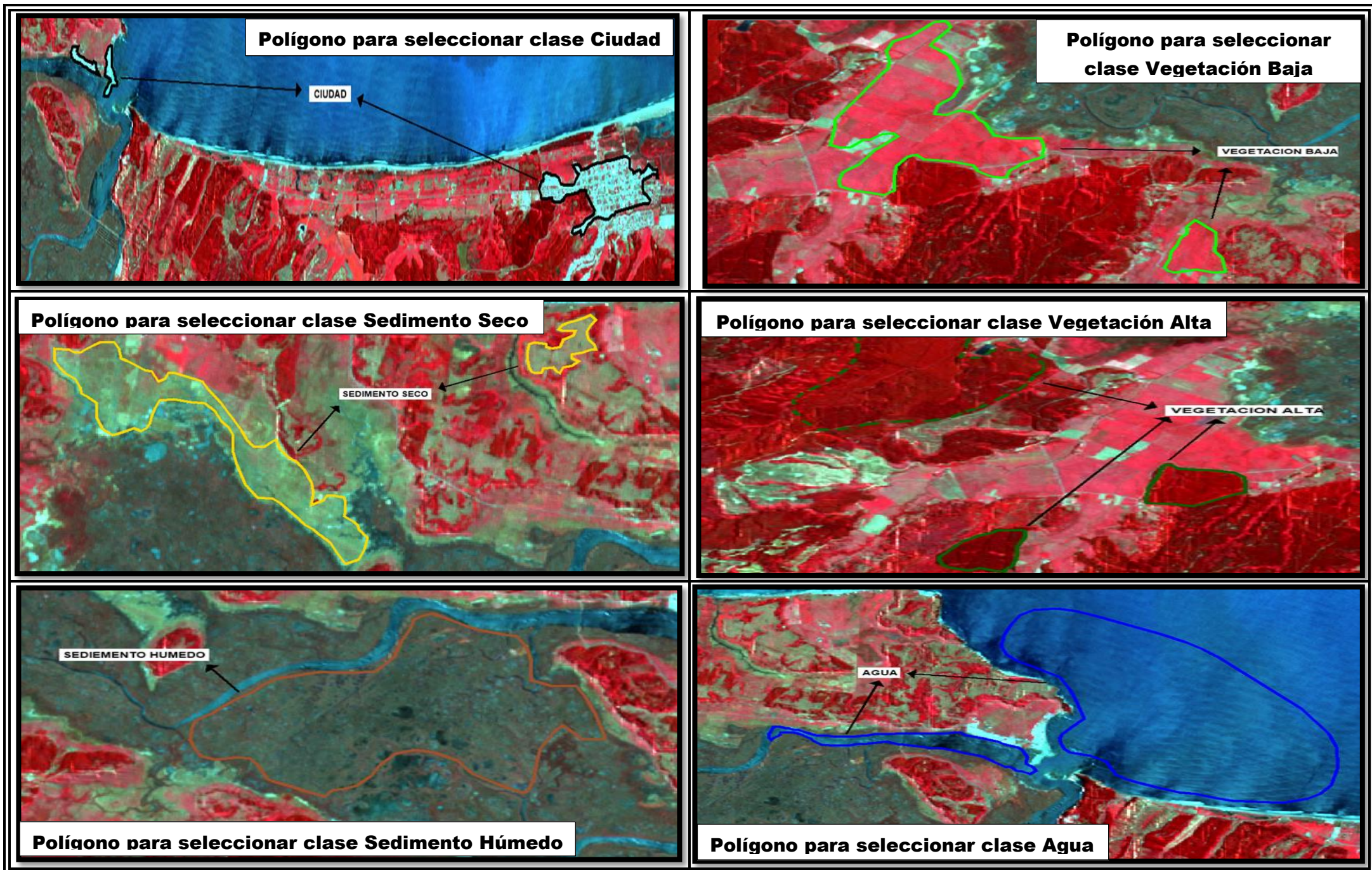


Figura 3.3.5.3- Áreas de entrenamiento en imágenes ASTER para periodo de tiempo previo al 27 de febrero del 2010.

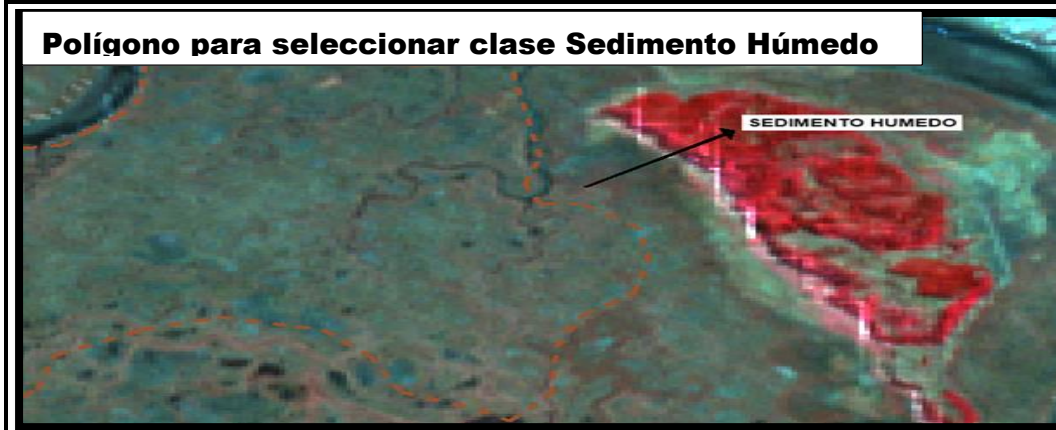


Figura 3.3.5.4- Áreas de entrenamiento en imágenes ASTER para periodo de tiempo posterior al 27 de febrero del 2010.

La creación de los polígonos nos permitirá seleccionar los píxeles con valores de reflectancia similares para identificar un tipo de cobertura de suelo. Por lo que se tendrá un archivo en donde se combinen la información espectral extraída de las áreas de entrenamiento de cada clase, a través de la herramienta de ERDA: “extract value pixel”. Procedimiento que entregara un archivo .xml, con los valores de los píxeles para cada una de las bandas que componen cada composición de bandas.

Para visualizar si nuestras áreas de entrenamiento seleccionadas a través de polígonos son representativas y útiles para diferenciar las diferentes coberturas, se utiliza la herramienta “Signature Editor” de ERDAS, mediante la cual se crean graficas que representan las respuestas espectrales de cada una de las coberturas a lo largo de las regiones espectrales que representan cada banda, para cada una de las imágenes y separándolas para los dos periodos de tiempo, antes y después del 27F.

En la figura 3.3.5.5 se pueden visualizar las respuestas espectrales de las diferentes coberturas de interés a lo largo de las 6 bandas de la composición de bandas para las imágenes LandSat previo y posterior al tsunami del 2010. Se puede distinguir que las categorías se logran diferenciar de buena forma en todos los rangos espectrales.

Se observa comportamientos similares para categorías que representan coberturas de similares características físicas y biológicas, como lo es entre la vegetación baja y vegetación alta, los cuales alcanzan su pick de reflectancia alrededor de la banda 4 y de ahí decaen, comportamiento típico de coberturas vegetales ([Chuvieco E., 2006](#)).

También existe un comportamiento similar entre las coberturas de suelo (sedimento seco, sedimento húmedo y ciudad), los cuales tienen una curva ascendente, que llega a su punto más alto de reflectancia en torno a la banda 4 donde empieza a decaer un poco, comportamiento típico de coberturas de suelo sin vegetación ([Chuvieco E., 2006](#)).

Para el caso del agua, representa una curva que tiene sus mayores valores de reflectancia en el rango del visible (bandas 1, 2 y 3) y luego decae hasta valores muy cercanos a 0.

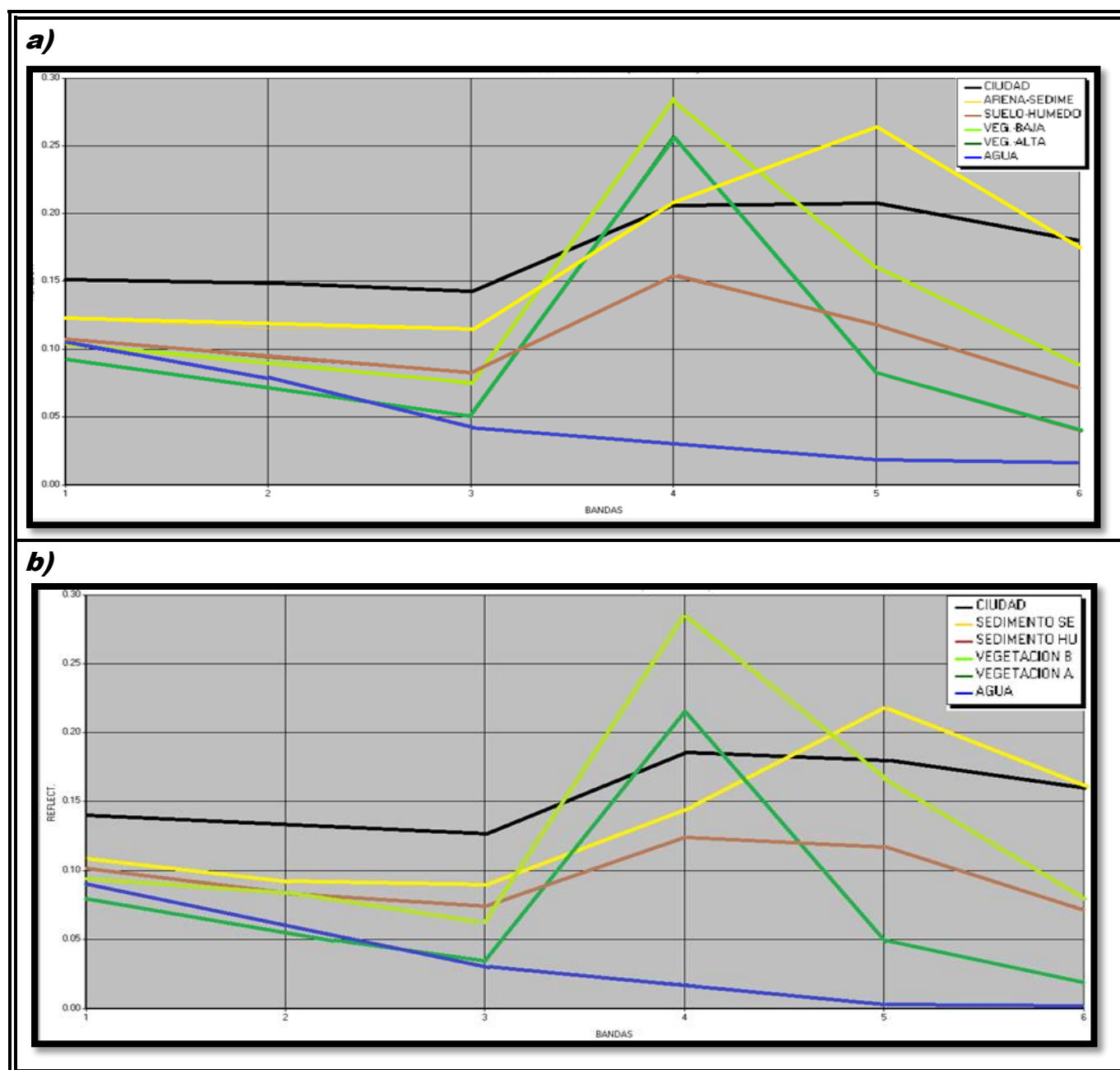


Figura 3.3.5.5.- Gráficos que representan las respuestas espectrales de las diferentes coberturas de interés a lo largo de las 6 bandas para imágenes Landsat. a) Previo 27/F; b) Posterior 27/F.

En la figura 3.3.5.6 se pueden visualizar las respuestas espectrales de las diferentes coberturas de interés a lo largo de las 3 bandas para las imágenes ASTER previo y posterior al tsunami del 2010.

Se puede distinguir que las categorías se logran diferenciar de buena forma en todos los rangos espectrales, observándose comportamientos similares para categorías que representan coberturas de similares características físicas y biológicas. Ejemplo de lo anterior es la vegetación baja y vegetación alta, los cuales van alcanzando sus mayores valores de reflectancia a medida que se llega a la banda 3, es decir a los rangos espectrales del infrarrojo cercano, comportamiento típico de coberturas vegetales.

También existe un comportamiento similar entre las coberturas de suelo (sedimento seco, sedimento húmedo y ciudad), los cuales tienen una curva ascendente, comportamiento típico de coberturas de suelo sin vegetación. Para el caso del agua, representa una curva que tiene sus mayores valores de reflectancia en el rango del visible y luego decae hasta valores muy cercanos a 0.

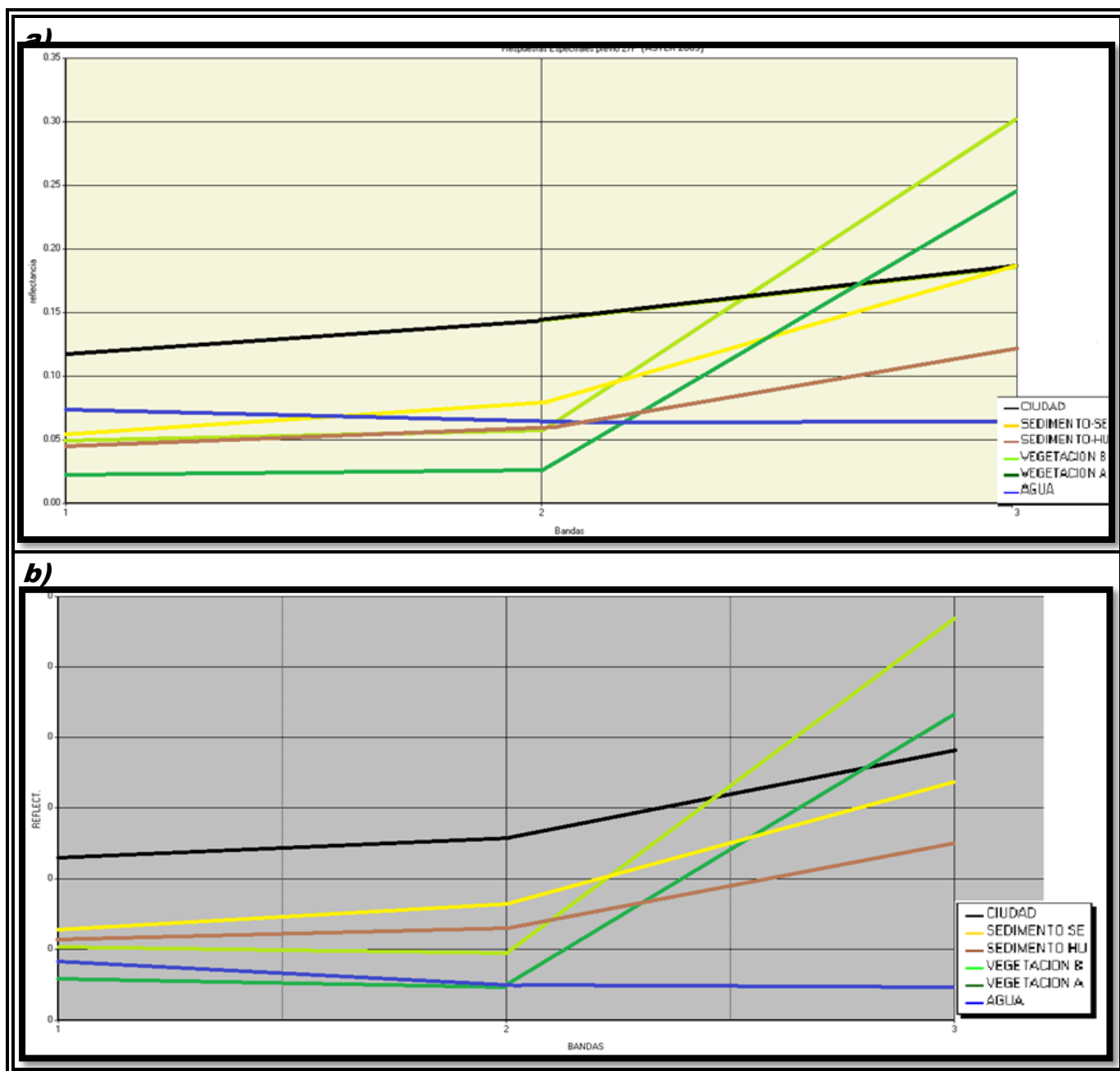


Figura 3.3.5.6- Gráficos que representan las respuestas espectrales de las diferentes coberturas de interés a lo largo de las 3 bandas para imágenes ASTER. a) Previo 27/F; b) Posterior 27/F.

3.4- Preparación de datos:

Se utilizará el software R para realizar el proceso de clasificación SVM y RF. Los paquetes del software R son libres y de fuente abierta ([R Core Team \(2013\) URL: http://www.R-project.org](http://www.R-project.org)). Estas proveen una gran variedad de funciones para análisis estadísticos. Recientemente R ha desarrollado importantes herramientas en percepción remota, ya que ofrece implementaciones de una variedad de métodos de clasificación de imágenes en los que destacan SVM y RF (["http://cran.r-project.org"](http://cran.r-project.org)). También se puede utilizar para crear matrices de error y evaluar la exactitud de la clasificación.

Aunque R puede leer imágenes en formato raster, necesariamente se debe preparar archivos de entrenamiento y validación (evaluación de precisión). Esencialmente lo que se necesita es un archivo en el cada pixel sea separado por una fila de los datos y que las columnas separen cada valor de reflectancia en las diferentes bandas para cada uno de los pixeles, pero que la primera columna identifique la clase a la que corresponde cada uno de los pixeles.

Con este objetivo se realiza el proceso de selección de áreas de entrenamiento usando los shape previamente creados para cada una de las coberturas de interés. En ERDAS se leen los shape creados como capas AOI (área de interés) y luego usando la herramienta "Subset" en la opción Raster, se realiza el recorte de las áreas de interés en las composiciones de bandas. Luego se utiliza la herramienta "Pixel to ASCII" en la opción Manage Data, el cual genera un archivo .asc, donde se tendrá la información de los pixeles que correspondan al área del shape, asignándole un número que identifique su clase a cada pixel. Se repite el proceso para los shape de todas las clases y se combinan todos los archivos en uno solo (allpix.csv). Como es necesario tener dos conjuntos, el de entrenamiento y el de validación, se debe dividir el archivo "allpix.csv", para esto se ordenan de forma aleatoria todos los pixeles para asegurar que los dos archivos contengan pixeles correspondientes a todas las clases. El archivo de entrenamiento (Train.csv) contiene el 25% de los datos de "allpix" y el archivo de validación (validación.csv) el otro 75% de los pixeles.

3.5- Procesamiento de datos

Preparados los datos, se comienza a trabajar en R:

- 1- Se instalan los siguientes paquetes:
 - 1.1- e1071 (implementación de SVM).
 - 1.2- randomForest (implementación de RF).
 - 1.3- sp, raster, rgdal (estos 3 paquetes permiten leer datos raster en R)
- 2- Luego se cargan los datos de entrenamiento y validación. Se necesita especificar que la primera columna corresponde a las clases de los pixeles en los datos de validación por lo que deben ser ignorados, ya que el programa de clasificación espera sólo variables predictoras en los datos de validación y porque al usar valores numéricos para identificar las clases, se debe definir el campo "Clase" como categórico, como oposición a una variable continua/numérica.
- 3- Por último se carga la imagen a ser clasificada.

3.5.1- Entrenando el clasificador

Se entrena el clasificador mediante los datos de entrenamiento, usando la función “svm()” disponible en el paquete e1071. SVM requiere un número de parámetros definidos por el usuario. El primero es el tipo de núcleo, para este proyecto se utiliza la función gaussiana de base radial (radial basis). La función base radial tiene dos parámetros adicionales a definir: “gamma” y “coast”. Los cuales en una primera etapa van a tener sus valores predeterminados por el programa. Para gamma va a ser el inverso del número de variables predictoras (el número de clases en el que se dividen los datos de entrenamiento). Para “coast” será 1.

Para entrenar el modelo RF, se utiliza la función randomForest() disponible en el paquete randomForest. RF requiere dos parámetros definidos por el usuario, “mtry” (el número de variables disponibles para la división de cada nodo) por default es la raíz cuadrada del número de variables predictoras. “mtree” (el número de árboles de decisión producidos) 500 por default.

3.5.2- Optimizando los algoritmos

La optimización de los algoritmos es un paso necesario para la creación de algoritmos basados en máquinas de aprendizaje, debido a ser muy sensibles a los parámetros definidos por el usuario. Por lo que se debe realizar una optimización de parámetros definidos por el usuario para lograr una mejora en la precisión de la clasificación. Esto se logra usando el método de testeo de combinaciones de parámetros mediante validación cruzada (10-fold) a través del paquete e1071. Este testeo, en términos generales, va prueba combinaciones de parámetros y evalúa la reducción en precisión de los algoritmos, entregando la combinación con los mejores resultados. Los parámetros optimizados obtenidos fueron:

SVM → gamma: 10

coast: 1000

RF → mtree: 200

mtry: 3

Después se corren nuevamente los modelos originales usando estos parámetros optimizados en orden a mejorar la precisión de la clasificación.

Se puede visualizar en la tabla 3.5.1.1. los resultados de la evaluación de importancia de las variables de entrada, es decir las bandas de las imágenes, utilizando la función “importance()” sobre el modelo de RF. Al desplegarse los resultados, se muestran en la primera columna los números de las bandas. Las siguientes columnas reflejan la importancia de cada banda en cada una de las clases. Estas mediciones deben ser consideradas como mediciones relativas de importancia. En la penúltima columna están los valores “obb” (out of the bag) decrecimiento en mediciones de precisión (MeanDecreaseAccuracy) es usada para evaluar la precisión en contraste con la medición de Gini, que se encuentra en la última columna (MeanDecreaseGini). La idea del decrecimiento medio en precisión es que representa la pérdida promedio de precisión cuando una banda en particular es eliminada desde los modelos. Cuanto mayor es el valor para obb-decrecimiento medio en precisión, más importante es la variable predictora que produjo el modelo.

Tabla evaluación de importancia de las variables (LANDSAT-2009)								
	1	2	3	4	5	6	MeanDecreaseAccuracy	MeanDecreaseGini
B1	3.692386	3.902352	2.029935	3.679340	4.809794	101.542683	5.694730	4336.3190
B2	39.524859	8.463213	4.688177	12.758969	3.920246	4.180170	14.202765	7506.3995
B3	2.287679	2.312215	2.558448	22.231551	157.366731	16.511300	34.828518	7956.7637
B4	2.096242	1.955600	79.198343	3.057107	7.653799	3.748519	6.846918	2255.6842
B5	1.329415	1.997284	40.645900	4.133271	1.823668	2.873235	5.387613	1510.5899
B6	1.477645	1.898702	2.856702	1.483897	1.345034	1.244970	1.675450	711.5265
Tabla evaluación de importancia de las variables (LANDSAT-2010-previo)								
	1	2	3	4	5	6	MeanDecreaseAccuracy	MeanDecreaseGini
B1	4.809756	103.468503	2.980204	19.126375	6.596850	9.880541	13.802911	5555.96661
B2	66.441275	6.205750	8.748942	8.792006	1.631282	3.730405	10.236417	3374.85137
B3	2.775058	1.525711	3.262770	3.911450	2.676623	10.262061	6.873431	1140.82993
B4	1.759260	2.594571	6.646881	2.451349	4.601836	2.681837	4.688748	2176.53492
B5	2.305233	3.173813	9.431151	5.591704	1.641661	3.893912	4.738993	2867.99296
B6	1.081121	1.261720	2.045627	1.998208	1.561699	2.108896	3.002679	61.61006
Tabla evaluación de importancia de las variables (LANDSAT-2010-posterior)								
	1	2	3	4	5	6	MeanDecreaseAccuracy	MeanDecreaseGini
B1	3.950057	3.256675	1.081246	3.468570	3.834674	57.178580	5.853617	2795.520
B2	40.871588	12.891535	1.774750	2.681341	1.387864	1.048987	3.581749	3234.255
B3	4.739289	3.573159	1.472462	9.992512	11.397472	8.920961	13.104885	5281.562
B4	1.340202	2.366916	45.745712	22.649821	10.268593	3.349643	12.189804	1897.385
B5	1.301532	1.950395	18.888807	15.209777	2.152659	4.333672	8.236121	2931.857
B6	1.155673	1.622740	3.131116	1.164329	2.105038	1.166741	1.473158	1028.695
Tabla evaluación de importancia de las variables (ASTER-2009)								
	1	2	3	4	5	6	MeanDecreaseAccuracy	MeanDecreaseGini
B1	10.48406	9.408566	2.581321	3.154136	0.0000	0.0000	5.341528	1123.716
B2	5.26931	1.054093	3.852811	7.689056	311.6115	0.0000	23.516484	1805.123
B3	143.35366	525.301060	332.254920	195.192868	415.3149	238.1002	314.945304	19320.480
Tabla evaluación de importancia de las variables (ASTER-2010)								
	1	2	3	4	5	6	MeanDecreaseAccuracy	MeanDecreaseGini
B1	4.110122	3.447863	3.738200	0.0000	0.0000	0.0000	5.149673	504.8944
B2	62.203545	158.388467	2.536255	143.5449	121.4076	0.0000	115.739023	4815.0315
B3	36.043956	118.718739	880.079788	223.2877	322.5711	448.5167	575.162073	16461.9090

Tabla 3.5.1.1- Evaluación de importancia en los modelos RF de las variables (bandas) de las composiciones de bandas LandSat y ASTER.

En la visualización de los resultados de la evaluación de importancia de las variables sobre los modelos RF desarrollados para las imágenes raster mostrados el tabla 3.5.1.1., señalan que los valores altos de MeanDecreaseAccuracy se encuentran en las bandas 2-3-4 de LANDSAT, las que contribuyen en mayor medida a la elaboración del modelo RF. Estos valores altos concuerdan con lo vistos en las gráficas de respuesta espectral, en las que se diferencian claramente los valores de reflectancia promedio para cada tipo de cobertura.

La banda 2 tiene un gran valor de importancia para la clase ciudad, la banda 3 destaca su valor de importancia para la clase de vegetación alta y para la banda 4 destaca su valor de importancia para la clase sedimento húmedo. Siendo estas razones las que hacen que contribuyan mayormente a la formación del modelo. La banda 6 posee

bajos valores de MeanDecreaseAccuracy, lo que señala que incluso se podría prescindir de ella, siendo esto ratificado en las respuestas espectrales de las diferentes coberturas, siendo estas bastantes similares en el rango espectral de la banda 6.

Para las imágenes ASTER, la banda 3 es la que posee los mayores valores de MeanDecreaseAccuracy, contribuyendo fuertemente en la creación del modelo, debido a su fuerte importancia para todas las coberturas estudiadas. Coincidiendo lo anterior con lo visto en las gráficas de respuesta espectral, mostrándose en estas que en el rango espectral de la banda 3 (rojo-IRC), todas las coberturas alcanzan su mayor distinción entre ellas. Esto nos habla que para la creación del modelo, podría realizarse con una buena exactitud usando solo la banda 3.

4- Resultados

Una vez entrenados los algoritmos SVM y RF, se procede a validar estos modelos (realizar una evaluación de precisión) usando los datos de validación previamente generados. Esto nos entregara como resultado la precisión de nuestros modelos, visualizando de esta manera su potencial uso como clasificador de las coberturas de suelo en las imágenes. Primero se necesita generar la clasificación utilizando los datos de validación y los modelos (argumentos requeridos en la clasificación), que representan la regla de clasificación. El siguiente paso será aplicarla. Mediante el paquete “epicalc” se crean matrices de error, con lo que se puede llegar al valor de la confiabilidad global o concordancia total de los modelos y el valor de índice kappa. En la tabla 4.1. se representaran las matrices de error en conjunto a los valores de concordancia total e índice kappa para cada uno de los modelos generados a partir de las diferentes imágenes satelitales ocupadas y distintas fechas de adquisición.

LANDSAT_2009 (PREVIO 27 F)				SVM modelo		
	1	2	3	4	5	6
1	483	7	0	3	1	0
2	21	1805	2	9	11	0
3	1	10	2799	1	5	14
4	12	6	0	653	88	0
5	0	4	0	42	1929	0
6	2	0	8	0	0	3134
CONCORDANCIA TOTAL: 97.76%						
KAPPA: 0.972						
LANDSAT_2009 (PREVIO 27 F)				RF modelo		
	1	2	3	4	5	6
1	488	6	1	7	0	0
2	25	1801	2	11	9	0
3	0	12	2795	4	2	16
4	5	5	2	618	81	0
5	1	8	1	67	1941	0
6	0	0	8	1	1	3132
CONCORDANCIA TOTAL: 97.51%						
KAPPA: 0.968						

LANDSAT_2010 (PREVIO 27 F)				SVM modelo		
	1	2	3	4	5	6
1	492	5	0	30	6	0
2	4	2005	2	38	33	0
3	2	22	2843	7	2	7
4	16	43	0	3646	123	0
5	4	22	1	198	4396	0
6	0	2	6	0	0	2937
CONCORDANCIA TOTAL: 96.61%						
KAPPA: 0.957						
LANDSAT_2010 (PREVIO 27 F)				RF modelo		
	1	2	3	4	5	6
1	489	13	2	34	5	0
2	6	1975	4	49	27	0
3	1	24	2836	6	3	3
4	19	45	2	3665	109	0
5	3	40	1	165	4416	0
6	0	2	7	0	0	2941
CONCORDANCIA TOTAL: 96.63%						
KAPPA: 0.958						
LANDSAT_2010 (POSTERIOR 27 F)				SVM modelo		
	1	2	3	4	5	6
1	631	28	0	2	3	0
2	15	1925	14	12	39	7
3	4	47	2786	1	8	0
4	2	9	0	1839	99	0
5	19	23	5	98	21652	0
6	0	1	1	0	0	2406
CONCORDANCIA TOTAL: 98.62%						
KAPPA: 0.973						
LANDSAT_2010 (POSTERIOR 27 F)				RF modelo		
	1	2	3	4	5	6
1	614	29	5	3	8	0
2	37	1911	23	17	52	4
3	5	65	2763	2	5	0
4	1	6	0	1799	94	0
5	14	22	15	131	21639	0
6	0	0	0	0	3	2409
CONCORDANCIA TOTAL: 98.29%						
KAPPA: 0.966						

ASTER_2009 (PREVIO 27 F)				SVM modelo		
	1	2	3	4	5	6
1	1447	22	3	18	1	4
2	22	1341	27	77	29	0
3	5	45	6882	5	8	60
4	11	73	2	9497	230	0
5	5	30	3	88	14338	0
6	10	1	26	0	4	13697
CONCORDANCIA TOTAL: 98.32%						
KAPPA: 0.978						
ASTER_2009 (PREVIO 27 F)				RF modelo		
	1	2	3	4	5	6
1	1445	30	2	25	1	10
2	22	1301	47	81	41	0
3	7	55	6855	2	8	72
4	16	83	3	9507	204	0
5	1	42	5	70	14351	1
6	9	0	31	0	5	13678
CONCORDANCIA TOTAL: 98.18%						
KAPPA: 0.976						
ASTER_2009 (PREVIO 27 F)				MV modelo		
	1	2	3	4	5	6
1	1121	27	2	23	80	1305
2	139	1245	63	393	23	0
3	16	83	6798	1	15	577
4	91	95	1	9017	414	0
5	0	59	6	251	14076	0
6	123	0	73	0	2	11879
CONCORDANCIA TOTAL: 91.19%						
KAPPA: 0.883						
ASTER_2010 (POSTERIOR 27 F)				SVM modelo		
	1	2	3	4	5	6
1	489	15	0	0	6	4
2	40	1309	101	141	50	0
3	3	212	8142	7	24	25
4	20	121	2	3481	350	0
5	1	9	0	194	5810	0
6	0	0	1	0	0	9037
CONCORDANCIA TOTAL: 95.53%						
KAPPA: 0.941						
ASTER_2010 (POSTERIOR 27 F)				RF modelo		
	1	2	3	4	5	6
1	485	16	0	10	4	0
2	33	1336	137	146	52	0
3	4	179	8077	6	17	21
4	25	116	5	3405	291	0
5	6	22	3	256	5876	0
6	0	1	24	0	0	9041
CONCORDANCIA TOTAL: 95.36%						
KAPPA: 0.939						

ASTER_2010 (POSTERIOR 27 F)			MV modelo			
	1	2	3	4	5	6
1	428	13	0	0	6	693
2	69	1337	417	264	94	0
3	4	199	7826	8	19	9
4	22	103	1	3239	407	0
5	0	14	1	312	5191	429
6	30	0	1	0	523	7933
CONCORDANCIA TOTAL: 0.877%						
KAPPA: 0.823						

Tabla 4.1- Matrices de confusión e índices kappa para los modelos de clasificación SVM, RF y MV, utilizados en las imágenes y distintos periodos estudiados. (1-Ciudad; 2-Sedimento seco; 3-Sedimento Húmedo; 4-Vegetacion Baja; 5-Vegetacion Alta; 6-Agua).

Como salida del procesamiento de validación de los modelos SVM, RF y MV, a través de la clasificación aplicando los modelos creados sobre los denominados conjuntos de validación, se obtienen como salida unas matrices de confusión e índices kappa que dieron distintos valores de confiabilidad sobre los modelos que se aplicaran a las imágenes completas.

Los valores entregados de precisión son bastantes óptimos para los modelos SVM y RF, variando entre un 95% y 98%. Estas variaciones se debieron principalmente a los tamaños de las muestras de entrenamiento usadas, dado que para los conjuntos con menor cantidad de muestras la precisión estuvo cercana al 95%, en cambio para los conjuntos más numerosos la precisión global supero el 98%.

Por otra parte el modelo de máxima verosimilitud tiene una precisión menor (cercana al 90%) comparado a los otros dos modelos. Esto es debido, principalmente, a que depende en un alto grado del número de muestras que se tengan disponibles para cada categoría, con el fin de determinar con suficiente precisión tanto la media como la matriz de la covarianza. Además que requiere de un mayor tiempo de proceso. Cabe señalar que los niveles de precisión son bastante óptimos en la mayoría de los casos, independiente del modelo usado, lo cual se debería principalmente a que las clases temáticas a utilizar separan coberturas de suelo claramente diferenciables entre sí, lo que ayuda en gran medida que no exista una gran confusión entre las clases. Por otra parte el número de muestras de entrenamiento son abundantes en todos los casos, ayudando a la optimización de los algoritmos.

Una vez obtenidos los modelos óptimos para una mejor clasificación, se procede a aplicar los métodos de clasificación SVM y RF para todas las imágenes corregidas. El método MV solo fue aplicado en las imágenes ASTER, ya que requiere de un mayor tiempo de proceso. El resultado de esto son mapas raster o mapas temáticos que pueden visualizarse mediante cualquier software geo-espacial. Estos mapas temáticos contarán con las seis clases en las que se ha deseado categorizar la

cobertura de suelo, dando cuenta de cómo se distribuía la cobertura de suelo en la región para los periodos de tiempo determinados.

A continuación se presentan los mapas temáticos resultantes, en las figuras 4.1. a 4.5. Las figuras 4.1 y 4.2 representan los mapas temáticos de las imágenes previas al 27/F a partir de la clasificación con el modelo SVM y RF, respectivamente. Las figuras 4.3 y 4.4 representan los mapas temáticos de las imágenes posteriores al 27/F a partir de la clasificación con el modelo SVM y RF, respectivamente. Por último la figura 4.5 representa los mapas de las imágenes previas y posteriores al 27/F a partir de la clasificación utilizando los modelos de MV.

En cuanto a los colores, estos tendrán las mismas significancias para todas las imágenes, donde el negro representa cobertura de ciudad o roca expuesta, el amarillo para suelo seco o con escasa cobertura vegetal, el café para suelos húmedos, el verde claro para vegetación baja, el verde oscura para vegetación alta y el azul para representar el agua.

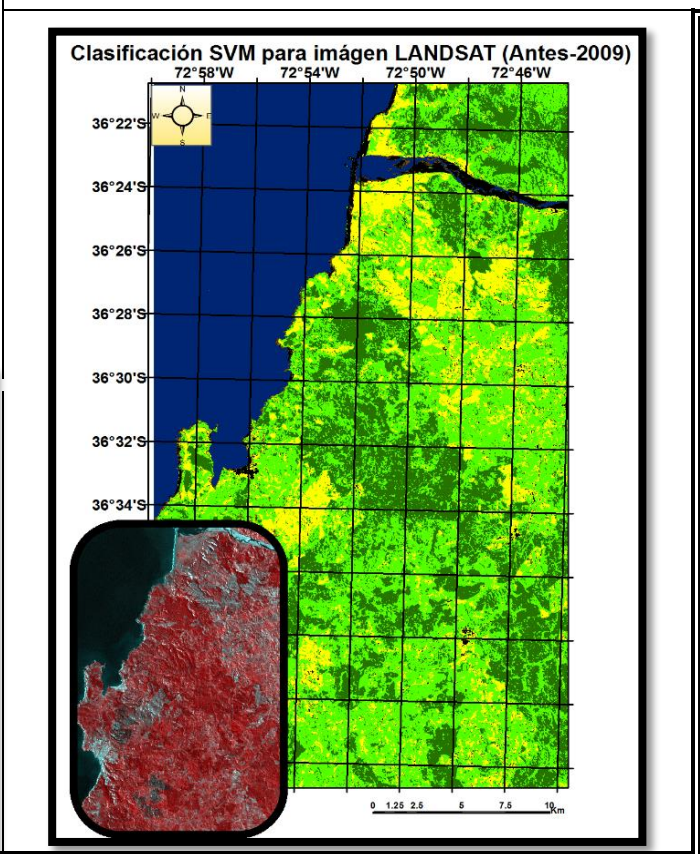
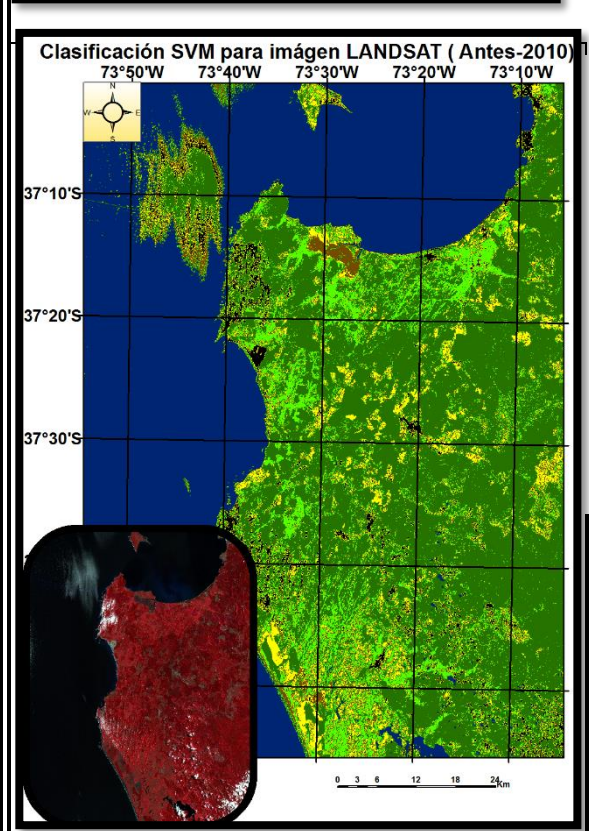
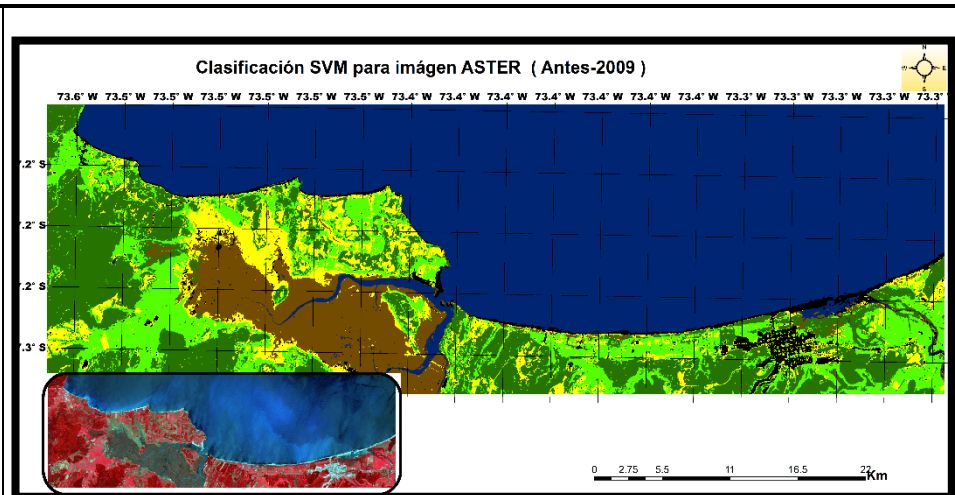
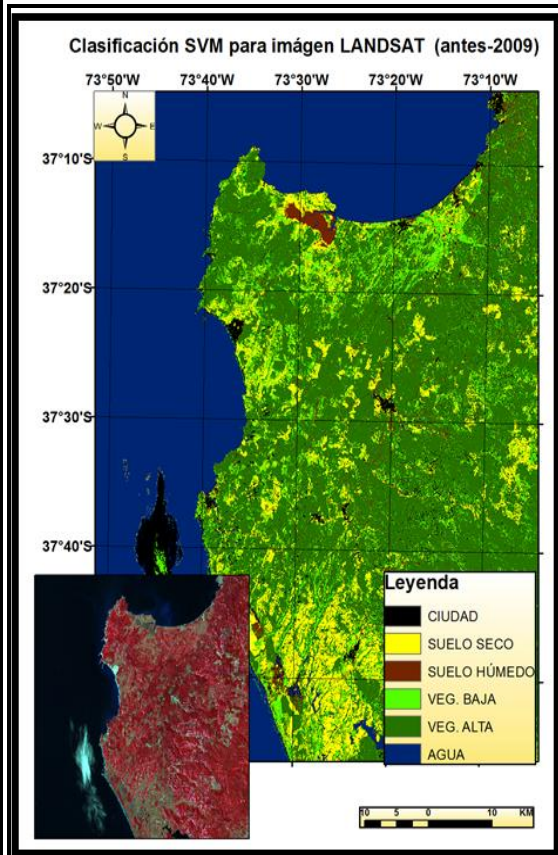


Figura 4.1 - Mapas tematicos a partir de las composiciones de bandas de las imágenes satelitales, usando la clasificación del modelo SVM, para periodo previo al 27 de febrero del 2010.

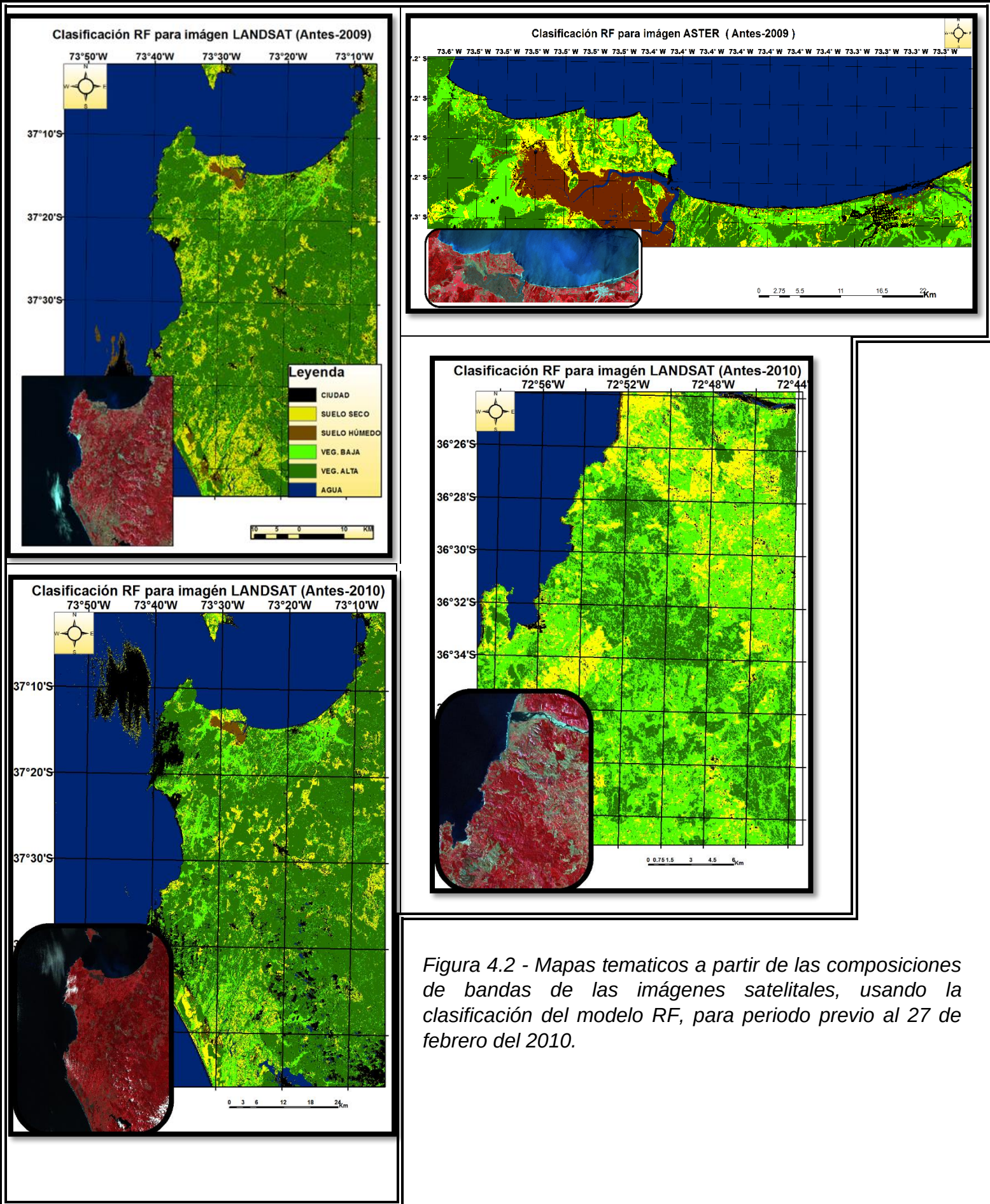


Figura 4.2 - Mapas temáticos a partir de las composiciones de bandas de las imágenes satelitales, usando la clasificación del modelo RF, para periodo previo al 27 de febrero del 2010.

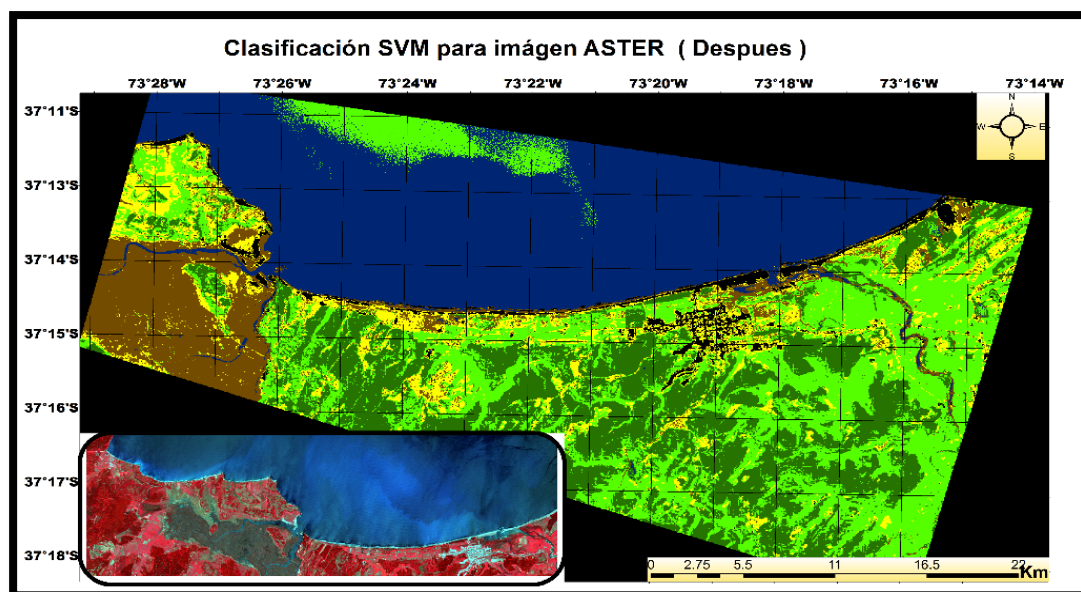
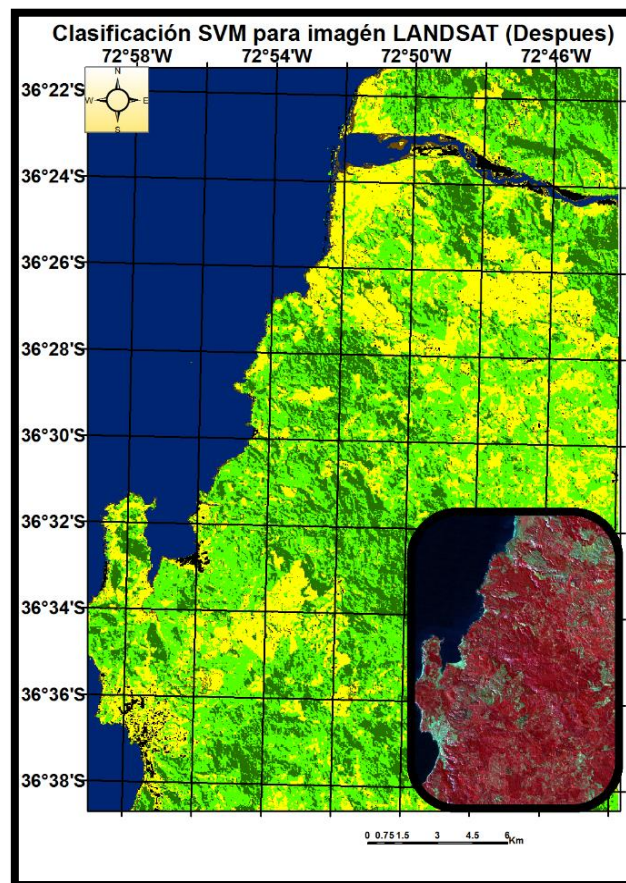
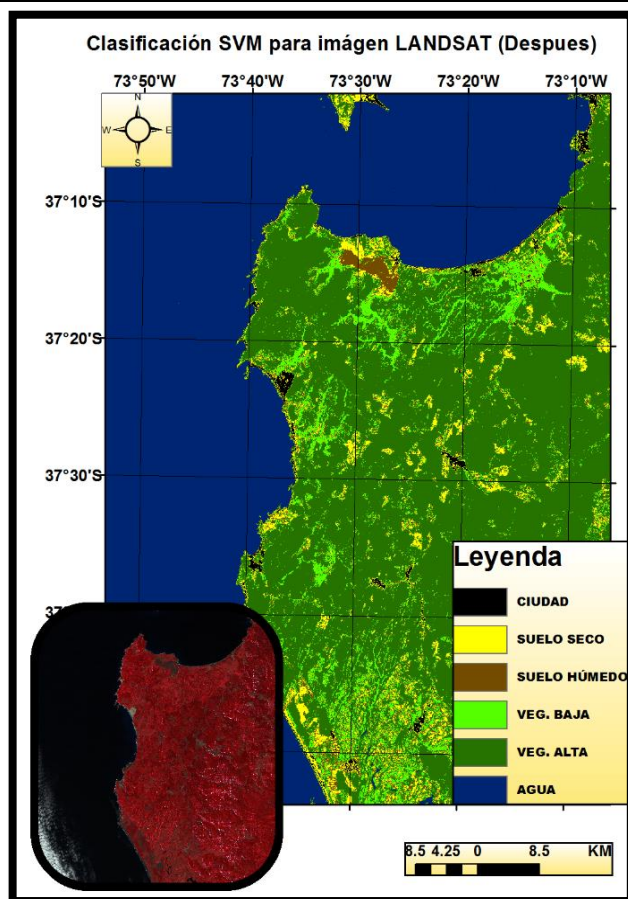


Figura 4.3- - Mapas tematicos a partir de las composiciones de bandas de las imágenes satelitales, usando la clasificación del modelo SVM, para periodo posterior al 27 de febrero del 2010.

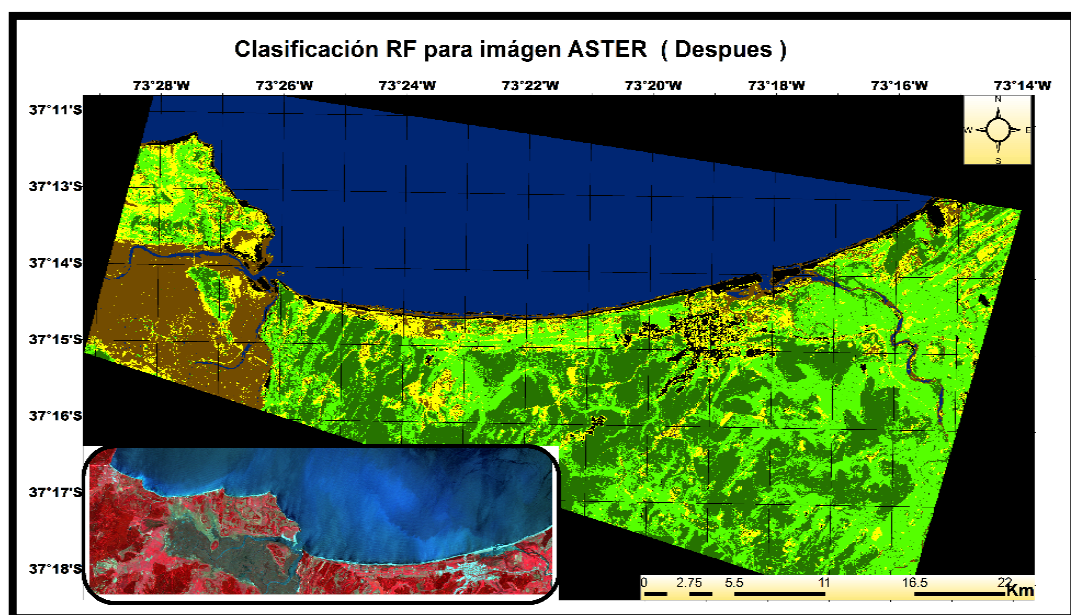
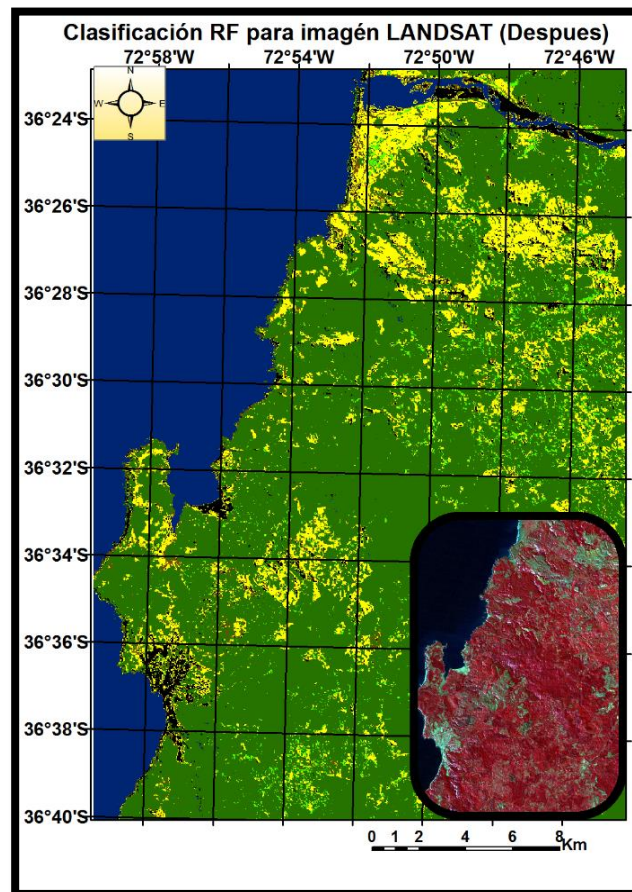
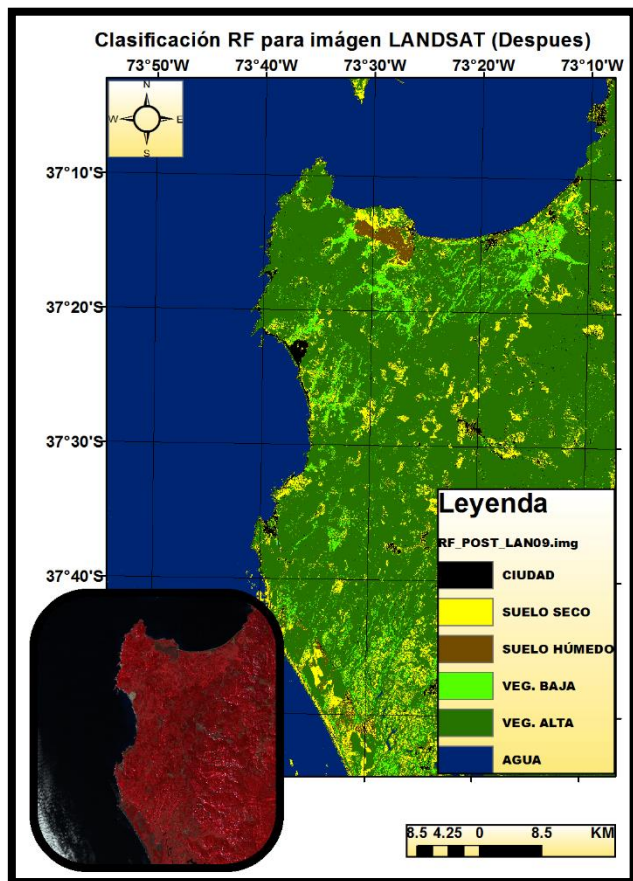


Figura 4.4- Mapas temáticos a partir de las composiciones de bandas de las imágenes satelitales, usando la clasificación del modelo RF, para periodo posterior al 27 de febrero del 2010.

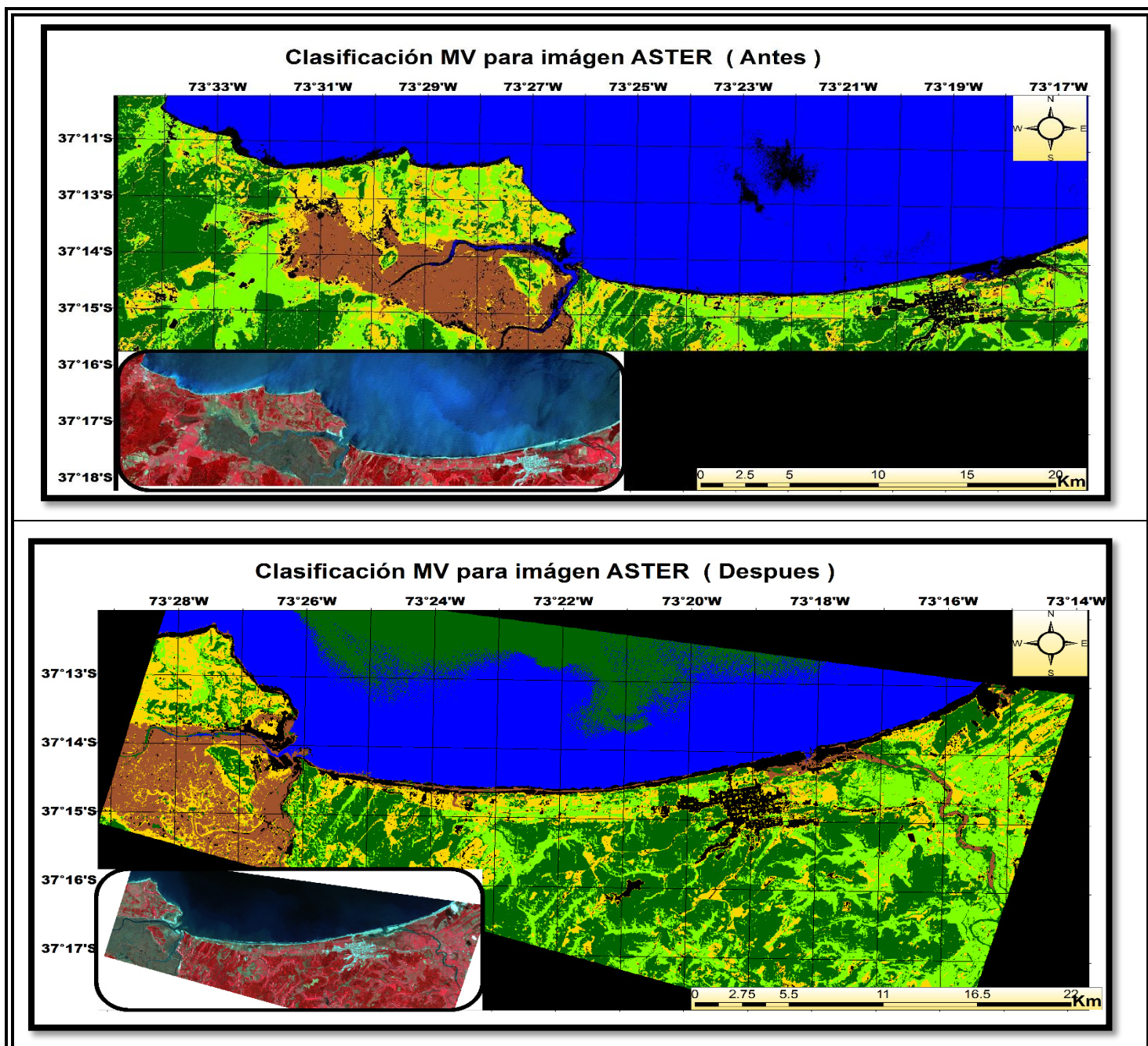


Figura 4.5- Mapas temáticos a partir de las composiciones de bandas de las imágenes satelitales ASTER, usando la clasificación del modelo MV, para los periodos previo y posterior al 27 de febrero del 2010.

5- Discusión

Para la clasificación de las imágenes los tres algoritmos muestran ser muy óptimos, sobre todo los algoritmos de aprendizaje SVM y RF.

El SVM, desde la perspectiva algorítmica, se tiene controversia en cuanto al núcleo a utilizar y la selección de parámetros, lo que trae consecuencias en la precisión. En este caso la elección del núcleo de “base radial” y la utilización de un optimizador de parámetros muestra tener muy buenos resultados. Esto viene a sumarse a los múltiples trabajos que muestran que hay evidencia empírica para apoyar la formulación teórica y la motivación detrás de SVM ([Friedl et. al., 2002](#); [Sesnie et. al., 2008](#); [Ghimere et. al., 2010](#)).

La característica más importante es la capacidad de SVM para generalizar bien a partir de una cantidad limitada y/o una mala calidad de datos de entrenamiento, como se ve reflejado en los altos niveles de confiabilidad global entregado por los modelos, a pesar de que algunos conjuntos de datos de entrenamiento eran más limitados y los pixeles representativos para cada clase no eran necesariamente los más representativos para cada categoría de suelo.

La capacidad de obtener una gran precisión de SVM, pese a las limitantes previamente descritas, está en la línea con el concepto de “vector de soporte” que se basa solo en unos pocos puntos de datos para definir el hiperplano del clasificador, siendo este proceso computacionalmente más livianos que otros métodos.

SVM son resistentes a quedar atrapados en mínimos locales debido a la convexidad de la función “coste” (cost) que permite al clasificador identificar constantemente la solución óptima, lo que se ve reflejado en la buena y consistente discriminación de coberturas similares que podrían llevar a errores de asignación y que SVM permite evitar. Una ventaja adicional es que no hay necesidad de repetir el entrenamiento del clasificador utilizando diferentes inicializaciones o arquitecturas aleatorias.

Además, SVM no asume una distribución estadística conocida de los datos a clasificar. Esto es particularmente útil dado que los datos adquiridos a partir de imágenes de detección remota generalmente tienen una distribución desconocida y la normalidad no siempre da una suposición correcta de la distribución real de los pixeles en cada clase.

Todas las características descritas de SVM permiten que supere a las técnicas como los de Máxima Verosimilitud. Evidenciándose esta diferencia de exactitud en los resultados obtenidos, mostrándose un índice kappa promedio de 0,96 para el algoritmo SVM, muy superior al 0,86 obtenido para el algoritmo MV.

Por otra parte Random Forest genera un gran interés, ya que consiste en el “aprendizaje conjunto”, método que generan muchos clasificadores, comparando sus resultados.

La técnica de dividir cada nodo utilizando el mejor ajuste entre un subconjunto de predictores elegidos al azar en ese nodo, dio muy buenos resultados para este caso, entregando una precisión comparable al de SVM.

Además es muy fácil de utilizar, ya que solo tiene dos parámetros (el número de variables en el subconjunto aleatorio de cada nodo y el número de árboles en el conjunto de decisión) y no es muy sensible a sus valores ([Galiano V., et. al., 2012](#)). El algoritmo RF tiene como ventaja, la generación de una estimación interna no sesgada

del error de generalización (error obb) por lo que no implicaría necesariamente utilizar un subconjunto de prueba independiente del de entrenamiento o recurrir a la validación cruzada.

Por otra parte, el algoritmo permite calcular la importancia de las variables (bandas) para la clasificación general de las categorías de cobertura de suelo y la clasificación de cada categoría mediante el índice de Ginni y la estimación de obb. Esta estimación podría ser importante para la selección de variables en la clasificación de áreas complejas en las que es obligatorio utilizar grandes conjuntos con un gran número de variables, pero para este caso el número de bandas era bastante reducido (6 para LandSat y 3 para ASTER), por lo que la reducción de éstas no se hace necesario.

RF es relativamente robusto a la reducción del tamaño del conjunto de entrenamiento. Los resultados demuestran que es superior a los enfoques de clasificación estándar como MV, ya que permitió una mejor clasificación entre las diferentes categorías del área de estudio, teniendo unas exactitudes kappa de 0,95 y 0,86 para RF y MV respectivamente.

De los 3 métodos analizados, los algoritmos SVM y RF entregan una precisión de clasificación muy elevada (superior al 95%) y por sobre el entregado por el método más ampliamente conocido y utilizado en Percepción Remota, como lo es el de Máxima Verosimilitud. Por lo cual, se escogerá el método SVM para analizar el antes y después de las coberturas de suelo en las localidades más afectadas del interés de este trabajo y así poder graficar los efectos geográficos que tuvo el terremoto y tsunami de 2010, mediante una comparación de los mapas temáticos antes y después. Estos mapas estarán representados en las figuras 5.1., 5.2., 5.3. y 5.4., para las localidades de Dichato, Llico, Arauco y Tubul, respectivamente.

Con el fin de tener un mejor enfoque en los cambios de cobertura de suelo de las áreas afectadas, apoyaremos los resultados obtenidos a través de los procesos de clasificación con los mapas de áreas de inundación producto del tsunami del 27 de febrero de 2010, elaborados por el Servicio Nacional de Geología y Minería. A su vez se respaldarán los resultados con imágenes de terreno, que den cuenta, entre otras cosas, de los cambios costeros y en cuencas de ríos producidos por alzamientos cosísmicos.

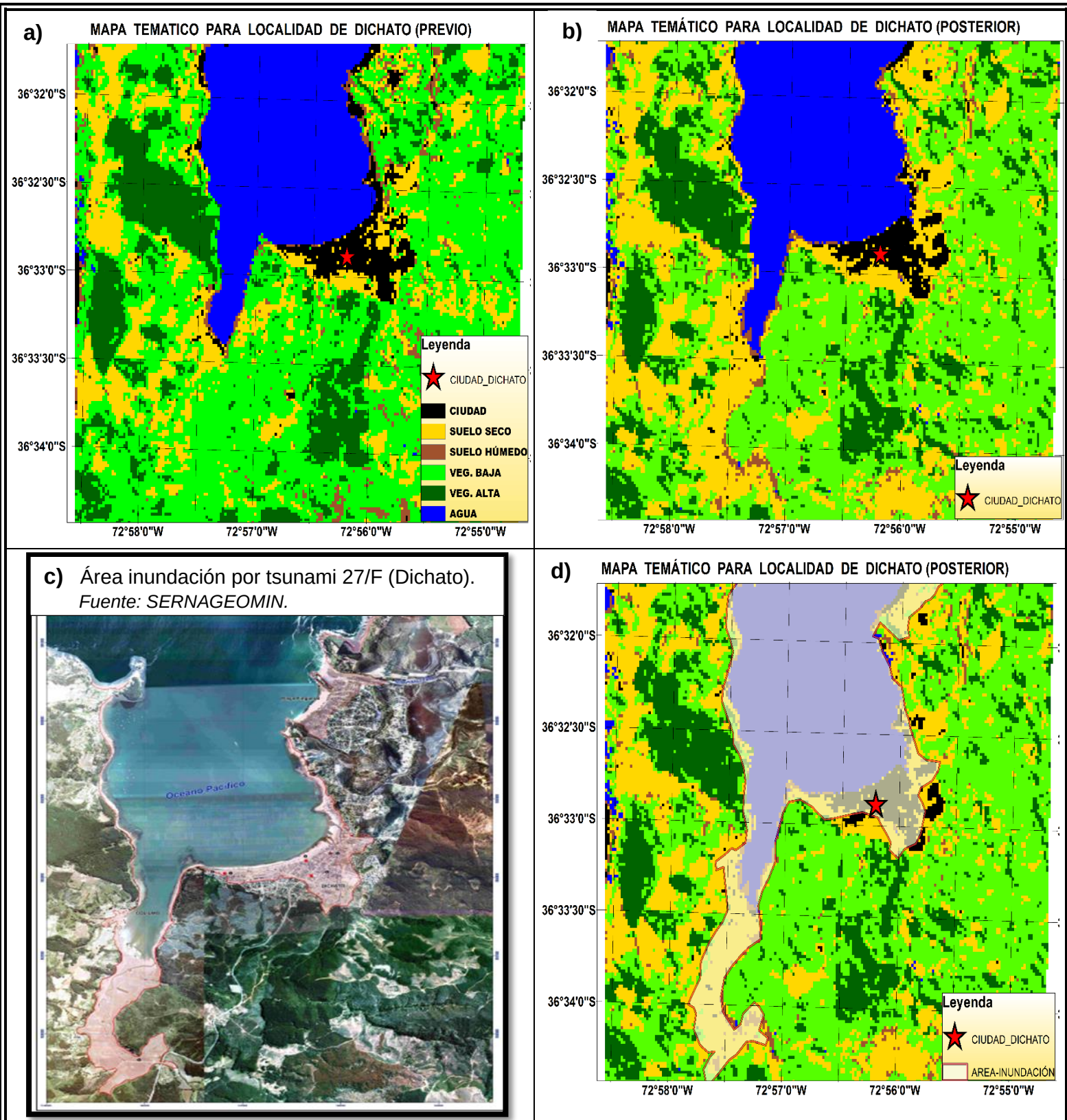


Figura 5.1- Mapas temáticos para localidad de Dichato y área de inundación por Tsunami del 27 de febrero del 2010:

a) Mapa temático de Dichato, para periodo previo al 27/F.

b) Mapa temático de Dichato, para periodo posterior al 27/F.

c) Mapa del área de inundación por tsunami del 27/F (SERNAGEOMIN).

d) Área de inundación sobre mapa temático de Dichato post 27/F, con el propósito de graficar la relación entre el alcance del tsunami tierra adentro y el cambio de cobertura de suelo en el área directamente afectado.

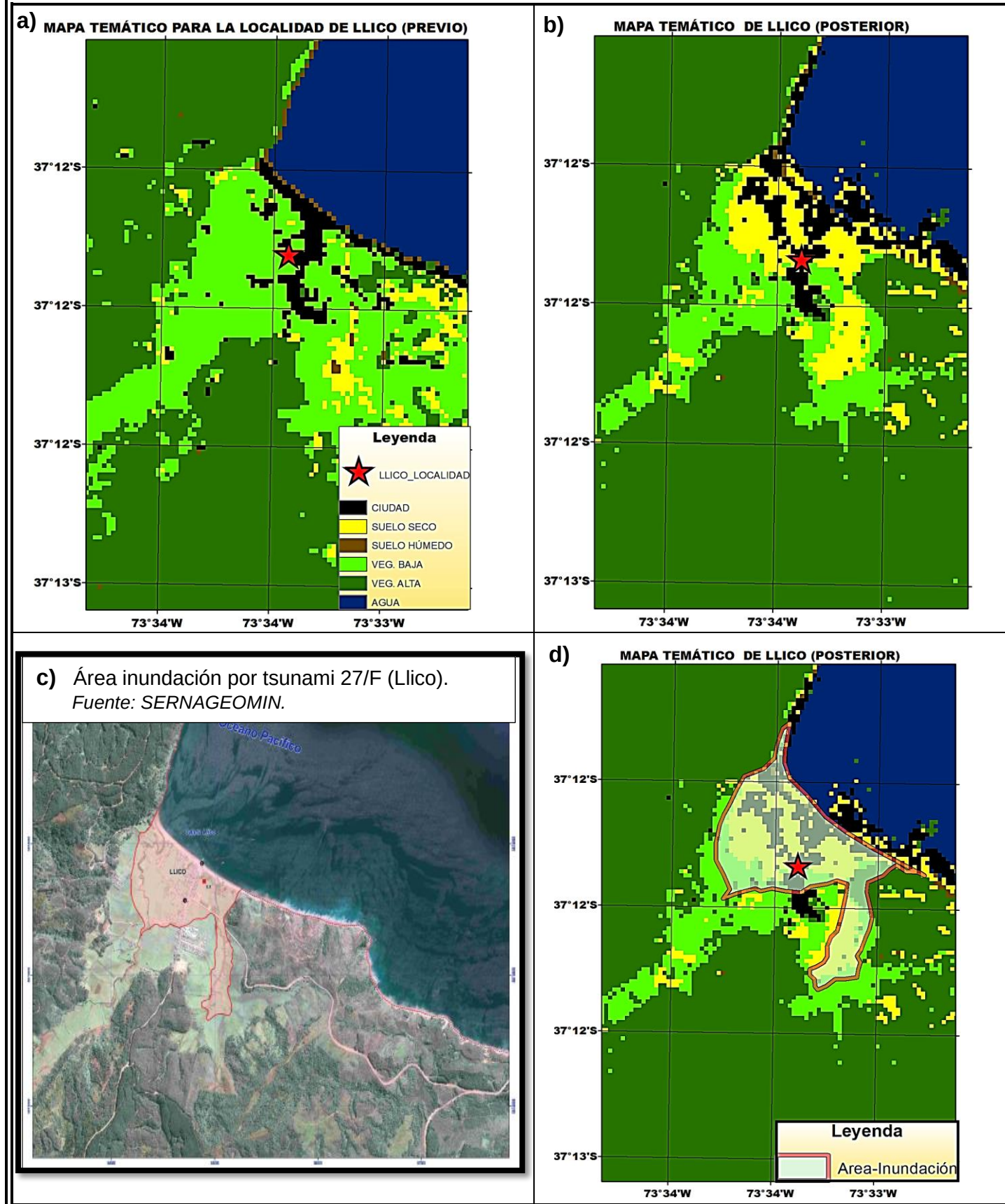
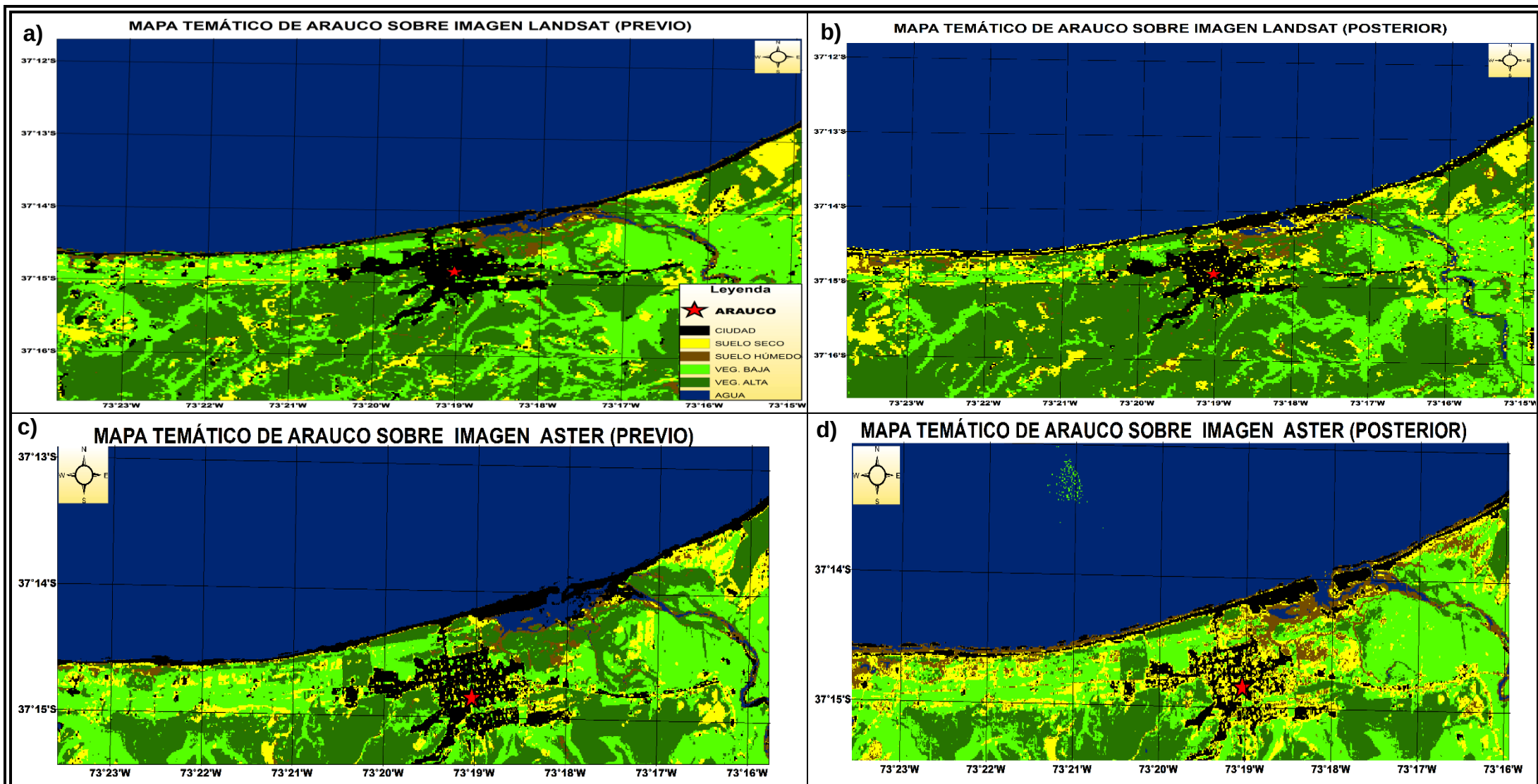


Figura 5.2- Mapas temáticos para localidad de Llico y área de inundación por Tsunami del 27 de febrero del 2010:

- a) Mapa temático de Llico, para periodo previo al 27/F.
- b) Mapa temático de Llico, para periodo posterior al 27/F.
- c) Mapa del área de inundación por tsunami del 27/F (SERNAGEOMIN).
- d) Área de inundación sobre mapa temático de Llico post 27/F, con el propósito de graficar la relación entre el alcance del tsunami tierra adentro y el cambio de cobertura de suelo en el área directamente afectado.



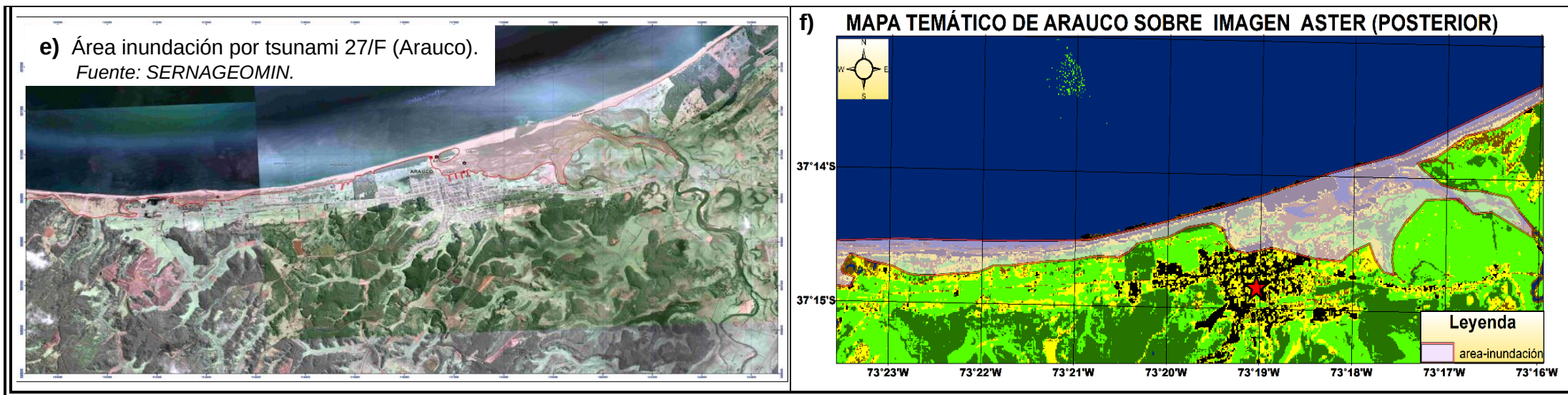
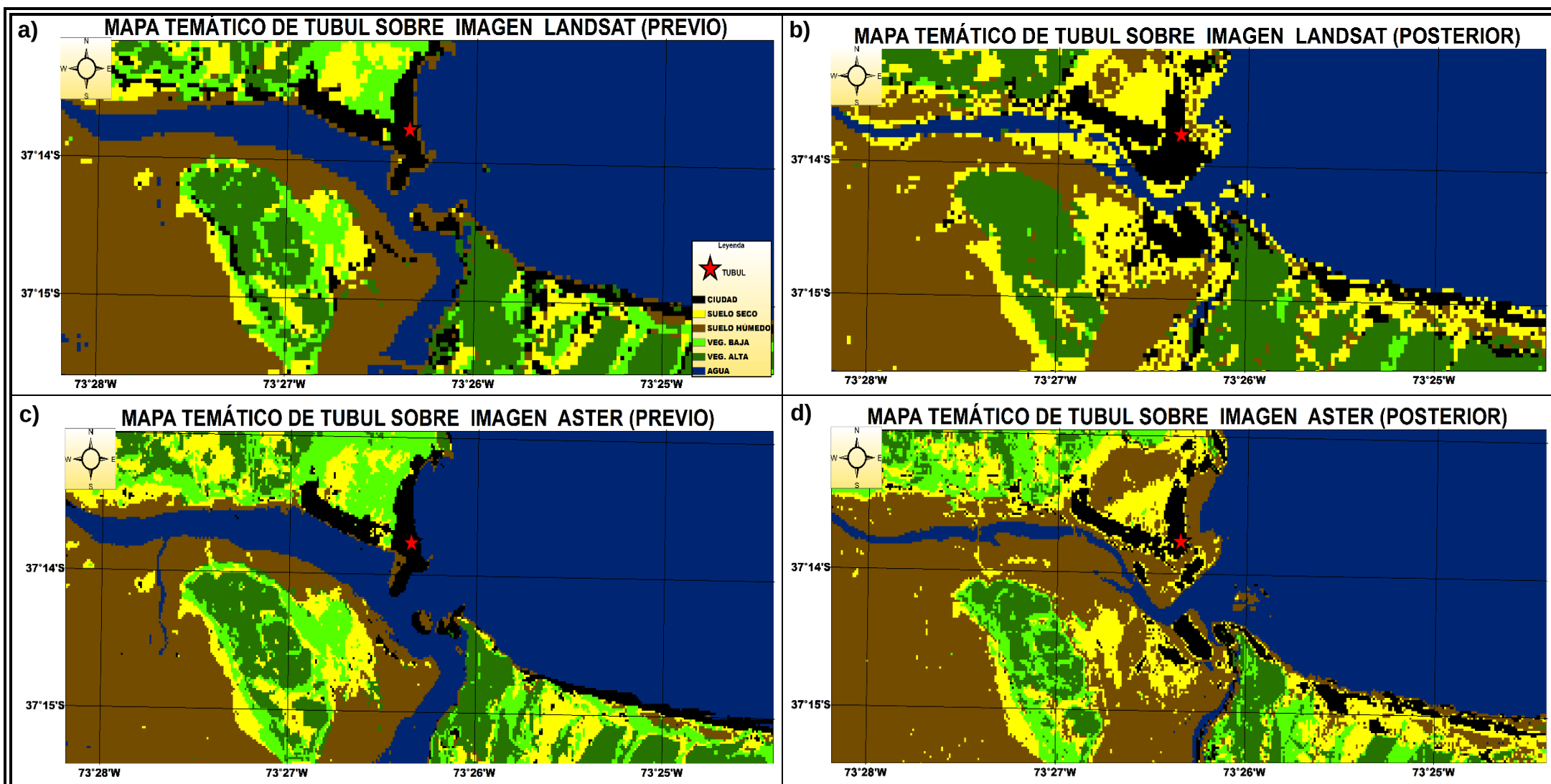


Figura 5.3- Mapas temáticos para localidad de Arauco y área de inundación por Tsunami del 27 de febrero del 2010:

- a) Mapa temático de Arauco, a partir de imagen LandSat, para periodo previo al 27/F.
- b) Mapa temático de Arauco, a partir de imagen LandSat, para periodo posterior al 27/F.
- c) Mapa temático de Arauco, a partir de imagen ASTER, para periodo previo al 27/F.
- d) Mapa temático de Arauco, a partir de imagen ASTER, para periodo posterior al 27/F.
- e) Mapa del área de inundación por tsunami del 27/F (SERNAGEOMIN).
- f) Área de inundación sobre mapa temático de Arauco post 27/F, con el propósito de graficar la relación entre el alcance del tsunami tierra adentro y el cambio de cobertura de suelo en el área directamente afectado.



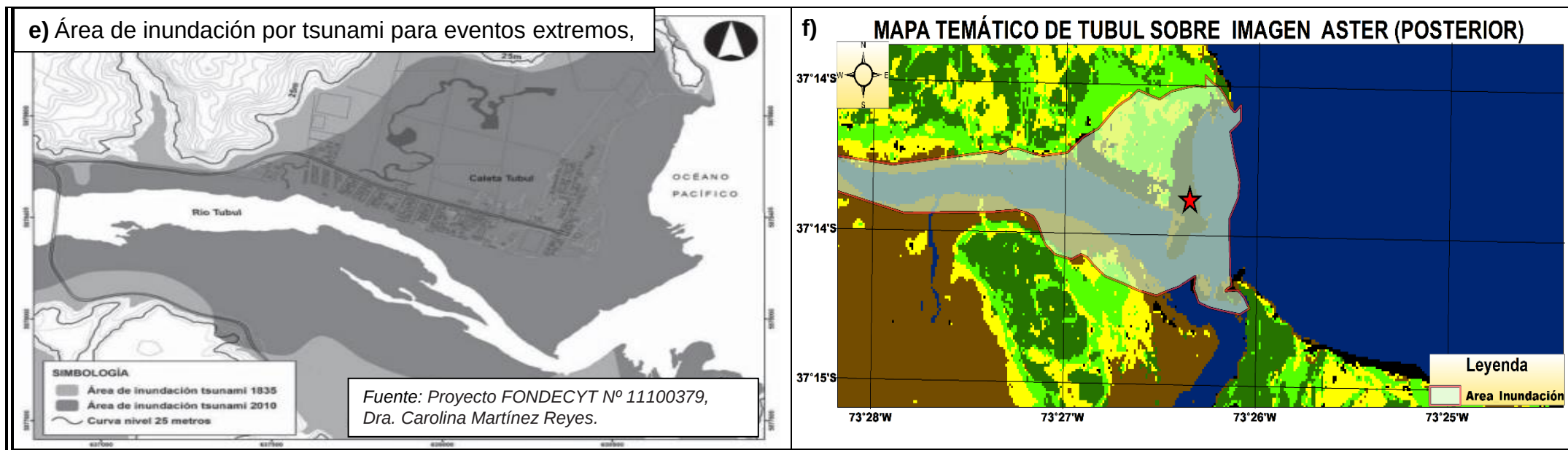


Figura 5.4- Mapas temáticos para localidad de Tubul y área de inundación por Tsunami del 27 de febrero del 2010:

- Mapa temático de Tubul, a partir de imagen LandSat, para periodo previo al 27/F.
- Mapa temático de Tubul, a partir de imagen LandSat, para periodo posterior al 27/F.
- Mapa temático de Tubul, a partir de imagen ASTER, para periodo previo al 27/F.
- Mapa temático de Tubul, a partir de imagen ASTER, para periodo posterior al 27/F.
- Mapa del área de inundación por tsunami para eventos extremos (Dra. Carolina Martínez).
- Área de inundación sobre mapa temático de Tubul post 27/F, con el propósito de graficar la relación entre el alcance del tsunami tierra adentro y el cambio de cobertura de suelo en el área directamente afectado.

Previo al terremoto, el aumento del nivel del mar era notorio en las localidades, evidenciando la subsidencia intersísmica. Estos efectos obligo a construir muros de protección en localidades de Tubul y Llico, ya que el mar estaba llegando hasta las viviendas y caminos más próximos al litoral, ocurriendo daños importantes durante las marejadas (Martínez C., 2011). En las décadas previas al terremoto del 2010, la subsidencia intersísmica produjo a su vez el ensanchamiento de ríos en la zona. Luego el alzamiento cosísmico, produjo desecamiento de ríos. En Tubul, se produjo un retroceso de 200 m del mar debido a su baja pendiente, así como en Llico, quedando una playa (Quezada et. al., 2012).

El tsunami del 27/F en la costa de la región del Biobío, generó alturas de ola entre 3 y 30 metros. El ingreso del tsunami a la costa fue favorecida por la presencia de cursos de agua locales como lo fue el río Carampangue en la localidad de Arauco y los ríos Tubul y Raqui en la localidad de Tubul. Los efectos geomorfológicos principales se expresan en el cambio en el nivel base de los ríos (Tubul), donde la costa experimentó alzamientos de hasta 2 metros (Quezada et.al., 2010). El litoral de la región del Biobío tuvo alzamientos co-sísmicos que generó el ensanchamientos de playas, la emersión de la plataforma de abrasión marina, desecamiento de ríos y humedales. El tsunami arrastró arenas litorales y vegetación hacia el interior. La morfología resultante del impacto directo del tsunami resultó ser principalmente erosiva (Martínez C., 2011).

En general existen cambios notorios en las coberturas de suelo para todas las áreas estudiadas, donde los mayores cambios se produjeron en la disminución de la cobertura vegetal costera, para dar paso a una cobertura sedimentaria (ver Figuras 5.1.a-b; 5.2.a-b). Este cambio impulsado principalmente por los efectos del tsunami que logro invadir amplias zonas, arrastrando consigo una gran cantidad de material como arena y escombros de todo tipo, además de la misma agua salada que afecta en gran medida a la vegetación.

Además el retroceso de la costa mar adentro en algunas zonas es evidente al ver el cambio de cobertura de suelo, al pasar de agua a una cobertura de suelo que en algunos casos como el de Tubul comprende una gran porción de terreno (ver Figuras 5.2.a-b; 5.4.c-d).

En promedio para nuestras 4 localidades y áreas circundantes (*mostradas entre figuras 5.1 a 5.4*), los cambios más significativos fueron el aumento de las coberturas sedimentarias (sedimento seco aumento un 7,4% y el sedimento húmedo un 6,6%) y la disminución de la cobertura vegetal (vegetación baja y alta) de un 8,2% promedio (ver *tablas 5.1-5.2-5.3-5.4*).

A continuación se presentan la descripción de las coberturas de suelo para cada una de las localidades, con un enfoque comparativo del antes y después del terremoto y tsunami. Para este objetivo, se muestran las tablas 5.1. a 5.4., donde para cada categoría de suelo se despliegan los números de píxeles clasificados en cada una de las categorías para el periodo previo al evento en la segunda columna (Nº píxeles (PRE)) y para el periodo posterior al evento en la tercera columna (Nº píxeles (POST)). También se mostraran los porcentajes que corresponden los números de píxeles seleccionados para cada cobertura con respecto al total de píxeles en la imagen, donde la cuarta columna (%Clase(PRE)) corresponderá al porcentaje de píxeles pertenecientes a las coberturas previo al evento y la quinta columna (%Clase(POST)) para el periodo posterior a este. Finalmente en la última columna (%POST-%PRE) se

representaran las diferencias en términos de porcentaje de los pixeles seleccionados para cada categoría, comparando los dos periodos de estudios, lo que nos proporciona un valor para discriminar si existió un aumento o disminución de cada categoría debido a los efectos del terremoto y tsunami.

Para un mejor análisis de los cambios ocurridos en cada localidad, se representarán por una parte, imágenes comparativas de las líneas de costa previa al tsunami sobre los mapas temáticos posteriores al evento, con el objetivo de visualizar la existencia de ensanchamientos del borde costero o acumulación de escombros en este. A sus ves se representaran las imágenes de los mapas temáticos de cada localidad, enmarcando las áreas especificadas donde se detectaron los mayores cambios de cobertura de suelo. Todas estas imágenes se encuentran en las figuras 5.5. a 5.8.

- Ha sido ampliamente demostrado y reconocido, tanto por catastros como estudios realizados en terreno ([Quezada et.al., 2010](#); [Martínez C. et.al., 2011](#)), que el tsunami afecto a la ciudad de Dichato prácticamente en toda su extensión, alcanzando una máxima distancia de propagación de hasta 1 km dentro de la ciudad y de 2,7 km tierra adentro entrando por la bahía de Coliumo, arrasando con toda la cobertura vegetal del área. La vegetación costera también se vio fuertemente afectada. Todo esto produjo una disminución cercana al 6% de la vegetación. Cambios en el suelo urbano no es mayormente apreciable en el mapa temático (figura 5.5), siendo percibido mayormente en la zona costera, lo que produjo una disminución de un 5% a un 4% en esta clase temática, dando paso a material sedimentario (*ver figura 5.5.b*). En la línea de costa no se aprecia que exista algún tipo de avance o retroceso significativo, por lo que descartaría algún tipo de alzamiento o hundimiento co-sísmico en la zona (*ver figura 5.5.a*).

LOCALIDAD DICHATO					
CATEGORIA	Nº pixeles (PRE)	Nº pixeles (POST)	% Clase (PRE)	% Clase (POST)	%POST - %PRE
CIUDAD	916	718	5.0%	3.9%	-1.1%
SUELO SECO	2473	3787	13.5%	20.7%	7.2%
SUELO HUMEDO	474	450	2.6%	2.5%	-0.1%
VEG. BAJA	8148	6987	44.6%	38.2%	-6.3%
VEG. ALTA	2142	2168	11.7%	11.9%	0.1%
AGUA	4136	4179	22.6%	22.8%	0.2%
TOTAL	18289	18289	100.0%	100.0%	0.0%

Tabla 5.1- Tabla comparativa de cambios en las clases temáticas previo y posterior al terremoto del 2010 para la localidad de Dichato.

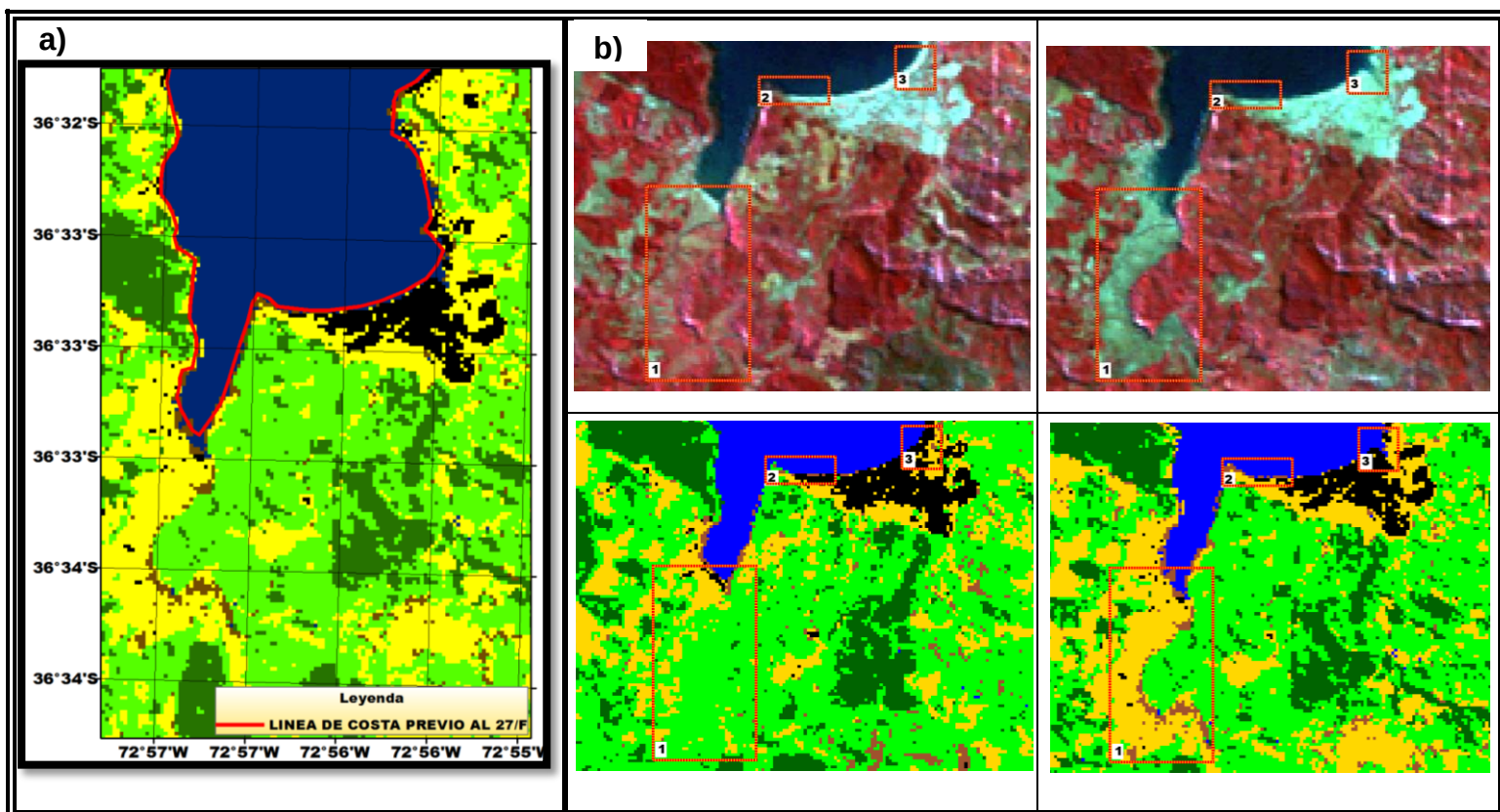


Figura 5.5- Cambios en cobertura de suelo y morfológicos en la localidad de Dichato:

- a) Se contrasta la línea de costa previa al 27/F con el mapa temático posterior al evento, comprobando que no existió un avance o retroceso significativo en cuanto a la costa en Dichato.
- b) Cuadros de las composiciones de bandas y mapas temáticos previo y posterior al terremoto y tsunami. Debido a la conformación geográfica en donde está situada la localidad, produjo una fuerte entrada de ola. Se visualizan algunas pequeñas entradas de mar, marcadas en la parte superior del rectángulo rojo nº1 y en el centro de rectángulo nº3. También se evidencian en los depósitos sedimentarios que se muestran en el recuadro nº1 donde el tsunami tuvo un gran alcance tierra adentro. El borde costero de la ciudad de Dichato evidencia los efectos del evento en los recuadros 2 y 3 donde suelo urbano y vegetación casi desaparecen.

- Llico fue una localidad muy afectada por el evento del 2010, con una altura de ola de hasta 10 metros y una distancia de propagación de 800 m tierra adentro (Martínez C. et.al., 2011). La vegetación de la zona fue fuertemente afectada, dando paso a roca expuesta y sedimentos arenosos, siendo fácilmente apreciable en los mapas temáticos de cambio de cobertura de suelo (figura 5.6). Se produjo una disminución de un 7% en la vegetación y un aumento en el 10% en la cobertura del sedimento (ver figura 5.6.b). Otro hecho importante es el cambio en la línea de costa que avanzó mar adentro hasta unos 100 metros, que deja una playa de arena (ver figura 5.6.a). Esto se justifica producto el alzamiento co-sísmico sufrido en la zona, el cual según estudios, fue de $1,9 \pm 0,2$ metros (Quezada et.al., 2010).

LOCALIDAD LLICO					
CATEGORIAS	Nº pixeles (PRE)	Nº pixeles (POST)	% Clase (PRE)	% Clase (POST)	%POST - %PRE
CIUDAD	350	370	9.0%	9.5%	0.5%
SUELO SECO	325	728	8.3%	18.7%	10.3%
SUELO HUMEDO	59	34	1.5%	0.9%	-0.6%
VEG. BAJA	1949	865	49.9%	22.2%	-27.8%
VEG. ALTA	1047	1856	26.8%	47.6%	20.7%
AGUA	172	49	4.4%	1.3%	-3.2%
TOTAL	3902	3902	1	1	0

Tabla 5.2- Tabla comparativa de cambios en las clases temáticas previo y posterior al terremoto del 2010 para la localidad de Llico.

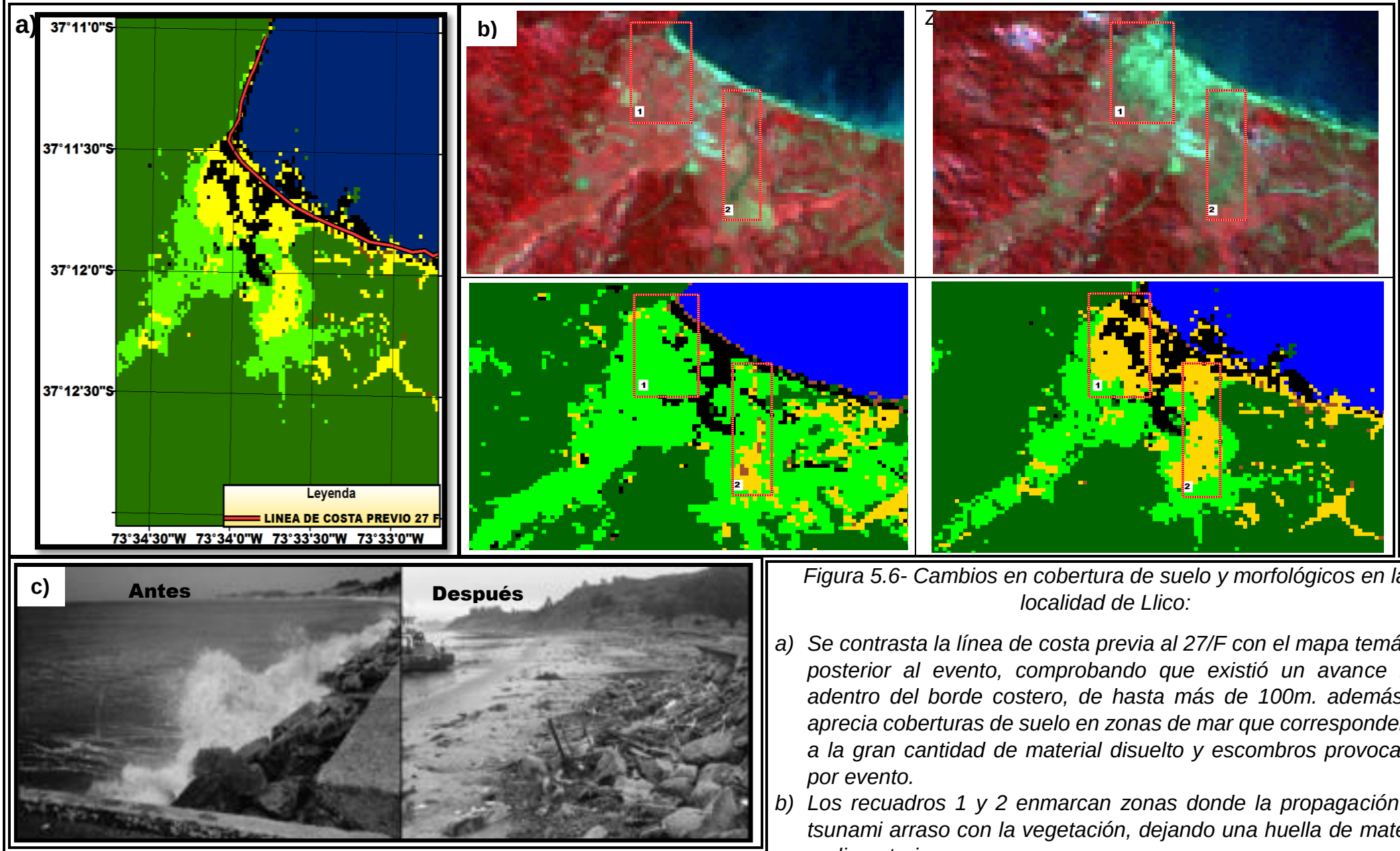


Figura 5.6- Cambios en cobertura de suelo y morfológicos en la localidad de Llico:

- a) Se contrasta la línea de costa previa al 27/F con el mapa temático posterior al evento, comprobando que existió un avance mar adentro del borde costero, de hasta más de 100m. además se aprecia coberturas de suelo en zonas de mar que corresponderían a la gran cantidad de material disuelto y escombros provocados por evento.
- b) Los recuadros 1 y 2 enmarcan zonas donde la propagación del tsunami arrasó con la vegetación, dejando una huella de material sedimentario.

c) Imágenes en terreno antes y después del 27/F, que evidencian el ensanchamiento de la costa y la gran cantidad de escombros vistos en los mapas temáticos. (Quezada, 2012).

- La ciudad de Arauco no presenta grandes daños comparado con otras localidades ubicadas en el golfo de Arauco, debido principalmente al efecto mitigador que tuvo el estuario del río Carampangue y la protección del cordón dunario que antecede al casco urbano. Por otra parte la altura de ola en esta zona superó los 2 metros y una distancia máxima de propagación de 1500 metros (Quezada et.al., 2010), alcanzando a áreas urbanas de la ciudad de Arauco. La zona del estuario es la que se percibe con los mayores cambios, pasando de una cobertura de agua a una cobertura de suelo dominada por sedimentación húmeda, lo que se evidencia en el aumento de esta última cobertura en casi un 7% (ver figura 5.7.b). La vegetación igualmente se vio disminuida en un 8%, dando paso a suelo sedimentario. La línea de costa no registra algún tipo de avance o retroceso evidente (ver figura 5.7.a).

LOCALIDAD ARAUCO					
CATEGORIA	Nº pixeles (PRE)	Nº pixeles (POST)	% Clase (PRE)	% Clase (POST)	%POST - %PRE
CIUDAD	13038	9491	20.7%	15.0%	-5.6%
SUELO SECO	6220	11475	9.9%	18.2%	8.3%
SUELO HUMEDO	2539	6809	4.0%	10.8%	6.8%
VEG. BAJA	17094	21222	27.1%	19.4%	-7.7%
VEG. ALTA	21427	12214	34.0%	33.6%	-0.4%
AGUA	2750	1857	4.4%	2.9%	-1.4%
TOTAL	63068	63068	100.0%	100.0%	0.0%

Tabla 5.3- Tabla comparativa de cambios en las clases temáticas previo y posterior al terremoto del 2010 para la localidad de Arauco.

- En Tubul se produjo un retroceso de 200 metros del mar, debido principalmente a la baja pendiente de la zona que se ve fuertemente afectado por el alzamiento co-sísmico de $1,4 \pm 0,1$ metros (Quezada et.al., 2010), formando una playa de arena (ver figura 5.8.c). Analizando la línea de costa previa al terremoto sobre el mapa temático posterior al evento, se vislumbra que la línea de costa retrocediera unos 200 metros aproximadamente (ver figura 5.8.a). Otro efecto claro es el desecamiento de los ríos Tubul y Raqui. Todos estos retrocesos de cobertura de agua, fue reemplazada mayormente por la clase temática de sedimento húmedo, registrándose un aumento de más del 20% de esta clase, en desmedro de la clase temática de agua y vegetación. A su vez apareció coberturas de roca expuesta, producto a deslizamientos en zonas escarpadas y traslado de rocas en áreas costeras (ver figura 5.8.b).

Localidad Tubul					
Clase	Nº pixeles (PRE)	Nº pixeles (POST)	% Clase (PRE)	% Clase (POST)	%POST - %PRE
CIUDAD	1412	1817	5.9%	7.6%	1.7%
SUELO SECO	3123	4054	13.0%	16.9%	3.9%
SUELO HUMEDO	7215	12093	30.0%	50.3%	20.3%
VEG. BAJA	4007	2561	16.7%	10.6%	-6.0%
VEG. ALTA	2922	1646	12.1%	6.8%	-5.3%
AGUA	5373	1884	22.3%	7.8%	-14.5%
TOTAL	24055	24055	100.0%	100.0%	0%

Tabla 5.4- Tabla comparativa de cambios en las clases temáticas previo y posterior al terremoto del 2010 para la localidad de Tubul.

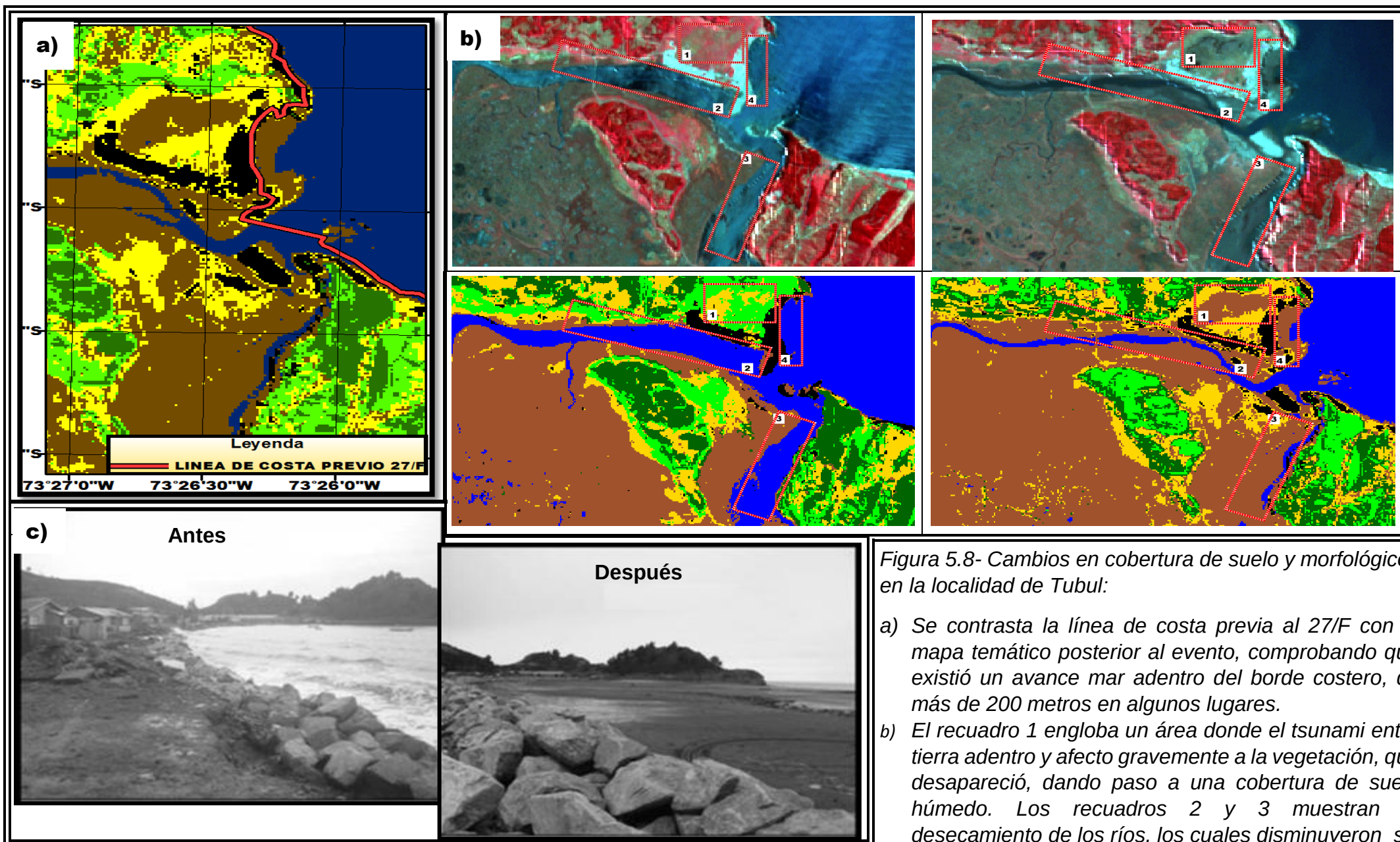


Figura 5.8- Cambios en cobertura de suelo y morfológicos en la localidad de Tubul:

- a) Se contrasta la línea de costa previa al 27/F con el mapa temático posterior al evento, comprobando que existió un avance mar adentro del borde costero, de más de 200 metros en algunos lugares.
- b) El recuadro 1 engloba un área donde el tsunami entro tierra adentro y afecto gravemente a la vegetación, que desapareció, dando paso a una cobertura de suelo húmedo. Los recuadros 2 y 3 muestran el desecamiento de los ríos, los cuales disminuyeron su

cuenca de manera notoria en los mapas temáticos, que pasan del color azul de la clase agua a ser dominados por el café del sedimento húmedo. El recuadro 4 muestra la aparición de una gran área de sedimento mar adentro, evidenciando un ensanchamiento de la playa frente a Tubul.

c) Imágenes en terreno antes y después del 27/F, que evidencian el ensanchamiento de la costa, conformándose una gran área de arena frente a Tubul. (Quezada, 2012)

6- Conclusión:

Este estudio tuvo como uno de sus objetivos principales evaluar el desempeño de los clasificadores de algoritmos de aprendizaje SVM y RF y compararlos con el método MV, el más utilizado en percepción remota a lo largo de los años. Todo esto sobre la clasificación de coberturas de suelo en áreas heterogéneas, como lo son las localidades de estudio. Mediante la incorporación de un conjunto de datos LandSat y ASTER multitemporales y el uso de un DEM, los algoritmos de aprendizaje se desempeñaron bien en el contexto de clasificación con 6 categorías, espectralmente bastantes diferenciables entre sí. Teniendo como posibles fuentes de disminución en la precisión de clasificación la elección no optima de parámetros ingresados por el usuario, necesarios para los algoritmos y por otra parte la reducción de los datos de entrenamiento. Los resultados de esta investigación proporcionaron nuevas perspectivas sobre el desempeño de algoritmos SVM y RF en el contexto de la cartografía de coberturas de suelo en áreas extensas y heterogéneas.

A su vez se suma a otras investigaciones que demuestran la superioridad de los algoritmos de aprendizaje frente a los métodos más tradicionales, estableciéndolos como la mejor opción para la clasificación de coberturas de suelo en zonas heterogéneas ([Melgani y Bruzzone, 2004](#); [Marcal et.al., 2005](#); [Dalponte et.al., 2009](#)).

En cuanto a los cambios visualizados en cobertura de suelo y análisis de efectos producto al terremoto y tsunami, se obtienen resultados comparables a los obtenidos en los estudios de similares características. El estudio realizado sobre detección de cambios inducidos por tsunamis desde imágenes satelitales de alta resolución para las costas de Sumatra, donde clasificaciones mediante algoritmos de máxima verosimilitud, detectaron regiones específicas de cambio de cobertura de suelo ([Sumer E. y Celebi F., 2006](#)). En este mismo sentido, otro estudio ilustró mediante generación de mapas temáticos, los daños producidos por el tsunami en las costas de Sumatra, demostrando que la clasificación de imágenes satelitales son una fuente para evaluar los efectos de desastres naturales de manera rápida ([Chen P. et.al., 2004](#)).

Por otra parte este estudio da un enfoque a escala más local sobre estudios de cambios en la cobertura de suelo antes y después del evento del 2010, logrando captar cambios en áreas más acotadas, lo cual destaca frente a otros trabajos realizados para analizar este evento mediante generación de mapas temáticos, que se enfocaron en una escala más regional ([Rojas C. et.al., 2013](#)). A su vez los resultados obtenidos concuerdan con los mostrados mediante otros tipos de análisis geográficos y morfológicos, los cuales en su mayoría realizan catastros en terreno para la cuantificación de cambios en las áreas afectadas ([Martinez C. et.al., 2011](#); [Quezada et.al., 2012](#)), los que concuerdan en el alzamiento co-sísmico que produjo ensanchamientos de playas y desecamiento de ríos, además de la entrada de material sedimentario, arrastrado por el tsunami, invadiendo zonas con una cobertura vegetal y también en zonas urbanas.

Analizando la línea de riesgos naturales, tales como estudios de desastres naturales, en Chile esta línea de trabajo es episódica, con interés particular en casos emblemáticos que han generado impacto mediático. Es de suma importancia investigar sobre escenarios de cambio a escala local y proponer alternativas de adaptación con el propósito de evitar consecuencias negativas, sobre todo en ambientes costeros y en sociedades donde los procesos productivos y desarrollo económico dependen fundamentalmente de los recursos naturales.

El primer paso es por lo tanto demostrar que existen cambios en los procesos naturales a escala de tiempo pre-definidas, que compromete el uso del territorio en sus distintas dimensiones y su proyección en el tiempo ([Rojas C. et.al., 2013](#)). En este aspecto existe mucho por investigar dada la escasa sistematización del conocimiento sobre procesos físico-naturales de la costa chilena. En este sentido, el terremoto y tsunami del 27 de febrero de 2010 ha permitido conocer los procesos y cambios de cobertura de suelo, además de cambios morfológicos en la costa en relación a estos tipos de eventos, lo cual puede visualizarse y desarrollarse como una buena herramienta en la planificación territorial, como una mejor proyección en la infraestructura y localización de asentamientos costeros.

Llevado a la toma de decisiones, el problema final radica en cómo se incorporan estos elementos a instrumentos de planificación de lenta renovación (en algunas localidades inexistentes aún), con expansiones urbanas y urbanización más rápidas, incapaces de absorber los cambios y que redundan en una planificación urbana rígida e incapaz de incorporar y trabajar con escenarios de cambio.

El terremoto $M_w=8.8$, del 27 de febrero, provocó alzamientos y subsidencias co-sísmicas que modificaron el relieve litoral. Los alzamientos fueron evidentes en el litoral de la región del Biobío, por tener su litoral más cercano a la fosa. Estos alzamientos provocaron la emersión de la plataforma de abrasión marina, ensanchamiento de playas y desecamientos de ríos y humedales costeros, efectos que se evidenciaron en el contraste de líneas de costa antes y después del 27F en Tubul y Llico, y así mismos en el contraste del grosor de los ríos Tubul y Raquí. El tsunami por su parte, removió las áreas litorales y sub-litorales, siendo estas en algunas partes depositadas tierra adentro constituyendo depósitos del tsunami, afectando en gran medida a suelo que tenía previamente una cobertura vegetal, pasando a suelos dominado por sedimentos; en otras zonas las arenas sueltas fueron redepositadas durante las marejadas posteriores, reconstruyendo o ensanchando las playas, como lo ocurrido en Tubul y Llico, donde los mapas temáticos posterior al 27/F muestran una cobertura de suelo húmedo en remplazado de una cobertura antes dominada ocupada por agua. Además el tsunami utilizó los cursos locales de las pequeñas localidades costeras (principalmente bahías y estuarios) para propagar la energía al interior de estas, generando una mayor superficie inundada, como lo ocurrido en Arauco, donde el estuario del río Carampangue encauso el tsunami, recibiendo gran cantidad de material, que se vislumbra en el crecimiento de la cobertura sedimentaria en las zonas del estuario.

Nuestras localidades estudiadas ubicadas en el golfo de Arauco sufrieron una fuerza mayor del tsunami debido a como se explicó por [Martínez C. \(2011\)](#), la ubicación de la isla Santa María al interior del golfo de Arauco, generó procesos de difracción que tenderían a desviar el frente de onda al lado sur del golfo de Arauco con impacto directo a las localidades de Arauco, Tubul y Llico, lo cual coincide con las alturas de ola mayores de acuerdo con observaciones locales ([Quezada et.al., 2010](#)) que produjeron un gran cambio en la cobertura de suelo registrada por este estudio para el antes y después del evento.

Para las localidades analizadas de manera particular se tiene en Dichato, en términos generales, existió una disminución en las coberturas de ciudad y vegetación baja y, por otra parte, un fuerte crecimiento en la cobertura de sedimento seco, principalmente vinculada a materiales, escombros y arenas arrastradas por el tsunami y también a suelo expuesto producto de la muerte de la vegetación.

En Llico, en general, la superficie con cobertura urbana se mantuvo constante, debido presumiblemente a que los escombros y rocas arrastradas por la fuerza del agua provocaron que el suelo presentara coberturas urbanas o de alto albedo donde antes no estaba presente, compensando las pérdidas de este tipo de cobertura en el sector costero. También se produjo una importante disminución en la cobertura vegetal, contrastando con el aumento de la cobertura sedimentaria, lo que se explica debido a las arenas arrastradas por la ola hacia el interior, a su vez por la pérdida de vegetación que dio paso a suelo expuesto y por el retroceso de la línea de costa, proceso registrado en esta localidad y que dio paso a la exposición de suelo marino.

En Arauco, la zona del estuario del río Carampangue fue la que percibió los mayores cambios en cobertura, pasando a presentar después del evento una mayor cobertura de sedimento húmedo producido por el suelo pantanoso del mismo humedal que fue removido por la fuerza del tsunami y trasladado a otras partes en las que se encontraba presente otro tipo de coberturas como la de vegetación baja, siendo esta la que percibió la mayor disminución. También en Arauco se evidenció una pérdida en cobertura de ciudad, vinculada mayormente a las áreas urbanas más próximas al humedal, donde algunas construcciones fueron gravemente dañadas por el terremoto debido a estar construidas en terrenos de relleno sobre el sedimento pantanoso ([Belmonte A., 2013](#)). A esto se le suma que el área de inundación llegó hasta algunas zonas de cobertura de ciudad ([Rojas C. et. al., 2013](#)), arrastrando gran cantidad de material sedimentario del humedal y a su vez, producto de la cercanía del sector urbano a estas superficies invadidas con material sedimentario, al pasar de los días este mismo material fue trasladado a zonas urbanas por factores como el viento o lluvias, generando confusión entre las coberturas de ciudad y sedimento.

En Tubul existió un gran aumento en la clase de sedimentación húmeda. Efecto que responde mayormente al ensanchamiento de la playa registrada en esta localidad y al desecamiento de los ríos Tubul y Raquí, dejando al descubierto el sedimento del fondo, antes cubierto por agua

Todo lo anterior nos señala claramente que el terremoto y tsunami generó cambios morfológicos drásticos en la costa de la región del Biobío, lo que generó una nueva realidad geográfica. Esta realidad en conjunto con las transformaciones territoriales que constantemente experimentan las localidades estudiadas, hace que permanentemente se deba modelar realidades dinámicas que hacen que tanto la peligrosidad y vulnerabilidad cambien en el tiempo. Como ejemplo claro para nuestras localidades afectadas, se tiene que fueron localidades fuertemente intervenidas producto de la reconstrucción y se observó un rápido proceso de urbanización en áreas expuestas, así como la construcción de costaneras y ciclovías dentro de zonas intermareales, donde eventos comunes en las áreas tales como marejadas y olas de tormenta tienen la capacidad de producir cambios morfológicos. La caleta Tubul en este sentido se localiza en uno de los principales humedales de la región, y ocupa la antigua planicie de playa construida por procesos fluviométricos y tectónicos sobre la cual se ha desarrollado la marisma salobre, por lo cual es una zona frágil y que requiere de manejo ambiental. Los procesos de reconstrucción en este sentido, deben considerar la naturaleza y fragilidad de los sistemas costeros en donde se intervienen con el fin de evitar la degradación ambiental de estos y consecuencias negativas ante futuras situaciones de riesgo de desastres, que como se ha vislumbrado a lo largo de la historia, son eventos recurrentes en la costa chilena. Además se hacen

necesario de incorporar a la forma de ocupar territorio, caracterizado por una intensa antropización de la zona costera.

Este trabajo se suma a la consigna de que terremotos y tsunamis son eventos que pueden ser observados mediante imágenes satelitales, lo que ayuda de gran manera a aprender de estos eventos y sus efectos en el territorio. Utilizando información cartográfica y datos provenientes de imágenes satelitales, se tienen los insumos básicos para entender los cambios medioambientales que ha sufrido el territorio. Esta información se está utilizando cada vez más para la generación de catastros de coberturas y uso de suelo, desarrollando también métodos para identificar de manera rápida la extensión de la superficie que fue inundada por un tsunami, haciendo una estimación de cambios comparando imágenes antes y después del evento.

Bibliografía

- Andermann, C., Gloaguen, R., 2009. Estimation of erosion in tectonically active orogenies. Example from the Bhotekoshi catchment, Himalaya (Nepal). *International Journal of Remote Sensing* 30 (12), 3075–3096.
- Andy Liaw and Mathew Wiener, 2002. Classification and Regression by RandomForest. Vol.2/3, December 2002.
- Bala, G., Caldeira, K., Wickett, M., Phillips, T., Lobell, D., Delire, C., Mirin, A., 2007. Combined climate and carbon-cycle effects of large-scale deforestation. *Proceedings of the National Academy of Sciences* 104 (16), 6550–6555.
- BELMONTE, A. (2013). Identificación y análisis de riesgos naturales en la costa de la región del Biobío. Informe “Análisis de Riesgos de Desastres y Zonificaciones Costeras Región del Bío-Bío”. Proyecto FNDR 30098326, Gobierno Regional del BioBío (inédito).
- Betts, R., Falloon, P., Goldewijk, K., Ramankutty, N., 2007. Biogeophysical effects of land use on climate: model simulations of radiative forcing and large-scale temperature change. *Agricultural and Forest Meteorology* 142 (2–4), 216–233.
- Bonan, G., 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* 320 (5882), 1444–1449.
- Breiman L., 1996. Bagging Predictors. *International Journal of Remote Sensing* 20 (16), 2075–2096.
- Breiman, L., 2001. Random forests. *Machine Learning* 45 (1), 5–32.
- Burges B. Schölkopf and A. Smola. *Advances in kernel methods: Support vector machines*. Cambridge, MA: MIT Press, 1999.
- Burrough, P., 1998. *Principles of Geographical Information Systems*, 2nd Edition. Oxford University Press.
- Candade, N., 2004. Multispectral classification of Landsat images: a comparison of support vector machine and neural network classifiers. *ASPRS Annual Conference Proceedings*, Denver, Colorado.
- Chander G., Markham B., Helder D., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors.
- Chen P. Samek, David L. Skole, and Walter Chomentowski, “Assessment of Impact of the December 26, 2004 Tsunami In Aceh Province Indonesi”,
http://www.globalchange.msu.edu/trfic/tsunami2004/Assessment_of_Tsunami04.pdf

- CHUVIECO E., 2006. Teledetección ambiental. La observación de la tierra desde el Espacio. 2 ed. Barcelona: Ariel, 2006. p.242.
- CIENFUEGOS R., VILLAGRAN M., AGUILERA J., CATALAN P., CASTELLE, B., & ALMAR R. (2014). Video Monitoring and Field Measurements of a Rapidly Evolving Coastal System: the River Mouth and Sand Spit of the Mataquito River in Chile. *Journal of Coastal Research*, 70, 639-644. <http://dx.doi.org/10.2112/SI70-108.1>
- Congedo L (2013). Semi-Automatic Classification Plugin for QGIS User Manual Version 1.0.0. octubre 2013.
- Crammer and Singer (2000). On the learnability and design of output codes for multiclass problems. In *Computational Learning Theory*, pages 35–46, 2000.
- Cristian Araya-Cornejo & Matias Carvajal, 2016. Efectos Geomorfológicos del Tsunami de Chile de 2010 Frente a la Zona de Máximo Slip Revelado por Imágenes Satelitales y Observaciones de Campo: El Caso del Litoral Arenoso La Trinchera, Región del Maule. *Invetg. Geograf. Chile*, 52: 5-24 (2016).
- Dalponte, M., Bruzzone, L., Vescovo, L., Gianelle, D., 2009. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment* 113 (11), 2345–2355.
- Delfin D., 2010. Cuanto le costó el terremoto a Chile. Fuente: LA TERCERA.
- Diao, F., Xiong, X. (2014). *Overlapping post-seismic deformation processes: afterslip and viscoelastic relaxation following the 2011 Mw 9.0 Tohoku (Japan) earthquake*. *Geophysical Journal International*, 196(1), 218-229. [http:// dx.doi.org/10.1093/gji/ggt376](http://dx.doi.org/10.1093/gji/ggt376)
- Esch, T., Himmeler, V., Schorcht, G., Thiel, M., Wehrmann, T., Bachofer, F., Conrad, C., Schmidt, M., Dech, S., 2009. Large-area assessment of impervious surface based on integrated analysis of single-date Landsat-7 images and geospatial vector data. *Remote Sensing of Environment* 113 (8), 1678–1690.
- Fariña L., Opazo C., Vera P., 2012. Impactos ambientales del terremoto y tsunami en Chile: Las réplicas ocultas del 27/F. Fundación TERRAM.
- Feller, W., 1968. *An Introduction to Probability Theory and its Application*, third ed. Willey, New York, USA.
- Ferrero S.B., Palacio M.G. y Campanella O.R., 2002. Análisis de componentes principales en teledetección. Consideraciones estadísticas para optimizar su interpretación.
- Filippi, A.M., Archibald, R., 2014. Support vector machine-based endmember extraction. *IEEE Transactions on Geoscience and Remote Sensing* 47 (3), 771–791.

- Flavia Yarella Landeros Caceres (2014). Análisis de áreas afectadas por terremoto y tsunami en Constitución-Chile en el año 2010 por medio de técnicas de percepción remota. Monografía de XXIV Curso Internacional en Sensoramiento Remoto y Sistemas de Información Geográfica.
- Foody, G.M., Mathur, A., 2004a. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42 (6), 1335–1343.
- Friedl, M., Brodley, C., Strahler, A., 2002. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing* 37 (2), 969–977.
- Galiano V., B. Ghimire, J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez, 2012. An assessment of the Effectiveness of a RandomForest Classifier for Land-Cover Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*.
- Ghimire, B., Rogan, J., Miller, J., 2010. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters* 1, 45–54.
- Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2004. Random forest classification of multisource remote sensing and geographic data, Igarss 2004. In: *IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, AK, 20– 24 September. pp. 1049–1052.
- Guzmán J., 2010. *La tragedia en 2010*". Por: Roció Montes. Fuente: EL PAIS.
- Gyanesh Chander , Brian L. Markham , Dennis L. Helder, 2009. Summary of current Radiometric Calibration Coefficients for Landsat MSS, TM, ETM+ and EO-1 ALI sensors. *Remote Sensing of Environment* 113 (2009) 893-903.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *Random Forests, The Elements of Statistical Learning*. Springer, New York. pp. 587–604.
- Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing* 23 (4), 725–749
- Karatyoglu A., Meyer D., Hornik K., 2006. *Support Vector Machine in R*. *Journal of Statistical Software*. April 2006, Vol. 15, Issue 9. <http://www.jstatsoft.org/>
- Karatzoglu, A., Smola, A. (2005). New support vector algorithms. *Neural Computation*, 12, 1207–1245.
- Lawrence, R., Wood, S., Sheley, R., 2006. *Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest)*. *Remote Sensing of Environment* 100 (3), 356–362.

- Lillesand, T.M., Dash, J., Mathur, A., Foody, G.M., Curran, P.J., Chipman, J.W., 2007. Land cover classification using multi-temporal MERIS vegetation indices. *International Journal of Remote Sensing* 28 (6), 1137–1159.
- Marcal, A.R.S, Borges, J.S., Gomes, J.A., Pinto Da Costa, J.F., 2005. Land cover update by supervised classification of segmented ASTER images. *International Journal of Remote Sensing* 26 (7), 1347–1362.
- Martínez C. Reyes, 2013. Vulnerabilidades y desastres socio-naturales: aportes desde la geografía al proceso de reconstrucción en la región del Biobío. *REV.GEO SUR* 4 (6), 2013: 33-57.
- Martinez C., Rojas O., Jaque E., Vasquez D. Efectos territoriales del 27 de febrero de 2010 en la costa de la región del Biobío, Chile. *Revista Geográfica de América Central. Número Especial EGAL, 2011-Costa Rica. II Semestre 2011* pp.1-16.
- Maulik, U., 2009. Unsupervised pixel classification in satellite imagery using multiobjective fuzzy clustering combined with SVM classifier. *IEEE Transactions on Geoscience and Remote Sensing* 47 (4), 1132–1138.
- Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* 42 (8), 1778–1790.
- Mingers, J., 1989b. An empirical comparison of pruning methods for decision tree induction. *Machine Learning* 4 (2), 227–243.
- Morales Muñoz, Roberto (2010). Terremoto y Tsunami del 27 de febrero de 2010. Efectos Urbanos en Localidades de la Provincia de Arauco. *Urbano*, vol.13, num.22, octubre, 2010 pp. 43-62.
- Moreira, E., Morisson M., 2014: “Application and evaluation of topographic correction methods to improve land-cover mapping using object-based classification”, *International Journal of Applied Earth Observation and Geoinformation*, Vol.32, pp. 208-217.
- Morton R., Gelfenbaum G., Buckley M., & Richmond, B. (2011). Geological effects and implications of the 2010 tsunami along the central coast of Chile. *Sedimentary Geology*, 24, 34-51. <http://dx.doi.org/10.1016/j.sedgeo.2011.09.004>
- Mountrakis, G. 2011. Support vector machines in remote sensing: A review. Department of Environmental Resources Engineering, SUNY College of Environmental Science and Forestry, 1 Forestry Dr, Syracuse, NY 13210, USA.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* 26 (1), 217–222.
- P.Chen, S.C.Liew and L.K. Kwoh, 2005. *Tsunami Damage Assessment Using High Resolution Satellite Imagery: A Case Study of Aceh, Indonesia*. IEEE.

- P.K. Kumaresan et.al., 2014. *Classification of Remote Sensing Data Using Support Vector Machine and Random Forest Classifier with Color, Textural nad Shape Features*. International Journal of Emerging in Management & Technology. (Vol-3, Issue-11) November 2014.
- Quezada J., Jaque E., Belmonte A., Fernandez A., Vasquez D., Martinez C., 2010. Movimientos Cosismicos Verticales y Cambios Geomorfológicos Generados Durante el Terremoto Mw=8.8 del 27 de febrero de 2010 en el centro sur de Chile. *Revista Geográfica del Sur* 2: 11-45.
- Quezada J., Jaque E., Fernandez A., Vasquez D., 2012. Cambios en el Relieve generados como consecuencia del terremoto Mw=8.8 del 27 de febrero de 2010 en el centro Sur de Chile. *Revista Geografica de Geografia Norte Grande*, 53: 35-55 (2012).
- Quinlan, J., 1993. *C4.5 Programs for Machine Learning*, first ed. Morgan Kaurmann, San Mateo, CA.
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ramirez P. y Sandoval J., 2010. *Tsunami paso a paso: los escandalosos errores y omisiones del SHOA y la ONEMI*". Fuente: CIPER.
- Richards (1999). *Remote sensing digital image analysis*. Springer, New York
- Richmond, B. (2012). Erosion, deposition and landscape change on the Sendai coastal plain, Japan, resulting from the March 11, 2011 Tohoku-oki tsunami. *Sedimentary Geology*, 282, 27-39. <http://dx.doi.org/10.1016/j.sedgeo.2012.08.005>
- Rodriguez J., 2005. Análisis multitemporal satelital de los bosques del Carare-opón, mediante imágenes Landsat de 1991 y 2002.
- Rojas C., Mauricio V., Opazo S., Stefan P. and Villaroel C., 2013. *Pre and Post earthquake Land Use and Land Cover Identification in Concepcion*. Springer-Verlag Berlin Heidelberg 2013.
- Roy, D., Borak, J., Devadiga, S., Wolfe, R., Zheng, M., & Descloitres, J. (2002). TheMODIS land product quality assessment approach. *Remote Sensing of Environment*, 83, 62–76.
- Rozenstein y Karnieli (2011). *Comparison of methods for land-use classification incorporating remote sensing and GIS inputs*. The Remote Sensing Laboratory, Jacob Blaustein Institute for Desert Research, Ben-Gurion University of the Negev, Sede Boker Campus 84990, Israel
- Sesnie, S., Gessler, P., Finegan, B., Thessler, S., 2008. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments. *Remote Sensing of Environment* 112 (5), 2145–2159.
- Soenen, S.A., Peddle, D.R., Coburn, C.A., 2005: "SCS+C: A Modified Sun-Canopy-Sensor Topographic Correction in Forested Terrain", *IEEE Transactions on Geoscience and Remote Sensing*, Vol.43, No.9, pp. 2148 - 2159.

- Sumer E. and Celebi F., "Preliminary Information on Impacts of the 26th December 2006 Tsunami on Selected Coastal Ecosystems in Aceh Province, Indonesia", 2005.
http://www.riverbasin.org/ev_en.php?ID=3760_201&ID2=DO_TOPIC
- Swain, P., and Davis S., 1978. *Remote Sensing: The Quantitative Approach*. New York: McGraw Hill Book Company.
- Tappin, R., Evans, H. (2012). Coastal changes in the Sendai area from the impact of the 2011 Tōhoku-oki tsunami: Interpretations of time series satellite images, helicopter-borne video and field observations. *Sedimentary Geology*, 282, 151-174. [http:// dx.doi.org/10.1016/j.sedgeo.2012.09.011](http://dx.doi.org/10.1016/j.sedgeo.2012.09.011)
- Teillet, P. M., Barker, J. L., Markham, B. L., Irish, R. R., Fedosejeves, G., & Storey, J. C. (2001). Radiometric cross-calibration of the Landsat-7 ETM+ and Landsat-5 TM sensors based on tandem data sets. *Remote Sensing of Environment*, 78(1–2), 39–54.
- Tso, B., Mather, P., 2009. *Classification Methods for Remotely Sensed Data*, 2nd ed. CRC Press, 376 p.
- Vargas, G. y Farías, M. (2011). Coastal uplift and tsunami effects associated to the 2010 Mw8.8 Maule earthquake in Central Chile. *Andean Geology*, N° 38: 219 -238.
- Vasquez-Jimenez R., et.al., 2016. *Topographic Correction to Landsat Imagery by Slope Classes*. ResearchGate: <http://www.researchgate.net/publications/301677429>
- Vapnik, V., 1979. *Estimation of Dependences Based on Empirical Data*. Nauka, Moscow, pp. 5165–5184, 27.
- Vapnick V., 1998. *The nature of Statistical learning theory*. Springer Verlag, New York.
- Vogeler, J.C.; Cohen, W.B., 2016. *A review of the role of active remote sensing and data fusion for characterizing forest in wildlife habitat models*. Spanish Association of Remote Sensing.
- Weingessel A (2004). "quadprog – Functions to Solve Quadratic Programming Problems." R package, Version 1.4-7. URL <http://CRAN.R-project.org/>.
- Yonezawa C., 2007. Maximum likelihood classification combined with spectral angle mapper algorithm for high-resolution satellite imagery.
- Zhu, G., Blumberg, D.G., 2002. *Classification using ASTER data and SVM algorithms; The case study of Beer Sheva, Israel*. *Remote Sensing of Environment* 80 (2), 233–240.

