



**UNIVERSITÀ DEGLI STUDI DI BARI “ALDO MORO”**

**FACOLTÀ DI INFORMATICA**

**Corso di Laurea in Informatica**

---

Tesi di Laurea Sperimentale in

**MODELLO E METODI PER LA SICUREZZA  
DELLE APPLICAZIONI**

**CLASSIFICAZIONE DI COLTURE  
AGRICOLE DA DATI TELERILEVATI  
SENTINEL-2**

Relatore:

Donato Impedovo

Co-Relatore

Giuseppe Satalino

Laureanda:

Marianna  
Abbattista

Matr. n. 663721

---

Anno Accademico 2018/2019

## **DICHIARAZIONE**

Dichiaro che la tesi intitolata “Classificazione di colture agricole da dati telerilevati Sentinel-2 ” da me presentata, per il conseguimento del titolo di Dottoressa in Informatica presso UNIVERSITÀ DEGLI STUDI DI BARI “ALDO MORO” è una registrazione del lavoro in buona fede svolto da me sotto la supervisione del professore associato, di Modelli e metodi per la gestione delle applicazioni, Donato Impedovo, Dipartimento di Informatica, Bari e del Dr. Giuseppe Satalino, ricercatore Consiglio Nazionale delle Ricerche (CNR) – Istituto Rilevamento Elettromagnetico dell’Ambiente (IREA), Bari.

Dichiaro inoltre che il lavoro riportato in questa tesi non è stato presentato e non sarà sottoposto, in tutto o in parte, al conseguimento di altri titoli o diploma in questo istituto o in qualsiasi altro istituto o università.

Sede: Bari

Data:

**Firma del Candidato**

# Sommario

1	Introduzione .....	4
2	I dati satellitari Sentinel-2.....	5
2.1	Le bande spettrali di acquisizione di Sentinel-2.....	6
2.2	Sensore MSI di Sentinel-2.....	9
2.3	Prodotti e Algoritmi .....	12
2.4	Calibrazione e Validazione .....	13
2.5	Qualità del prodotto.....	13
2.6	Il dato Raster .....	14
3	Metodi di classificazione .....	15
3.1	Metodo della Massima Verosimiglianza.....	15
3.2	Random Forest .....	19
4	Il toolbox Sentinel-2 SNAP .....	15
4.1	Sentinel-2 Toolbox.....	22
4.2	Download dei dati e relativa importazione .....	24
4.3	Analisi dell'immagine .....	29
4.4	Creazione e gestione dei Vector Data Container .....	32
5	Descrizione dei dati sperimentali .....	34
5.1	Immagini Sentinel-2.....	34
5.2	Classi delle colture .....	37
6	Esperimenti di classificazione .....	39
6.1	Pre-processing dei dati .....	39
6.2	Classificazione .....	43
6.3	Post-Processing .....	63
7	Conclusioni.....	69
8	Bibliografia.....	70
9	Ringraziamenti.....	74

# 1 Introduzione

In questo lavoro di tesi sono state utilizzate tecnologie innovative per l'elaborazione, l'analisi e la classificazione di dati telerilevati del Satellite Europeo Sentinel-2.

L'obiettivo di questo studio è stato quello di effettuare una classificazione dei dati telerilevati per ottenere una mappa delle colture presenti in un'area agricola del Tavoliere Pugliese. Per questo obiettivo sono stati effettuati diversi esperimenti di classificazione basati sui dati multi-banda Sentinel-2 acquisiti in diverse date (ovvero in Aprile 2018 e a Luglio 2018).

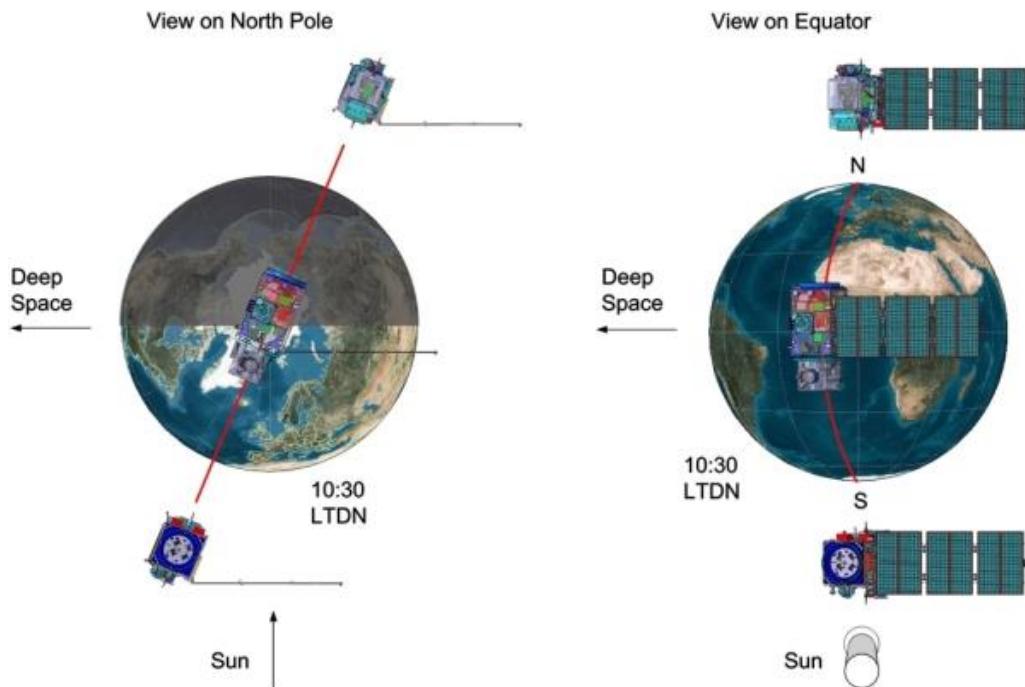
I dati Sentinel-2 utilizzati in questo lavoro di tesi sono stati scaricati dal sito Copernicus <https://scihub.copernicus.eu/>.

I dati di Sentinel-2 sono stati pre-processati usando il Toolbox SNAP. Questo Toolbox possiede inoltre anche degli algoritmi di classificazione supervisionata che sono stati applicati alle immagini multi-banda. Per questo scopo, sono stati utilizzati degli esempi di poligoni di verità a terra di colture della zona agricola di studio che fanno parte di un data base raccolto dal Consiglio per la Ricerca in Agricoltura e l'Analisi Economica (CREA) nell'ambito del progetto di ricerca Europeo "SENSAGRI".

La tesi è composta in tutto da nove capitoli. In particolare nel secondo capitolo viene presentata la missione Sentinel-2, con alcuni cenni sulla tipologia del sensore ottico, del dato misurato dal sensore, le bande elettromagnetiche di riferimento, oltre che alle sue caratteristiche geometriche e radiometriche. Nel terzo capitolo vengono analizzati dettagliatamente i metodi di classificazione utilizzati. Nel quarto capitolo viene descritta la Sentinel-2 Toolbox che è stata utilizzata per gli esperimenti di classificazione. Nel quinto capitolo vi è la descrizione dei dati sperimentali analizzati, ovvero le immagini Sentinel-2 e i poligoni di verità a terra, mentre nel sesto capitolo, vi è quella che spiega le operazioni di pre-processing iniziali, lo svolgimento della procedura di classificazione con i risultati di accuratezza e le operazioni di post-processing finali. Nel settimo capitolo vi sono le conclusioni. Infine, negli ultimi due capitoli sono riportate la bibliografia e i ringraziamenti.

## 2 I dati satellitari Sentinel-2

La missione Copernicus Sentinel-2 comprende una costellazione di due satelliti in orbita polare collocati nella stessa orbita eliosincrona, impostati fra loro a 180 gradi di distanza. Si dice eliosincrona un'orbita geocentrica che combina altezza e inclinazione in modo tale che un oggetto posto su quest'orbita, sorvoli ogni dato punto della superficie terrestre sempre alla stessa ora solare locale. Il sistema mira a monitorare la variabilità delle condizioni della superficie del terreno, e la sua ampia larghezza della traccia di acquisizione a terra (ovvero uno swath di 290 km) e i tempi di rivisita elevati (10 giorni all'equatore con un satellite, 5 giorni con 2 satelliti in condizioni senza nuvole, che si traduce in 2-3 giorni a metà latitudine) sosterranno il monitoraggio dei cambiamenti della superficie terrestre. I limiti di copertura sono tra latitudine a 56 gradi sud e 84 gradi a nord (Figura 1).



**Figura 1. La configurazione orbitale di Sentinel-2.**

Sentinel-2 trasporta un carico utile di strumenti ottici che campiona 13 bande spettrali: quattro bande a 10 m, sei bande a 20 m e tre bande con una risoluzione spaziale di 60 m.

I satelliti gemelli di Sentinel-2 (ovvero Sentinel-2A e Sentinel-2B) forniscono continuità ai dati immagine di tipo SPOT e LANDSAT, contribuiscono alle osservazioni multispettrali in corso e avvantaggiano i servizi e le applicazioni Copernicus come gestione del territorio, agricoltura e silvicoltura, controllo delle catastrofi, operazioni di soccorso umanitario, mappatura dei rischi e problemi di sicurezza.

Gli obiettivi della missione Sentinel-2 sono di fornire:

- acquisizione sistematica globale di immagini multispettrali ad alta risoluzione associate a un'alta frequenza di visita;
- continuità delle immagini multispettrali fornite dalla serie di satelliti SPOT e dallo strumento USGS LANDSAT Thematic Mapper;

- dati di osservazione per la prossima generazione di prodotti operativi, come mappe di copertura del suolo, mappe di rilevamento del cambiamento del suolo e variabili geofisiche.

## 2.1 Le bande spettrali di acquisizione di Sentinel-2

Lo spettro elettromagnetico (abbreviato spettro EM) indica l'insieme di tutte le possibili frequenze delle radiazioni elettromagnetiche.

Pur essendo lo spettro continuo, è possibile una suddivisione puramente convenzionale ed indicativa in vari intervalli o bande di frequenza, dettata a partire dallo spettro ottico. L'intero spettro è suddiviso nella parte di spettro visibile che dà vita alla luce e le parti di spettro non visibile a lunghezza d'onda maggiori e minori dello spettro visibile. Le onde di lunghezza nell'intervallo tra la luce visibile e le onde radio, a bassa intensità hanno poca energia e risultano scarsamente dannose, le radiazioni comprese tra l'ultravioletto e i raggi gamma invece hanno più energia, sono ionizzanti e quindi possono danneggiare gli esseri viventi.

In fisica lo spettro visibile è quella parte dello spettro elettromagnetico che cade tra il rosso e il violetto includendo tutti i colori percepibili dall'occhio umano che danno vita dunque al fenomeno della luce. La lunghezza d'onda della luce visibile nell'aria va indicativamente dai 390 ai 700 nm; le lunghezze d'onda corrispondenti in altri mezzi, come l'acqua, diminuiscono proporzionalmente all'indice di rifrazione. In termini di frequenze, lo spettro visibile varia tra i 770 ed i 430 THz.

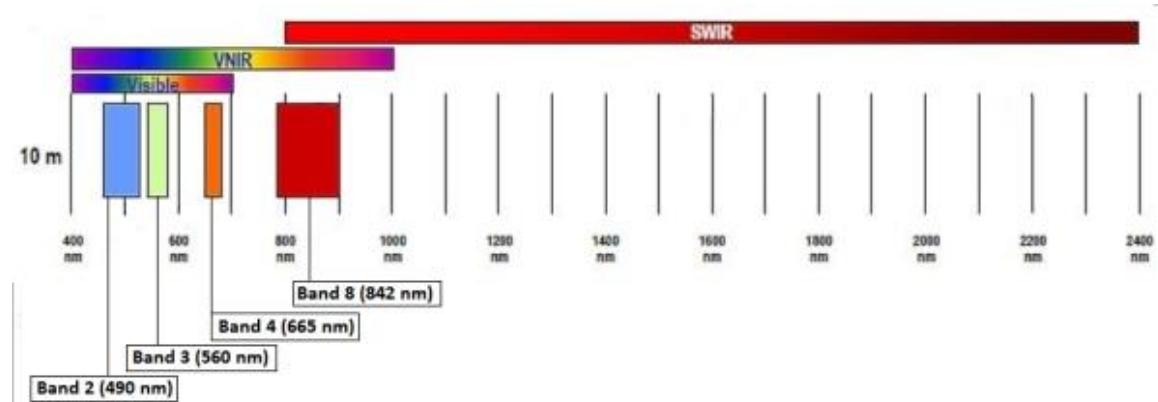
Il sensore ottico di Sentinel-2 (chiamato MultiSpectral Instrument –MSI e che è illustrato nella prossima sotto-sezione) fornisce misurazioni nello spettro EM con delle caratteristiche peculiari di:

- risoluzione temporale;
- risoluzione spaziale;
- risoluzione radiometrica.

La **risoluzione temporale** del sistema in orbita, ovvero la frequenza di visita del satellite su una determinata posizione, è di 10 giorni con un singolo satellite e di 5 giorni con due satelliti.

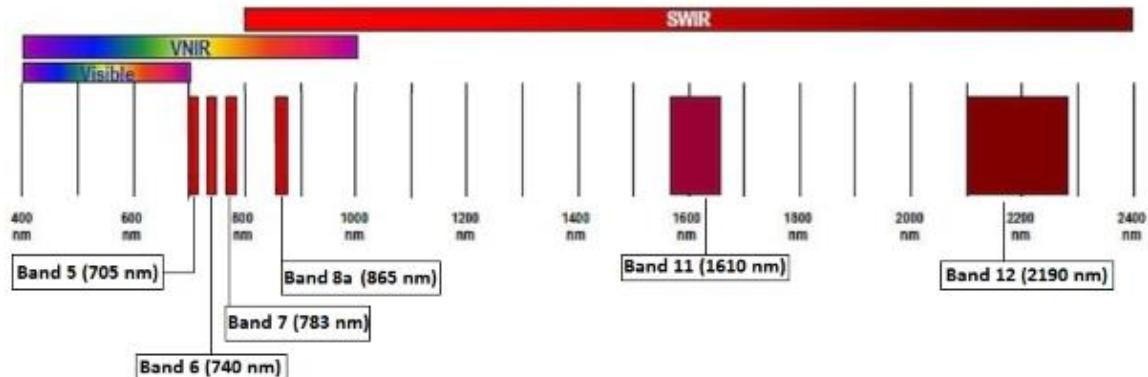
La **risoluzione spaziale** del sensore, ovvero la dimensione spaziale a terra corrispondente al singolo rivelatore in una matrice di sensori, dipende dalle particolari bande spettrali e i vari valori sono indicati nelle figure di seguito riportate:

*Risoluzione spaziale di 10 metri (Figura 2):*



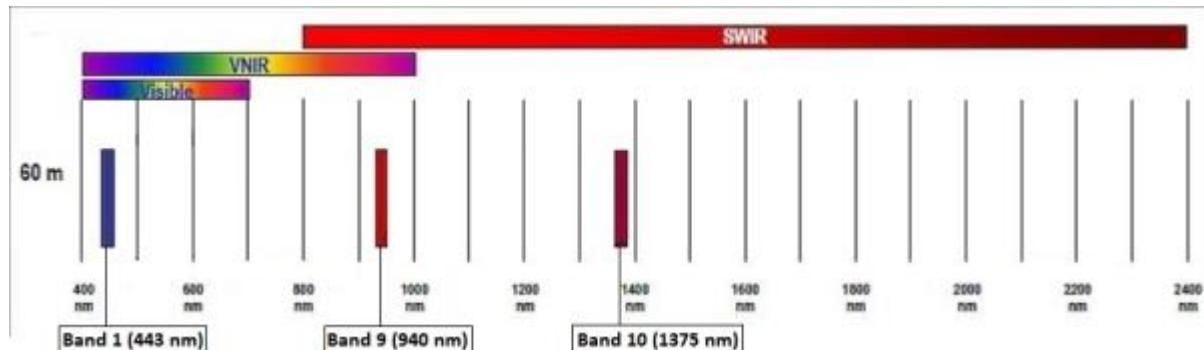
**Figura 2. Bande con risoluzione spaziale di 10 m: B2 (490 nm), B3 (560 nm), B4 (665 nm) e B8 (842 nm).**

Risoluzione spaziale di 20 metri (Figura 3):



**Figura 3.** Bande con risoluzione spaziale di 20 m: B5 (705 nm), B6 (740 nm), B7 (783 nm), B8a (865 nm), B11 (1610 nm) e B12 (2190 nm).

Risoluzione spaziale di 60 metri (Figura 4):



**Figura 4.** Bande con risoluzione spaziale di 60m: B1 (443 nm), B9 (940 nm) e B10 (1375 nm).

Per riassumere, le 13 bande spettrali di Sentinel-2 vanno da Visibile (VNIR) e Near Infra-Red (NIR) a Short Wave Infra-Red (SWIR) ed hanno le seguenti risoluzioni spaziali:

- 4 bande da 10 metri: le tre bande RGB classiche ((blu (~ 493nm), verde (560nm) e rossa (~ 665nm)) e una banda nel vicino infrarosso (~ 833nm);
- 6 bande da 20 metri: 4 bande strette nel dominio spettrale del bordo rosso della vegetazione VNIR (~ 704nm, ~ 740nm, ~ 783nm e ~ 865nm) e 2 bande SWIR più larghe (~ 1610nm e ~ 2190nm) per applicazioni come neve / ghiaccio / rilevamento delle nuvole o valutazione dello stress da umidità della vegetazione;
- 3 bande da 60 metri che sono usate principalmente per lo screening delle nuvole e per la correzione atmosferica (~ 443nm per gli aerosol e ~ 945nm per il vapore acqueo) e per il rilevamento dei cirri (~ 1374nm).

La **risoluzione radiometrica** del sensore rappresenta il più piccolo livello incrementale di intensità o riflettanza che può essere rappresentato o distinto dal sistema. In generale, maggiore è la risoluzione radiometrica, maggiore sarà la capacità del dispositivo di rilevare differenze di intensità o riflettanza.

La risoluzione radiometrica viene abitualmente espressa come un numero di bit, in genere nell'intervallo da 8 a 16 bit. La risoluzione radiometrica del sensore di Sentinel-2 è di 12 bit e consente di acquisire l'immagine in un intervallo da 0 a 4095 valori di intensità della luce potenziali. La precisione radiometrica per Sentinel-2 è inferiore al 5% (obiettivo 3%). La risoluzione radiometrica dipende anche dal rapporto segnale / rumore (SNR) del rivelatore. Nelle tabelle che seguono, vengono riportate per ogni banda spettrale, le larghezze di banda, la radianza di riferimento e il rapporto segnale / rumore (SNR) associato.

**Tabella 1. Bande con risoluzione spaziale di 10 m e rapporto segnale/rumore (SNR) associato.**

Band number	S2A		S2B		$L_{ref}$ (reference radiance) (W m <sup>-2</sup> sr <sup>-1</sup> μm <sup>-1</sup> )	SNR @ $L_{ref}$
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
2	492.4	66	492.1	66	128	154
3	559.8	36	559.0	36	128	168
4	664.6	31	664.9	31	108	142
8	832.8	106	832.9	106		

**Tabella 2. Bande con risoluzione spaziale di 20 metri e rapporto segnale/rumore (SNR) associato.**

Band number	S2A		S2B		$L_{ref}$ (reference radiance) (W m <sup>-2</sup> sr <sup>-1</sup> μm <sup>-1</sup> )	SNR @ $L_{ref}$
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
5	704.1	15	703.8	16	74.5	117
6	740.5	15	739.1	15	68	89
7	782.8	20	779.7	20	67	105
8a	864.7	21	864.0	22	52.5	72

11	1613.7	91	1610.4	94	4	100
12	2202.4	175	2185.7	185	1.5	100

**Tabella 3. Bande con risoluzione spaziale di 60 metri e rapporto segnale/rumore (SNR) associato.**

Band number	S2A		S2B		$L_{ref}$ (reference radiance) (W m <sup>-2</sup> sr <sup>-1</sup> $\mu\text{m}^{-1}$ )	SNR @ $L_{ref}$
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)		
1	442.7	21	442.2	21	129	129
9	945.1	20	943.2	21	9	114
10	1373.5	31	1376.9	30	6	50

## 2.2 Sensore MSI di Sentinel-2

Il sensore ottico di Sentinel-2 è il MultiSpectral Instrument (MSI), progettato per ottenere alte prestazioni di risoluzione geometrica e spettrali delle misurazioni della radianza riflessa della Terra nelle 13 bande spettrali da VNIR a SWIR. La Tabella 4 riporta le bande di acquisizione del sensore ordinate in modo crescente rispetto alla lunghezza d'onda. La larghezza di banda (in inglese Bandwidth in nm) è misurata alla “Larghezza massima metà massima” (in inglese Full Width Half Maximum - FWHM).

**Tabella 4: Bande spettrali per i sensori Sentinel-2 (S2A & S2B).**

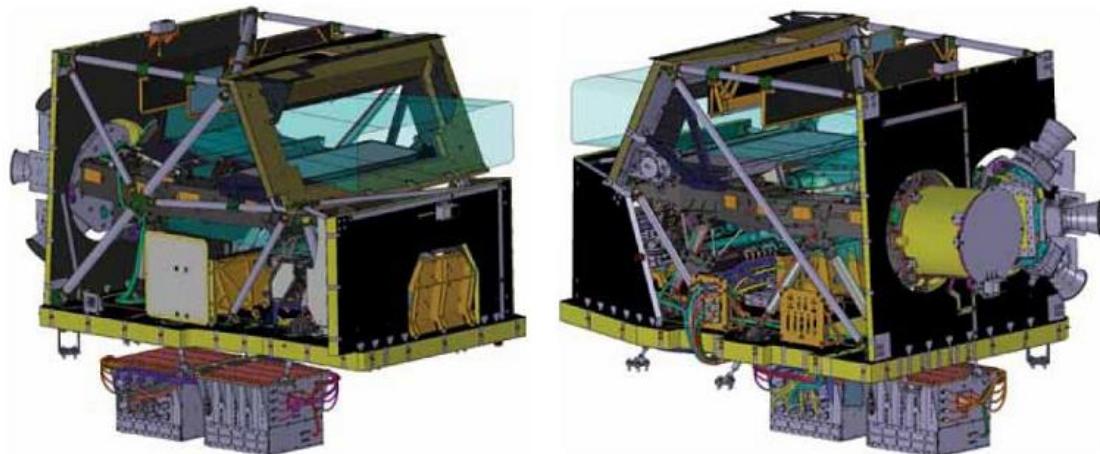
	S2A		S2B		
Band Number	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
1	442.7	21	442.2	21	60
2	492.4	66	492.1	66	10
3	559.8	36	559.0	36	10
4	664.6	31	664.9	31	10
5	704.1	15	703.8	16	20
6	740.5	15	739.1	15	20

<b>7</b>	782.8	20	779.7	20	20
<b>8</b>	832.8	106	832.9	106	10
<b>8a</b>	864.7	21	864.0	22	20
<b>9</b>	945.1	20	943.2	21	60
<b>10</b>	1373.5	31	1376.9	30	60
<b>11</b>	1613.7	91	1610.4	94	20
<b>12</b>	2202.4	175	2185.7	185	20

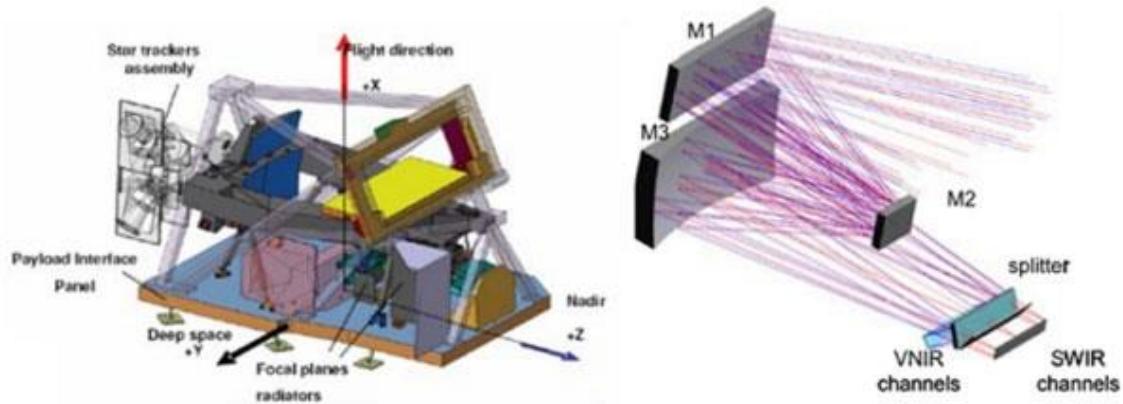
Il sistema MSI è progettato e realizzato da Airbus Defence and Space, Francia, e si basa su un concetto di push-broom (Figura 5).

Le caratteristiche del sistema MSI sono:

- Un telescopio a tre specchi Anastigmat (TMA) con un diametro della pupilla equivalente a 150 mm, montato isostaticamente sulla piattaforma per ridurre al minimo le distorsioni termoelastiche. Il design ottico è stato ottimizzato per ottenere delle immagini di qualità a-stato-dell'arte attraverso il suo campo visivo molto ampio. La sua larghezza di swath è di 290 km è maggiore rispetto alle precedenti missioni ottiche multispettrali come SPOT e LANDSAT. La struttura del telescopio e gli specchi sono realizzati in carburo di silicio, che offre un'altissima stabilità ottica rispetto al rapporto di massa e riduce al minimo la deformazione termica.
- Ha due piani focali basati su:
  - rivelatori monolitici CMOS per VNIR.
  - rivelatori di mercurio-cadmio-tellurio ibridati su un circuito di lettura CMOS per SWIR (il piano focale SWIR è termicamente controllato passivamente a temperature inferiori a 195 K).
- Uno splitter dicroico VNIR / SWIR. La separazione spettrale nelle varie bande è fornita da filtri a strisce montati sulla parte superiore dei rivelatori.
- Un diffusore di bordo a campo pieno e a pupilla intera per la calibrazione radiometrica per garantire prestazioni radiometriche di alta qualità.
- Due array distinti di 12 rivelatori montati su ciascun piano focale che coprono rispettivamente i canali VNIR e SWIR. I 12 rivelatori su ciascun piano focale sono in una configurazione sfalsata per coprire l'intero campo visivo.
- Un meccanismo di scatto che impedisce allo strumento di illuminare direttamente il sole. Questo meccanismo viene anche utilizzato come dispositivo di calibrazione raccogliendo la luce solare dopo il riflesso da un diffusore.



**Figura 5:** The MultiSpectral Instrument (Airbus Defence and Space).

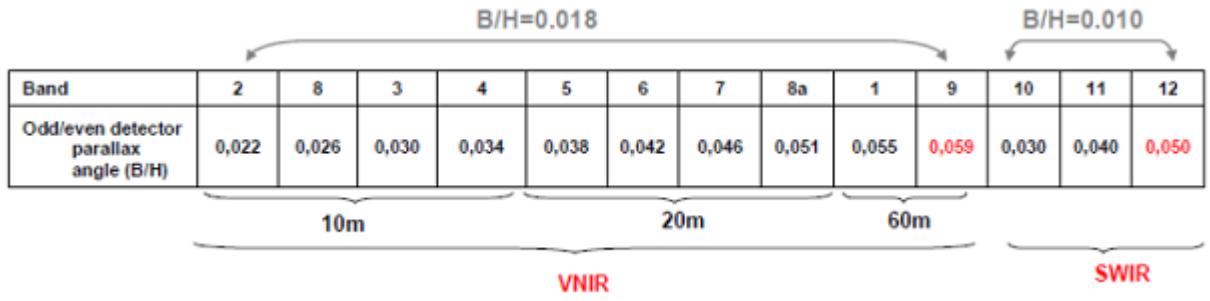


**Figura 6:** Configurazione Interna di MSI.

I 12 rivelatori su ciascun piano focale sono montati in una formazione sfalsata (Figura 6) per coprire l'intero campo visivo di dello strumento di  $20,6^{\circ}$ , risultando in uno swath avente una larghezza a terra di 290 km.

Come illustrato nella Figura 7, a causa del posizionamento sfalsato dei rivelatori sui piani focali, nelle misurazioni viene indotto un angolo di parallasse tra i due gruppi dispari e pari alternati di rivelatori, con conseguente spostamento lungo il binario di circa 46 km (massimo) interrivelatore. $20,6^{\circ}$  dello strumento, risultando in uno swath avente una larghezza a terra di 290 km.

Allo stesso modo, il design hardware di entrambi i rivelatori VNIR e SWIR impone uno spostamento relativo di ciascun sensore del canale spettrale all'interno del rivelatore, determinando una parallasse di misurazione interbanda pari a un massimo di spostamento della pista di circa 14 km.



**Figura 7: Configurazione del rivelatore sfalsato e angoli di parallasse inter-rivelatore / interbanda (figure di parallasse derivate dalla documentazione dello strumento MSI).**

A causa della particolare disposizione geometrica del piano focale, ogni banda spettrale dell'MSI osserva la superficie del terreno in momenti diversi.

## 2.3 Prodotti e Algoritmi

Il dato acquisito dal MSI viene processato per ottenere i seguenti prodotti:

- Level-0: è costituito da dati grezzi compressi. Il prodotto di livello 0 contiene tutte le informazioni necessarie per generare i livelli di prodotto di livello 1 (e superiore).
- Level-1A: è costituito da dati grezzi non compressi con bande spettrali registrate in modo grossolano e dati annessi
- Level-1B: sono dati di radianza corretti per via radiometrica. Il modello geometrico fisico viene perfezionato utilizzando i punti di controllo del suolo disponibili e aggiunto al prodotto, ma non applicato.
- Level-1C: fornisce la riflettanza top-of-atmosphere (TOA) orto rettificata nella proiezione cartografica UTM/WGS84, con registrazione multispettrale al sub-pixel. Nel prodotto è inclusa una maschera per le classi nuvole, terra e acqua.
- Level-2A: fornisce riflettanza Bottom-Of-Atmosphere (BOA) orto rettificata sulla geometria cartografica con proiezione UTM/WGS84, con registrazione multispettrale al sub-pixel. Nel prodotto è inclusa una mappa di classificazione delle scene (nuvola, ombre delle nuvole, vegetazione, suoli / deserti, acqua, neve, ecc.).

I prodotti Level-0, Level-1A e Level-1B non vengono divulgati agli utenti, mentre i prodotti Level-1C e Level-2A sono resi disponibili agli utenti tramite Copernicus Open Access Hub, con scene aventi dimensioni di 100x100 km<sup>2</sup>.

L'acquisizione, l'elaborazione, l'archiviazione e la diffusione dei dati dal livello 0 al livello 2A sono eseguiti da Sentinel-2 Ground Segment. Il livello 2A può anche essere eseguito dall'utente utilizzando Sentinel-2 Toolbox.

## 2.4 Calibrazione e Validazione

La calibrazione e la validazione (CalVal) sono processi vitali nel controllo di qualità dei prodotti generati e garantiscono che i dati elaborati siano significativi per gli utenti scientifici.

La calibrazione viene effettuata valutando la risposta dello strumento su aree note e identificando le eventuali variazioni che si verificano naturalmente nel tempo nei sistemi e nell'hardware, in modo che esse siano corrette tramite l'adattamento dei parametri di calibrazione, in modo da garantire la stabilità dell'output.

Dopo aver effettuato la calibrazione, il processo richiede la fase di validazione. La validazione consiste nel verificare che i nuovi parametri di calibrazione adottati producano il risultato atteso e che quindi il prodotto è stabile ed è quindi idoneo alla sua diffusione presso la comunità degli utenti.

La performance identifica i requisiti di qualità dell'immagine radiometrica e geometrica dei prodotti Sentinel-2.

## 2.5 Qualità del prodotto

Il Centro di coordinamento del Mission Performance Center (MPC-CC) fornisce uno stato mensile della qualità del prodotto Sentinel-2 attraverso la diffusione di un Rapporto sulla qualità dei dati (DQR). Il rapporto fornisce informazioni sul monitoraggio e la misurazione delle prestazioni del prodotto rispetto alle specifiche proposte. Documenta inoltre le anomalie osservate e i problemi noti, l'elenco dei pixel difettosi e qualsiasi miglioramento della catena di elaborazione che porta a un incremento della linea di base di elaborazione.

I DQR forniscono informazioni sul monitoraggio e sulla misurazione delle prestazioni dei prodotti L1C e L2A rispetto alle specifiche proposte, vale a dire:

Prestazioni geometriche:

- Stato di calibrazione geometrica
- Geolocalizzazione assoluta
- Registrazione multi-spettrale
- Registrazione multi-temporale

Prestazioni radiometriche:

- Stato di calibrazione radiometrica
- Incertezza radiometrica
- Rumore
- Modulazione Funzione di trasferimento

## 2.6 Il dato Raster

Il dato Raster (detto anche semplicemente bitmap), nella computer grafica, è una tecnica usata per descrivere un'immagine in formato digitale che si contrappone al dato vettoriale.

Il termine Raster ha origine nella tecnologia televisiva analogica, ovvero dal termine che indica le righe orizzontali (dette anche scan-line) dei televisori o dei monitor).

Nella computer grafica, Raster indica la griglia ortogonale di punti che costituisce un'immagine Raster, ovvero l'immagine viene vista come una scacchiera e ad ogni elemento della scacchiera, chiamato pixel, viene associato uno specifico colore. Il colore può essere definito con due tecniche:

- se l'immagine contiene pochi colori (massimo 256), si crea un elenco dei colori da utilizzare e nella scacchiera viene inserito l'indice che punta allo specifico colore del pixel;
- se l'immagine contiene molti colori, il singolo pixel non definisce l'indice con il quale si punta a una tavolozza di colori, ma direttamente il colore.

Il colore viene definito, ad esempio nel sistema RGB (R Rosso, G Verde, B Blu), come una combinazione di tre componenti: blu, rosso, verde. Questo non è l'unico modo di definire un colore, esistono altri modi che vengono chiamati spazi di colore, ma nel caso delle immagini generate al computer, il sistema RGB è il più diffuso dato che le schede grafiche lo utilizzano nativamente per generare il segnale da visualizzare con il monitor.

Nel campo dei Sistemi informativi territoriali o GIS, la locuzione è usata per indicare la tipologia di dato impiegata nella rappresentazione cartografica digitale. Con i dati Raster il territorio viene riprodotto attraverso una matrice di pixel di forma quadrata o rettangolare. A ciascun pixel è associato un attributo che definisce le caratteristiche dell'elemento rappresentato.

Ad esempio in un modello digitale di elevazione, a ciascun pixel è associato il valore della quota sul livello del mare in quel punto. La dimensione dei pixel è inversamente proporzionale alla precisione della carta. I dati Raster possono essere implementati in un sistema GIS mediante acquisizione diretta con apparecchiature a lettura ottica quali ad esempio scanner d'immagini o attraverso l'elaborazione di dati, Raster o vettoriali, già acquisiti.

Nel caso delle immagini acquisite da satellite, il Raster contiene i valori di radianza misurati dal sensore nelle varie bande spettrali secondo una scala di valori che può essere assoluta (a valori reali) o relativa (Digital Number a valori interi).

### 3 Metodi di classificazione

In questo capitolo vengono analizzati dettagliatamente i metodi di classificazione utilizzati. Essi sono: il metodo della Massima Verosimiglianza ed il metodo Random Forest.

#### 3.1 Metodo della Massima Verosimiglianza

Il metodo della massima verosimiglianza, in statistica, è un procedimento matematico per determinare uno stimatore. Caso particolare della più ampia classe di metodi di stima basata sugli estimatori d'estremo, il metodo consiste nel massimizzare la funzione di verosimiglianza, definita in base alla probabilità di osservare una data realizzazione campionaria, condizionatamente ai valori assunti dai parametri statistici oggetto di stima. Il metodo è stato sviluppato, originariamente, dal genetista e statistico sir Ronald Fisher, tra il 1912 e il 1922.

##### Definizione 1:

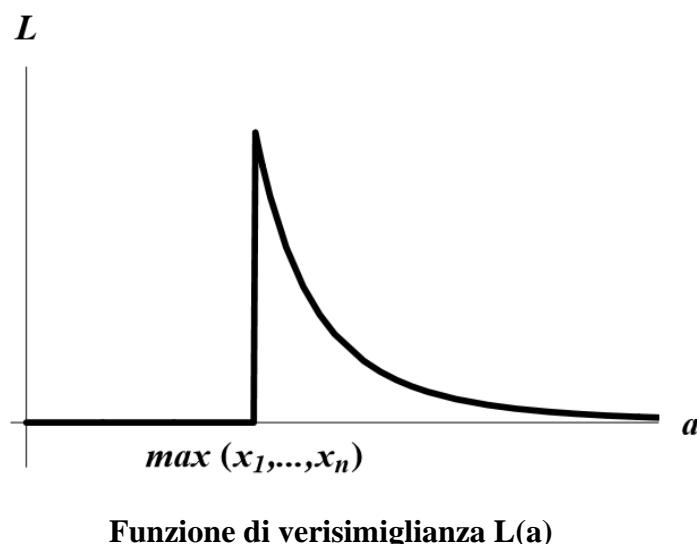
Dato un campione  $X_1, \dots, X_n$  di una variabile aleatoria  $X$ , con le  $X_k$  indipendenti e identicamente distribuite, chiameremo allora funzione di verosimiglianza la funzione  $L$  di  $\theta$ :

$$L(\theta) = P_\theta \{X_1 = x_1\} \cdot \dots \cdot P_\theta \{X_n = x_n\} = p_\theta(x_1) \cdot \dots \cdot p_\theta(x_n)$$

nel caso in cui  $X$  è una variabile aleatoria discreta con valori  $x_k$  e dove  $P$  è la funzione di probabilità congiunta, condizionata dal parametro  $\theta$ , oppure la funzione  $L$  di  $\theta$ :

$$L(\theta) = \text{pdf } \theta(x_1) \cdot \dots \cdot \text{pdf } \theta(x_n)$$

nel caso in cui  $X$  è una variabile aleatoria continua, dove  $\text{pdf}$  è la funzione di densità di probabilità congiunta, condizionata dal parametro  $\theta$ . Assegnando i valori  $x_1, \dots, x_n$  al campione,  $L(\theta)$  dipenderà solo dal parametro  $\theta$  e sarà possibile determinare il particolare valore  $\theta_1$  di  $\theta$  per il quale  $L(\theta)$  è massima. In questo modo  $\theta_1$  è il valore del parametro per il quale i valori  $x_1, \dots, x_n$  coincidono con la moda del campione. Quindi, scegliere  $\theta_1$  vuol dire calibrare il parametro incognito in modo tale che quel che si misura (il campione dato) sia sempre il risultato più probabile (moda).



Ovviamente  $\Theta_1 = \theta_1(x_1, \dots, x_n)$  dipenderà dai valori di  $x_1, \dots, x_n$  anche se in genere, per semplificare la notazione, noi eviteremo di metterlo in evidenza. Adotteremo quindi come stimatore di Massima Verosimiglianza la statistica:

$$\Theta_1 = \theta_1(X_1, \dots, X_n) \text{ dove } \theta_1 = \theta_1(x_1, \dots, x_n)$$

Con cui trovare il valore che rende massima la funzione di verosimiglianza  $L(\theta)$  per un dato campione  $x_1, \dots, x_n$ .

Il massimo della  $L(\theta)$  è in genere determinato con i tradizionali metodi analitici (annullamento della derivata); siccome però  $L(\theta)$  si presenta come un prodotto di  $n$  funzioni, e siccome non è sempre agevole derivare tali prodotti, spesso si preferisce determinare il massimo della verosimiglianza logaritmica  $\log L(\theta)$  che si presenta invece come una somma di funzioni di  $\theta$ . Le due procedure sono equivalenti dato che la funzione  $\log x$  è monotona crescente.

## Definizione 2:

Data una distribuzione di probabilità  $D$ , con funzione di massa (o densità, se continua) di probabilità  $\mathcal{L}_D$ , caratterizzata da un parametro  $\vartheta$ , dato un campione di dati osservati  $\{x_i\}_{i=1}^n$  di dimensione  $n$  si può calcolare la probabilità associata ai dati osservati:

$$P(\{x_i\}_{i=1}^n | \vartheta) = \mathcal{L}_D(\vartheta | \{x_i\}_{i=1}^n)$$

D'altra parte, può darsi che il parametro  $\vartheta$  sia ignoto, sebbene sia noto che il campione è estratto dalla distribuzione  $D$ . Un'idea per stimare  $\vartheta$  è allora utilizzare i dati a nostra disposizione:  $\{x_i\}_{i=1}^n$  per ottenere informazioni su  $\vartheta$ .

Il metodo della massima verosimiglianza ricerca il valore più verosimile di  $\vartheta$ , ossia ricerca, all'interno dello spazio  $\Theta$  di tutti i possibili valori di  $\vartheta$ , il valore del parametro che massimizza la probabilità di aver ottenuto il campione dato. Da un punto di vista matematico,  $\mathcal{L}_D(\vartheta | \{x_i\}_{i=1}^n)$  o equivalentemente  $\mathcal{L}_D(\vartheta | x_1, \dots, x_n)$  è una funzione di verosimiglianza, e lo stimatore di massima verosimiglianza è ottenuto come:

$$\hat{\vartheta} = \arg \max_{\vartheta \in \Theta} \mathcal{L}_D(\vartheta | x_1, \dots, x_n)$$

## Esempi

Al fine di illustrare il metodo della massima verosimiglianza si consideri un campione  $\{x_i\}_{i=1}^n$  di variabili casuali identicamente e indipendentemente distribuite, con distribuzione normale:  $x_i \sim N(\mu, \sigma^2) \forall i$ . La funzione di verosimiglianza associata è:

$$\mathcal{L}(\mu, \sigma^2 | \{x_i\}_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\}$$

La massimizzazione della funzione di verosimiglianza è equivalente a massimizzarne il logaritmo:

$$L(\mu, \sigma^2 | \{x_i\}_i) = \ln \mathcal{L}(\mu, \sigma^2 | \{x_i\}_i) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2$$

I parametri  $\mu$  e  $\sigma^2$  sono determinati risolvendo il problema di massimo:

$$\{\mu, \sigma^2\} = \arg \max_{\mu, \sigma^2} L(\mu, \sigma^2 | \{x_i\}_i)$$

Le condizioni del primo ordine per un massimo definiscono il seguente sistema di equazioni in  $\mu$  e  $\sigma^2$ :

$$\frac{\partial L}{\partial \mu} = \frac{1}{\hat{\sigma}^2} \sum_i (x_i - \hat{\mu}) = 0$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_i (x_i - \hat{\mu})^2 = 0$$

dove i segni di apice sopra i parametri denotano i loro stimatori. Dalla prima equazione discende immediatamente lo stimatore di massima verosimiglianza per la media:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

cioè la media campionaria. La varianza dello stimatore  $\hat{\mu}$  è data dalla seguente espressione:

$$\text{var}(\hat{\mu}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(x_i) = \frac{\sigma^2}{n}$$

Sostituendo  $\hat{\mu}$  nella seconda equazione, si ha lo stimatore di massima verosimiglianza per la varianza:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

cioè la varianza campionaria.

L'esempio è particolarmente calzante, perché consente di illustrare alcune proprietà degli stimatori di massima verosimiglianza. È immediato verificare la correttezza (o unbiasedness) di  $\hat{\mu}$ :

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[x_i] = \mu$$

D'altra parte,  $\hat{\sigma}^2$  non gode di tale proprietà. Ricordando che:

$$\sum_i (x_i - \mu)^2 = \sum_i (x_i - \hat{\mu})^2 + n(\hat{\mu} - \mu)^2$$

segue che:

$$\mathbb{E}[\hat{\sigma}^2] = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n (x_i - \hat{\mu})^2\right) = \frac{1}{n} \mathbb{E}\left[\sum_i (x_i - \mu)^2 - n(\hat{\mu} - \mu)^2\right] = \frac{n-1}{n} \sigma^2$$

Dunque  $\hat{\sigma}^2$  non è uno stimatore corretto; un tale stimatore sarebbe dotato della statistica:

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

Val la pena d'altra parte di osservare che lo stimatore di massima verosimiglianza è comunque uno stimatore asintoticamente corretto; infatti

$$\lim_{n \rightarrow \infty} E[\hat{\sigma}^2] = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2$$

In particolare, qualunque stimatore di massima verosimiglianza è asintoticamente corretto e asintoticamente normalmente distribuito.

L'espressione per la varianza dello stimatore  $\hat{\sigma}^2$  è al di là degli scopi di questo esempio.

È interessante osservare che gli stimatori derivati in questa sezione sono identici a quelli ottenibili, nelle stesse condizioni, impiegando il metodo dei momenti; a scanso di equivoci, si precisa che i due metodi di ricerca degli stimatori non conducono necessariamente a individuare gli stessi stimatori in condizioni più generali.

## Proprietà degli stimatori di massima verosimiglianza

### Invarianza funzionale

Se  $\hat{\vartheta}$  è lo stimatore di massima verosimiglianza per il parametro  $\vartheta$ , allora lo stimatore di massima verosimiglianza per  $\alpha = g(\vartheta)$  è  $\hat{\alpha} = g(\hat{\vartheta})$ , purché  $g$  sia una funzione biettiva.

### Distorsione

Gli stimatori di massima verosimiglianza, come illustrato negli esempi, possono essere distorti (cioè non corretti o all'inglese biased), anche in maniera consistente. D'altra parte essi sono asintoticamente corretti.

### Efficienza e comportamento asintotico

Gli stimatori di massima verosimiglianza non conseguono in generale il limite inferiore per la varianza stabilito dal risultato di Cramér-Rao, lo conseguono però asintoticamente, cioè la varianza si discosta dal limite inferiore di Cramér-Rao per una quantità infinitesima al crescere di  $n$ . Gli stimatori di massima verosimiglianza sono inoltre asintoticamente normalmente distribuiti.

## 3.2 Random Forest

Il metodo Random Forest è un classificatore d'insieme composto da molti alberi di decisione, che fornisce in uscita la classe associata ad un campione di input sulla base delle stime di classi effettuate dagli alberi presi individualmente. L'algoritmo per indurre a una foresta casuale fu sviluppato da Leo Breiman e Adele Cutler.

Leo Breiman lo definisce come un insieme di predizioni di alberi in modo tale che ogni albero dipenda dai valori di un vettore casuale campionato in maniera indipendente e con la distribuzione per tutti gli alberi della foresta.

Il nome viene dalle foreste di decisione casuali che furono proposte per primo da Tin Kam Ho dei Bell Labs nel 1995.

Il metodo combina l'idea dell'insaccamento di Breiman in cui si ha una selezione casuale delle caratteristiche, introdotto indipendentemente da Ho e Amit Geman per costruire una collezione di alberi di decisione con la variazione controllata.

La selezione di un sottoinsieme di caratteristiche è un esempio del metodo del sottoinsieme casuale che, nella formulazione di Ho, è un modo di implementare la discriminazione stocastica proposta da Eugene Kleinberg.

## Alberi di decisione

Un albero di decisione è l'elemento costitutivo di una foresta casuale ed è un modello intuitivo. Possiamo pensare a un albero di decisione come una serie di domande sì / no poste sui nostri dati che alla fine portano a una classe prevista (o valore continuo in caso di regressione). Questo è un modello interpretabile perché rende le classificazioni molto simili a noi: chiediamo una sequenza di domande sui dati disponibili che abbiamo fino a quando non arriviamo a una decisione (in un mondo ideale). I dettagli tecnici di un albero di decisione riguardano la modalità di formazione delle domande sui dati. Nell'algoritmo CART, viene creato un albero di decisione determinando le domande (dette suddivisioni dei nodi) che, una volta ottenuta la risposta, portano alla riduzione maggiore di Gini Impurity. Ciò significa che l'albero di decisione cerca di formare nodi contenenti un'alta percentuale di campioni (punti dati) da una singola classe trovando valori nelle caratteristiche che dividono i dati in modo pulito in classi.

La Gini Impurity di un nodo è la probabilità che un campione scelto casualmente in un nodo venga etichettato erroneamente se fosse etichettato dà la distribuzione di campioni nel nodo. L'equazione sottostante va ad indicare come trovare il valore della Gini Impurity di un nodo n:

$$I_G(n) = 1 - \sum_{i=1}^J (p_i)^2$$

Il valore della Gini Impurity di un nodo n è 1 meno la somma su tutte le classi J (per un'attività di classificazione binaria è 2) della frazione di esempi in ciascuna classe p\_i al quadrato.

## Random Forest

Il Random Forest è quindi un algoritmo di classificazione supervisionato. È un modello composto da molti alberi decisionali. Non si limita a fare una media della previsione degli alberi (che potremmo chiamare una "foresta"), ma inoltre, questo modello, utilizza due concetti chiave che gli danno il nome di casuale:

- Campionamento casuale di punti dati di allenamento durante la costruzione di alberi;
- Sottoinsiemi casuali di funzionalità considerate durante la divisione dei nodi.

Durante la fase di training, ogni albero della Random Forest impara da un campione casuale di punti dati. I campioni vengono disegnati con la sostituzione, nota come bootstrap, il che significa che alcuni campioni verranno utilizzati più volte in un singolo albero. L'idea è che addestrando ciascun albero su campioni diversi, sebbene ogni albero possa presentare una varianza elevata rispetto a una particolare serie di dati di training, nel complesso l'intera foresta avrà una varianza inferiore ma non aumenterà la distorsione.

Al momento del test, le previsioni vengono effettuate calcolando la media delle previsioni di ciascun albero decisionale. Questa procedura di formazione di ogni singolo allievo su diversi sottoinsiemi di dati bootstrap, e quindi la media delle previsioni, è nota come bagging, abbreviazione di aggregazione bootstrap.

L'altro concetto principale nel Random Forest è che viene considerato solo un sottoinsieme di tutte le funzionalità per suddividere ciascun nodo in ciascun albero decisionale. Generalmente questo è impostato su  $\sqrt{n\_features}$  per la classificazione, il che significa che se ci sono 16 caratteristiche, in ciascun nodo in ogni albero, saranno prese in considerazione solo 4 caratteristiche casuali per dividere il nodo. La foresta casuale può anche essere addestrata considerando tutte le funzionalità di ogni nodo come è comune nella regressione. Queste opzioni possono essere controllate nell'implementazione della foresta casuale di Scikit-Learn.

La differenza tra l'algoritmo Random Forest e l'algoritmo dell'albero decisionale è che: in Random Forest i processi di ricerca del nodo radice e suddivisione dei nodi funzione verranno eseguiti in modo casuale mentre nell'algoritmo dell'albero di decisione ho un solo albero che darà un solo risultato.

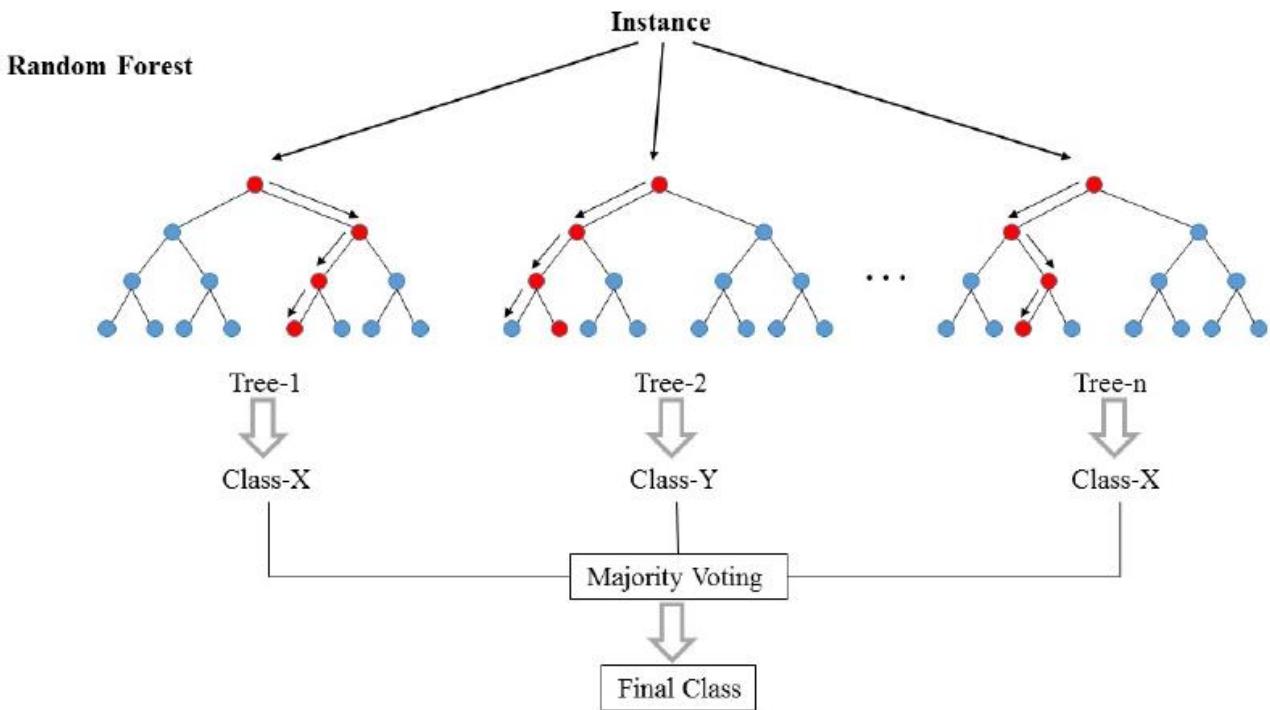
## Algoritmo

Ci sono due fasi nell'algoritmo Random Forest, una è la creazione di una foresta casuale, l'altra è fare una previsione dal classificatore Random Forest creato nella prima fase. L'intero processo è mostrato di seguito ed è illustrato in figura.

Pseudocodice di creazione della foresta casuale:

- a) seleziona casualmente le funzioni "k" dalle funzioni "m" totali in cui  $k \ll m$ ;
- b) tra le funzionalità "k", calcolare il nodo "d" utilizzando il miglior punto di divisione;
- c) dividi il nodo in nodi secondari usando la divisione migliore;
- d) ripetere i passaggi da a) a c) fino al raggiungimento del numero "l" di nodi;
- e) costruisci la foresta ripetendo i punti da a) a d) per "n" volte per creare "n" alberi.

La Figura seguente mostra il processo di selezione casuale delle funzionalità:



Nella fase successiva, una volta creato il classificatore Random Forest, può essere fatta la previsione. Il pseudocodice di previsione è il seguente:

- prendere le funzionalità di test e utilizzare le regole di ciascun albero decisionale creato casualmente per prevedere il risultato e memorizzare il risultato previsto (target);
- calcolare i voti per ciascun obiettivo previsto;
- considerare l'obiettivo previsto con il voto più alto come previsione finale dell'algoritmo Random Forest.

Da ciò si conclude che le Random Forest sono uno strumento di previsione molto efficace; data la Legge dei Grandi Numeri non andranno in overfitting; inoltre con la giusta dose di casualità la Random Forest è molto efficace come classificatore e regressore.

## Vantaggi dell'algoritmo

Rispetto ad altre tecniche di classificazione, ci sono tre vantaggi principali:

- per le applicazioni con problemi di classificazione, l'algoritmo Random Forest eviterà il problema di overfitting;
- sia per l'attività di classificazione che di regressione, è possibile utilizzare lo stesso algoritmo di foresta casuale;
- l'algoritmo Random Forest può essere utilizzato per identificare le funzionalità più importanti dal set di dati di training, in altre parole, ingegneria delle caratteristiche.

# 4 Il toolbox Sentinel-2 SNAP

## 4.1 Sentinel-2 Toolbox

La Toolbox per Sentinel-2 è un ambiente software che consiste in un set di strumenti di visualizzazione, analisi ed elaborazione di immagini Sentinel-2. Esso può integrare anche moduli analoghi per le altre tipologie di immagini dei satelliti ESA Copernicus, come quelle Sentinel-1 e Sentinel-3.

Come Toolbox di telerilevamento multi-missione, è anche di supporto alle precedenti missioni ESA Envisat, ERS, nonché i dati di terze parti come ad esempio di RapidEye, SPOT, MODIS, Landsat, e altri.

I vari tool possono essere eseguiti da un'intuitiva applicazione desktop oppure con un'interfaccia da linea di comando. Una ricca interfaccia di programmazione delle applicazioni consente lo sviluppo di plugin utilizzando Java o Python.

La Toolbox di Sentinel-2 è stata sviluppata per ESA da CS in partnership con Brockmann Consult, CS-Romania, Telespazio Vega Deutschland, INRA e UCL.

La prima versione di Sentinel-2 Toolbox è stata una toolbox autonoma destinata al processamento di dati ottici ad alta risoluzione. Sentinel-2 Toolbox fornisce tutti gli strumenti provenienti dal Toolbox ESA BEAM, oltre che un lettore di prodotti Sentinel-2, un lettore di prodotti SPOT da 1 a 5, e un lettore di prodotti RapidEye. In una versione successiva, tutti questi moduli sono inclusi nel comune Sentinel Application Platform (SNAP).

### Sentinel Toolbox: Dove trovarla?

Le Sentinel Toolbox e il codice sorgente completo sono distribuiti liberamente sotto la licenza GNU GPL. L'open source rende il software intrinsecamente indipendente da specifici fornitori, programmatore e fornitori. Il software può essere distribuito e condiviso liberamente da grandi comunità, incluso il codice sorgente e si può modificare liberamente. Ciò garantisce che non vi sia un'unica entità da cui dipende il futuro del software e ciò consente miglioramenti e messa a punto illimitati della qualità e della funzionalità del software. Come open source, l'evoluzione futura e la crescita delle Toolbox saranno possibili dalla comunità di utenti e sviluppatori che contribuiranno al progetto.

Il codice sorgente è disponibile su GitHub.

La Sentinel-2 Toolbox può essere scaricato da [Science Toolbox Exploitation Platform \(STEP\)](#).

Inoltre si può consultare il forum [Sentinel-2 toolbox forum](#) per contattare il Team di sviluppo, discutere processi e funzionalità con gli altri utenti e altro ancora.

### SNAP (SeNtinel Application Platform)

Un'architettura comune per tutti gli strumenti Sentinel è stata sviluppata congiuntamente da Brockmann Consult, Array Systems Computing e C-S chiamata Sentinel Application Platform (SNAP).

L'architettura SNAP è ideale per l'elaborazione e l'analisi dell'osservazione della Terra grazie alle seguenti innovazioni tecnologiche: estensibilità, portabilità, ricca piattaforma modulare dal lato

Client, astrazione di dati generici EO, gestione della memoria piastrellata e un framework di elaborazione dei grafici.

*Funzionalità principali:*

- Architettura comune per tutte le Toolboxes
- Visualizzazione delle immagini e navigazione molto veloce, anche delle immagini in gigapixel
- Graph Processing Framework (GPF): per la creazione di catene di elaborazione definite dall'utente
- La gestione avanzata dei livelli consente l'aggiunta e la manipolazione di nuove sovrapposizioni come immagini di altre bande, immagini da server WMS o shapefile ESRI
- Ricche definizioni della regione di interesse per statistiche e vari grafici
- Facile definizione e sovrapposizioni delle maschere di bit
- Aritmetica di banda flessibile che utilizza espressioni matematiche arbitrarie
- Accurata riproduzione e orto-rettificazione alle comuni proiezioni cartografiche
- Geo-codifica e rettifica mediante punti di controllo a terra
- Download automatico di SRTM DEM e selezione di riquadri
- Libreria di prodotti per la scansione e la catalogazione di archivi di grandi dimensioni molto efficiente
- Supporto per processori multithreading e multi-core
- Visualizzazione WorldWind integrata

Prodotti supportati: Sentinel-2 L1B, L1C, L2A and L3, Landsat-8, Spot 1 to Spot 7, Spot4 Take5, Spot5 Take5, RapidEye, Deimos

*Processori:*

- Sen2Cor per la correzione atmosferica
- Processore Sen2Three level-3 per la sintesi spazio-temporale del fondo dell'atmosfera corregge le immagini Sentinel-2 level 2°
- L2B processore biofisico (LAI, fAPAR, ...) di ATBD
- Riflettanza sul processore di radiosità
- Indici radiometrici:
  - Indici di vegetazione: DVI, RVI, PVI, IPVI, WDVI, TNDVI, GNDVI, GEMI, ARVI, NDI45, MTCI, MCARI, REIP, S2REP, IRECI, PSSRa
  - Indici del suolo: SAVI, TSAVI, MSAVI, MSAVI2, BI, BI2, RI, CI
  - Indici d'acqua: NDWI, NDWI2, MNDWI, NDPI, NDTI
- Processore IdePix: fornisce una classificazione dei pixel in proprietà come chiaro / nuvoloso, terra / acqua, neve, ghiaccio ecc.
- Processore MCI: utilizza l'indice di massima clorofilla sfruttando l'altezza di una misurazione su una linea di base specifica.
- Strumenti OTB: MultivariateAlterationDetector, Pansharpening-bayes, Pansharpening-lmvm, Pansharpening-rcs, Rasterization-image, Rasterization-manual, Segmentation-cc, Segmentation-meanshift, Segmentation-mprofiles, Segmentation-watershed ed SFSTextureExtraction.

## 4.2 Download dei dati e relativa importazione

Per poter utilizzare i prodotti Sentinel da importare nel programma SNAP vi è una politica sui dati libera, completa e aperta adottata per il programma Copernicus, la quale prevede l'accesso a tutti gli utenti per i prodotti di dati Sentinel, tramite una semplice auto registrazione.

Qui ci si può iscrivere self-registration.

Inoltre esistono due metodi per scaricare i dati in base alle proprie preferenze

### Metodo 1

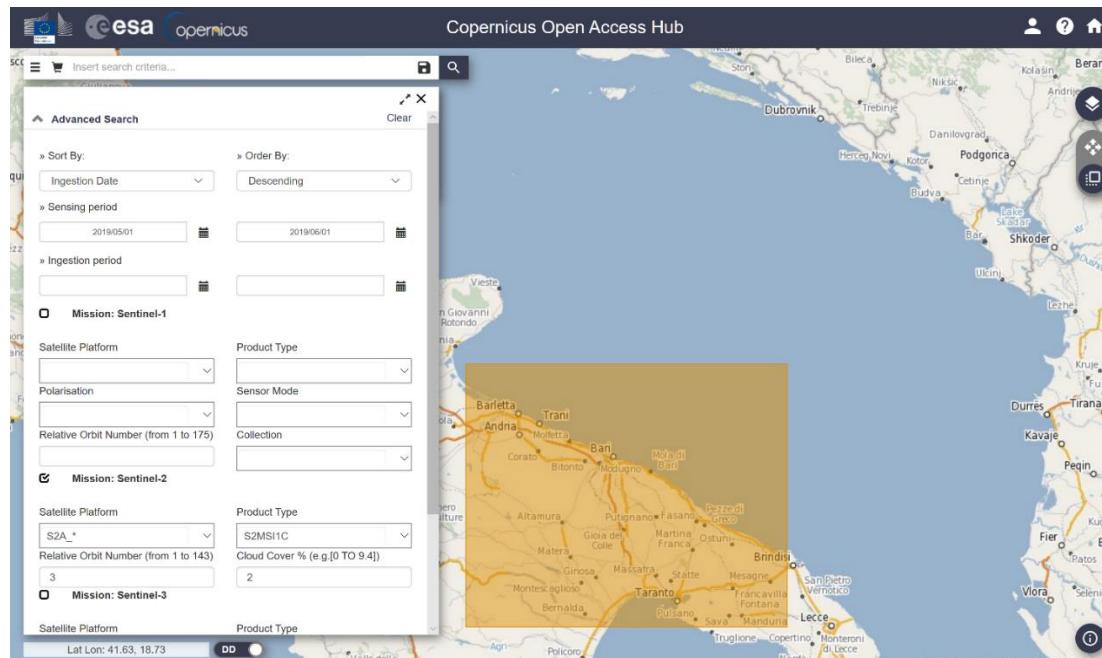
Una volta effettuata la registrazione, si accede al sito qui sotto:



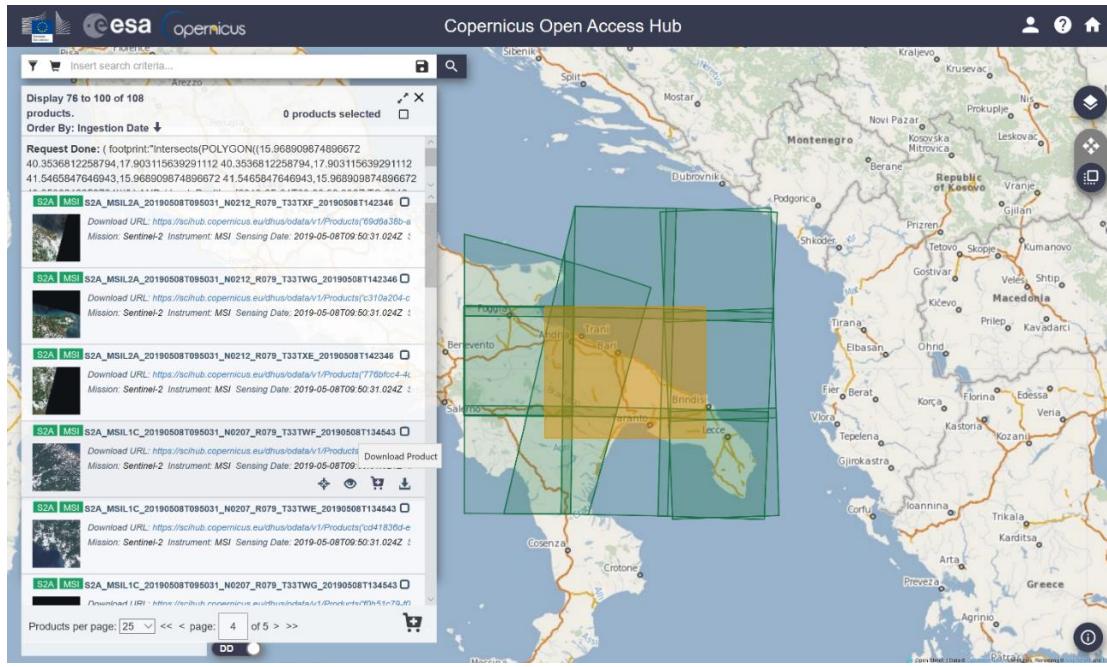
Il quale aprirà una mappa da cui si andrà a delimitare la zona che si vuole osservare su SNAP, inoltre è necessario effettuare il Login al sito.

Come primo parametro da inserire va selezionata la zona di rilevamento dei dati, poi dal menù laterale va selezionato la modalità di ordine dei risultati, bisogna poi indicare la data di inizio e fine del rilevamento dei dati.

Una volta effettuato ciò andranno indicate la tipologia di missione dal quale andare a prendere le immagini, nel nostro caso sarà Mission: Sentinel-2, e successivamente andremo ad indicare la piattaforma satellitare, il tipo di prodotto, il numero dell'orbita relativa (da 1 a 143) e la copertura di nuvolosità in percentuale % (ad es. [Da 0 a 9.4]):



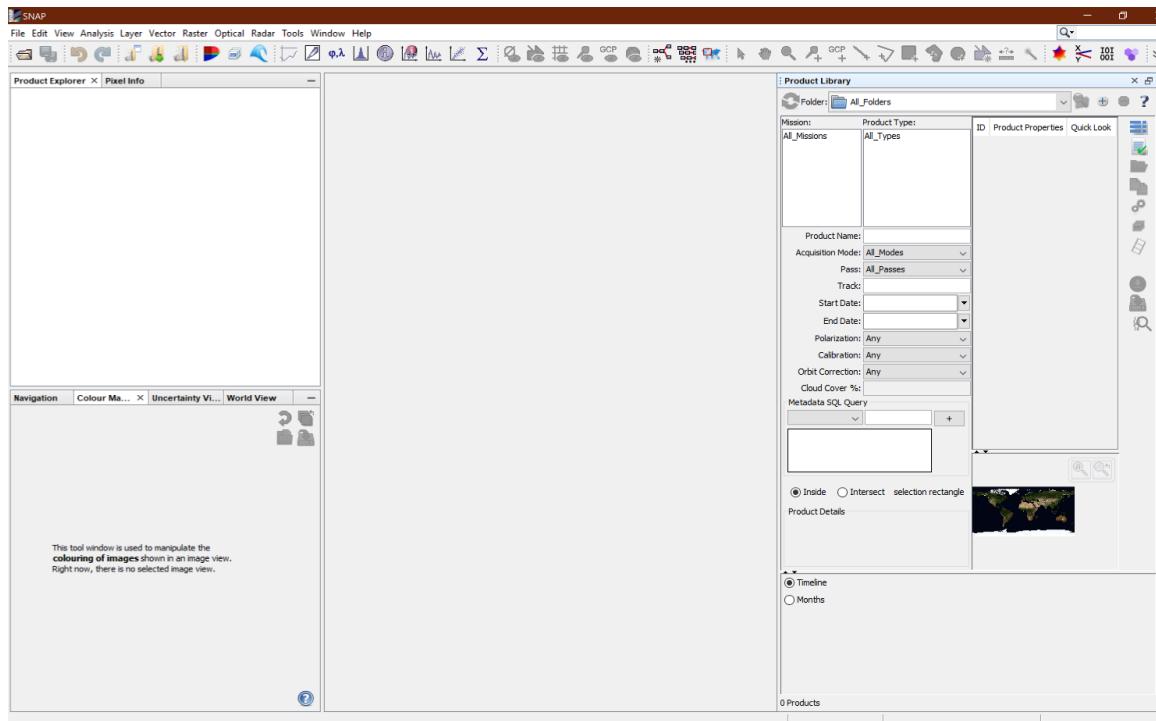
Dopo un breve caricamento verrà aperta una schermata laterale, in cui saranno elencate diverse immagini, e si potrà scegliere la più idonea alle proprie esigenze da scaricare, tramite Download Product, ed importare sul programma SNAP.



## Metodo 2

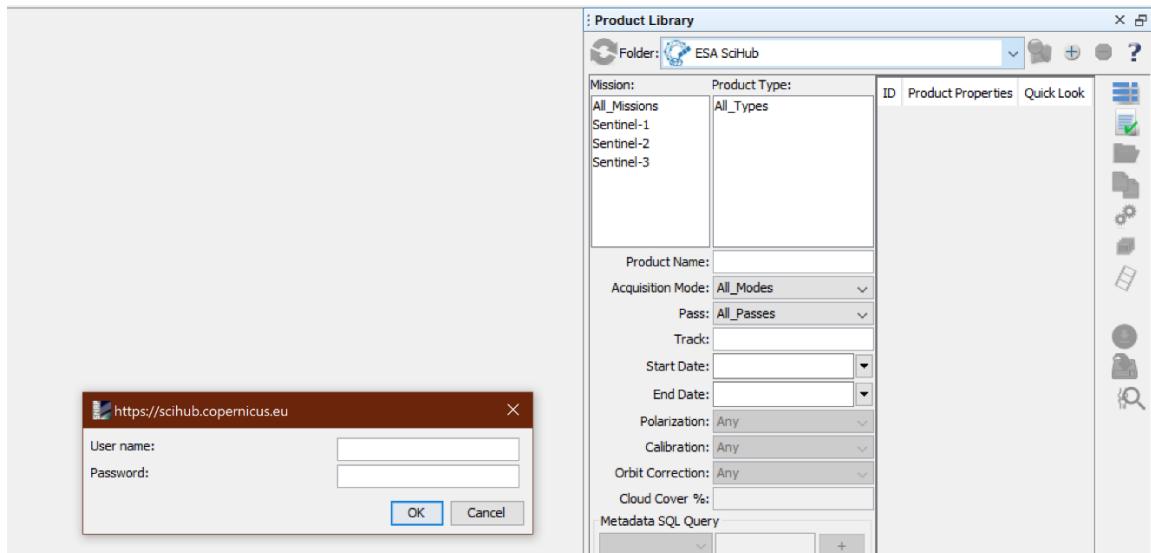
Il secondo metodo per scaricare i prodotti Sentinel consiste nell'utilizzare direttamente la piattaforma SNAP per il download.

Una volta aperta la schermata iniziale di SNAP va selezionata la voce laterale all'estrema sinistra “Product Library” la qual e aprirà la seguente schermata:



Nella voce “Folder”, bisogna passare da “All\_Folders”, che sta ad indicare la cartella contenente i dati Sentinel già scaricati, ad “ESA SciHub”.

Appena si effettua questo passaggio comparirà la casella di Login da compilare con i dati personali, ottenuti mediante l’auto-registrazione descritta in precedenza.

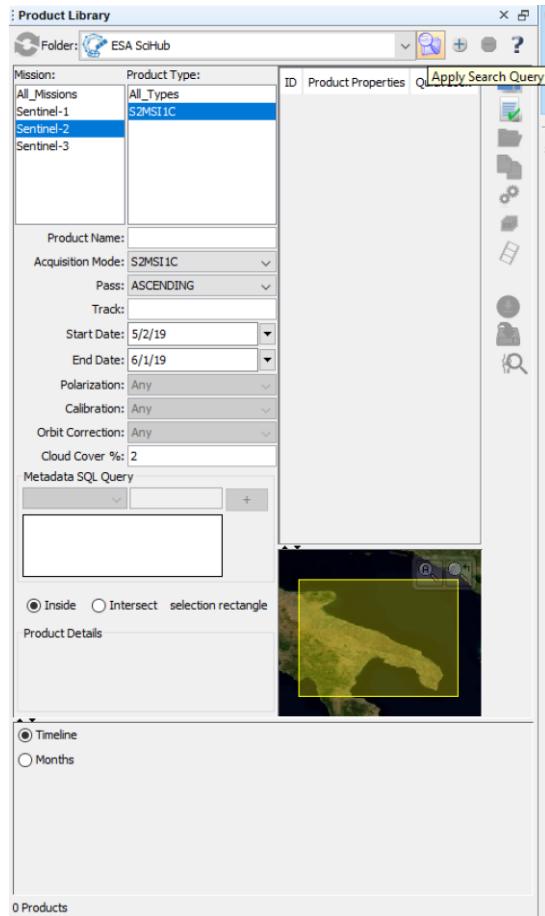


Successivamente va selezionato il tipo di missione dal quale andare a prendere le immagini, nel nostro caso sarà “Mission: Sentinel-2”, successivamente andremo ad indicare il tipo di prodotto, la modalità di acquisizione, la modalità di ordine dei risultati, bisognerà poi indicare la data di inizio e fine del rilevamento dei dati, la copertura di nuvolosità in percentuale % (ad es. [Da 0 a 9.4]) ed infine indicheremo la zona di rilevamento dei dati attraverso la mini mappa posta in basso a destra.

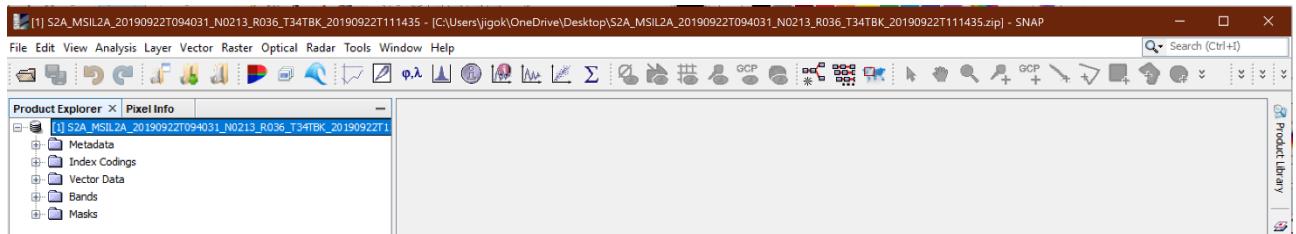
Inoltre è possibile selezionare se si vogliono rilevare i dati all’interno della zona selezionata o all’esterno utilizzando Inside o Intersect.

Una volta inseriti i parametri e formalizzata la query, bisognerà cercare dei risultati tramite il pulsante accanto alla casella “Folder”.

Di seguito è mostrato un esempio di inserimento dei parametri e di ricerca di risultati:



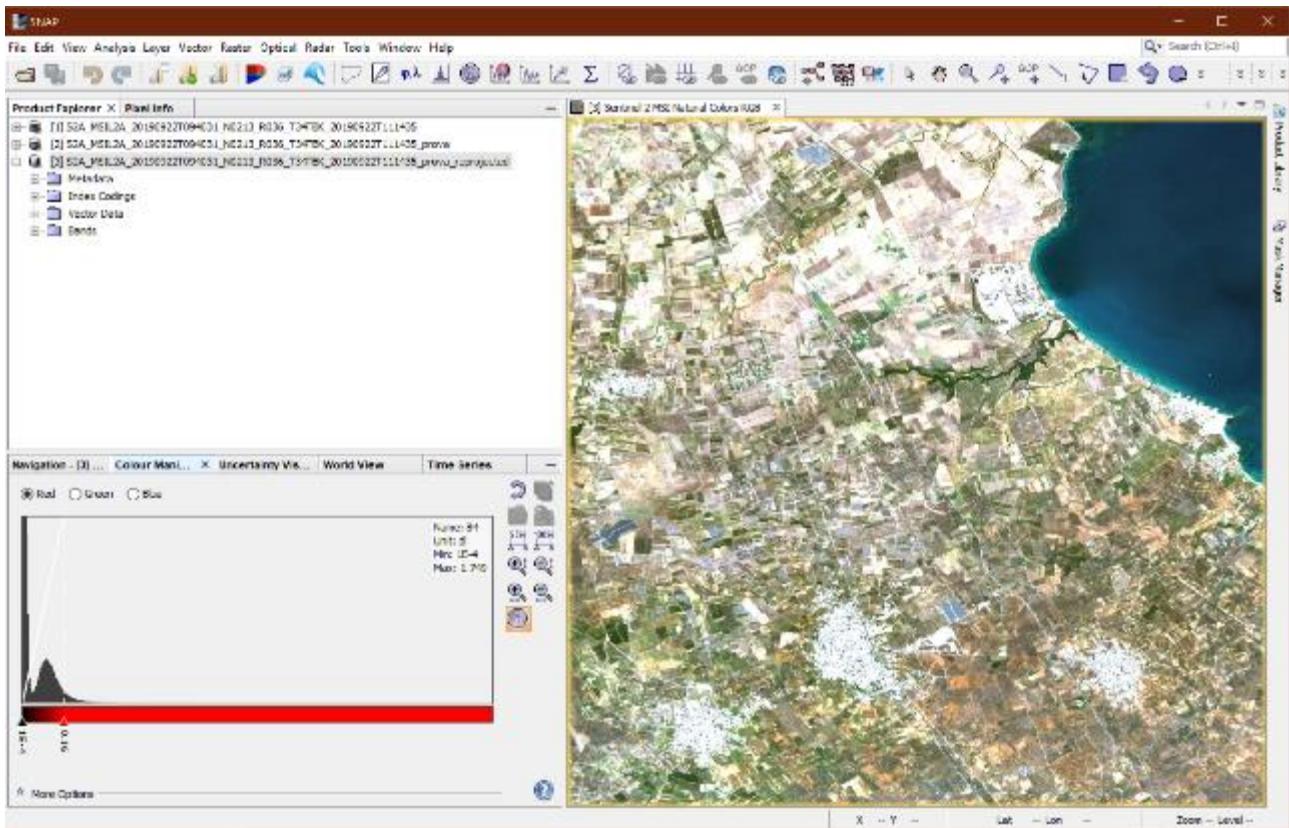
Dopo un breve caricamento compariranno i risultati che si potranno scaricare da Download ed in seguito compariranno nella parte all'estrema destra nella voce “Product Explorer”.



## 4.3 Analisi dell'immagine

Per creare la mappa RGB, si utilizza il comando Open RGB Image View, semplicemente effettuando tasto destro sull'ultimo progetto creato.

Successivamente comparirà una casella in cui saranno inserite le nostre 3 bande per la creazione dell'immagine.

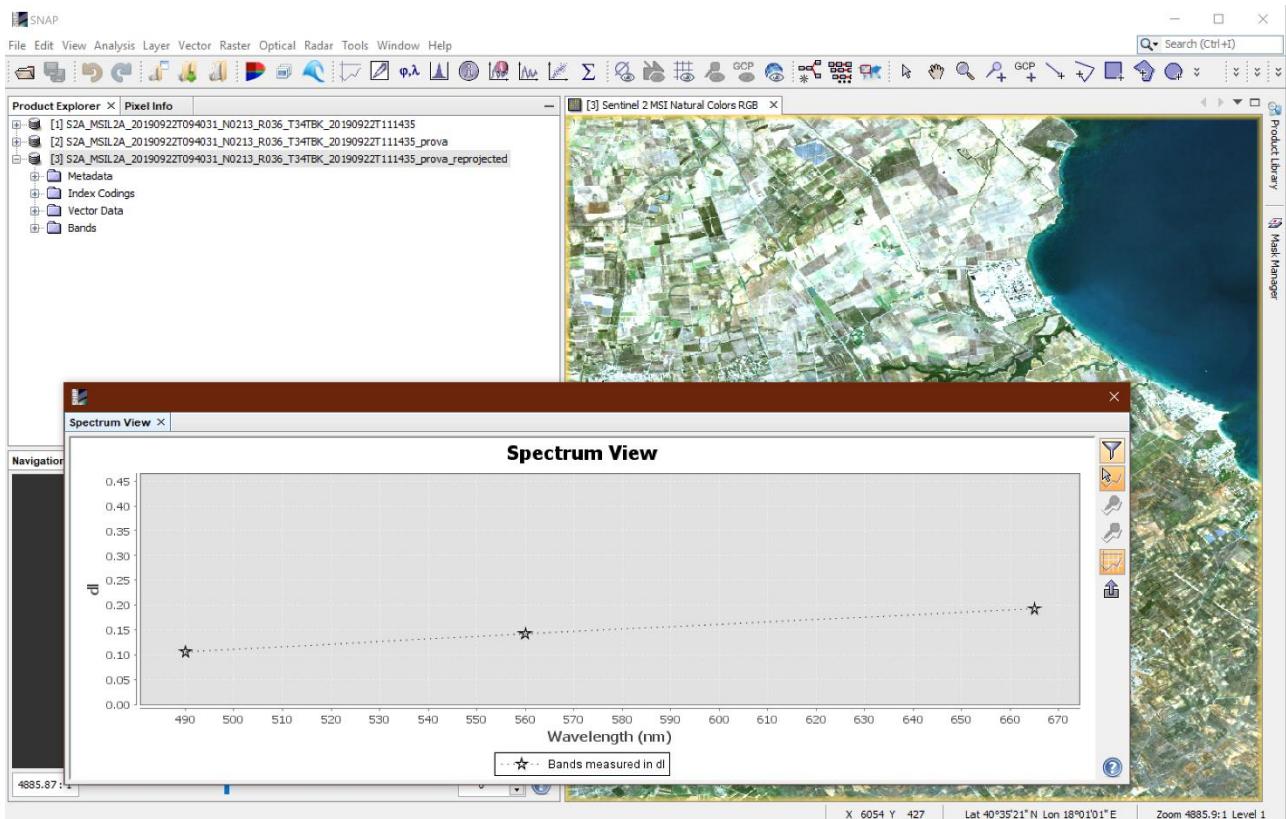


In basso a sinistra si può notare la Color Manipulation dei tre colori delle rispettive bande e la loro percentuale di colore all'interno della mappa, questa è possibile modificarla a piacimento.

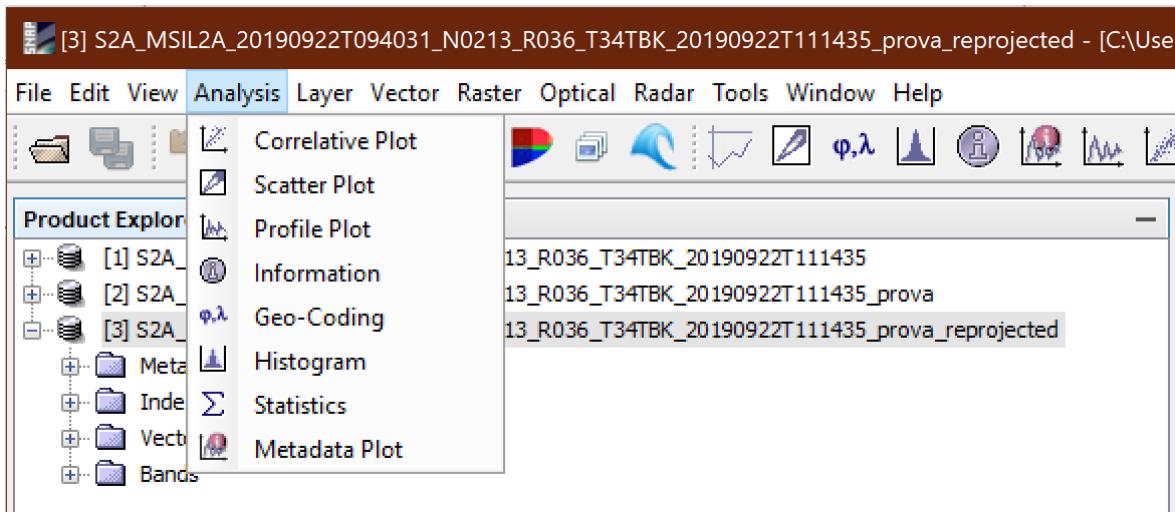
Inoltre è possibile spostarsi nella mappa tramite la schermata di Navigation o direttamente sull'immagine RGB utilizzando il mouse.

SNAP fornisce anche altri tools utili per l'analisi delle immagini. Se è stato aperto un prodotto che contiene diverse bande spettrali, è possibile utilizzare la Spectrum View per ispezionare gli spettri per una determinata posizione di pixel.

Se ora si sposta il cursore del mouse sull'immagine RGB, la Spectrum View visualizzerà gli spettri per la posizione attuale del pixel. Si noti che la vista spettro può solo visualizzare le intensità delle bande spettrali. La Spectrum View resterà vuota per i prodotti senza bande spettrali.



Nell’immagine sottostante sono elencati le varie tipologie di operazioni disponibili per l’analisi dettagliata dell’immagine disponibili su SNAP.



## Correlative Plot

Questa finestra di dialogo consente di tracciare una variabile importata (ad es. Una traccia navale) su una variabile Raster. Per tracciarle, queste variabili devono essere già importate nel prodotto che contiene i dati Raster correlati. È possibile selezionare una maschera ROI definita nuova o esistente per filtrare i dati da visualizzare sul grafico.

## **Scatter Plot**

Per ogni banda, l'intervallo di valori è uniformemente diviso in 512 bin. Lo Scatter Plot mostra questi contenitori tracciati uno contro l'altro, risultando in  $512 * 512$  celle. Per ogni cella, il numero di pixel all'interno degli intervalli di valori delle bande corrispondenti viene conteggiato e codificato a colori. Il giallo indica grandi quantità di pixel, il nero per quelli piccoli. Se nessun pixel è assegnato a una cella distinta, la cella non riceverà alcun colore. La codifica a colori può essere invertita utilizzando la casella di controllo "Inverti colori di stampa".

## **Profile Plot**

Questo strumento è disponibile solo se nel prodotto sono presenti dati vettoriali. Quindi, il grafico dei profili può operare in due modalità, la modalità classica basata su una forma geometrica e la modalità correlativa che consente all'utente di confrontare i dati satellitari con i dati in-situ.

## **Information**

Questa sezione mostra le proprietà generali di un prodotto caricato, una griglia a banda o tie-point e il relativo prodotto principale.

## **Geo-Coding**

Questa finestra di dialogo, mostra le informazioni di geo-codifica per il prodotto dati selezionato. La codifica geografica consente a Sentinel Toolbox di trasformare le coordinate dei pixel in coordinate geografiche e viceversa. La codifica geografica può essere basata su una proiezione cartografica (il prodotto è georeferenziato) o sulla base di griglie del punto di congiunzione (il prodotto è geo codificato). Se un prodotto non ha riferimenti geografici, la casella degli strumenti Sentinel utilizza le griglie dei punti di congiunzione "latitudine" e "longitudine" per la codifica geografica.

Per i geo-codificatori basati su griglia di punti di congiunzione, la trasformazione di una coordinata geografica in una posizione in pixel è più complicata rispetto al contrario. Sentinel Toolbox utilizza un algoritmo iterativo o un'approssimazione polinomiale a seconda dell'errore quadratico medio radice (RMSE) dell'approssimazione. Se RMSE si trova al di sotto di mezzo pixel, viene utilizzata l'approssimazione invece dell'iterazione perché quest'ultima a volte non può avere un chiaro punto di attrazione e cederebbe ad un loop infinito.

## **Histogram**

L'intervallo di valori dei dati di immagine geofisica è uniformemente diviso in bin. L'istogramma mostra quanti pixel dell'immagine cadono in ciascuno dei contenitori.

## **Statistics**

Questa sezione visualizza informazioni statistiche sulla banda attualmente selezionata. La sezione "Statistical accuracy" consente all'utente di specificare l'accuratezza con cui deve essere calcolato l'istogramma. Numeri più alti indicano una maggiore precisione ma possono portare a tempi di calcolo considerevolmente più lunghi. Più precisamente, questo parametro determina il numero di celle bin utilizzate per calcolare l'istogramma: il numero di celle bin viene calcolato da  $10^a$  con un parametro. Poiché le soglie della mediana e della percentuale sono derivate dall'istogramma, anche l'accuratezza influenza queste misure. Il valore del parametro deve essere compreso tra 0 (istogramma con un solo bin) e 6 (istogramma con 1.000.000 di bin). Quando la precisione automatica è abilitata, è impostata su 3.

## Metadata Plot

Questo grafico può essere utilizzato per tracciare campi dei set di dati dei metadati l'uno contro l'altro. I campi devono essere valori scalari o array. Se sono disponibili più record per un set di dati di metadati, il campo può essere tracciato su tutti i record. Un clic con il tasto destro sul diagramma aprirà un menu contestuale con alcune opzioni aggiuntive.

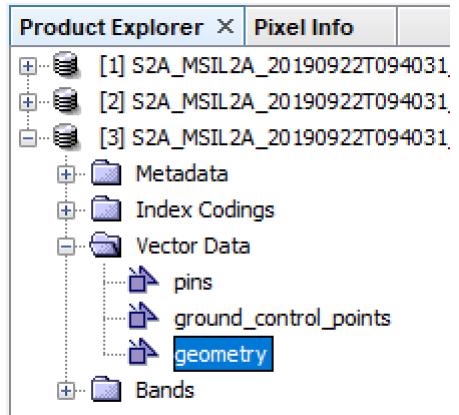
## 4.4 Creazione e gestione dei Vector Data Container

Oltre alle bande e alle griglie del punto di ancoraggio contenute nei prodotti di dati satellitari - che sono tutti i tipi di dati Raster - il modello di prodotto SNAP include anche dati vettoriali. Pertanto, i dati vettoriali vengono archiviati e ripristinati dai file di tipo BEAM-DIMAP. Un prodotto può comprendere qualsiasi numero di contenitori di dati vettoriali con un nome. Ogni contenitore può comprendere virtualmente un numero qualsiasi di nodi di dati vettoriali. Esistono due ricorrenze generali di dati vettoriali nella Sentinel Toolbox, Geometry e Data Correlative.

### Geometry

Le Geometry sono punti, linee, multilinea, poligoni o multi poligoni. Le nuove Geometry vengono sempre aggiunte al contenitore “geometry” selezionato nel nodo “Vector Data” della finestra dello strumento “Product Explorer”, come mostrato nella schermata. Se non esiste ancora un “Vector Data Container”, SNAP creerà un contenitore predefinito denominato “geometry”. (Questo comportamento predefinito può essere personalizzato nel file di configurazione di Sentinel Toolbox).

Inoltre si può creare un proprio contenitore tramite la barra di comando in alto denominata “Vector” e successivamente “New Vector Data Container”, in questo modo si può dare un nome e una descrizione al Vector Data più specifici.

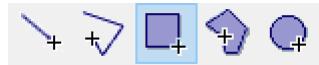


Le Geometry in un contenitore di geometria possono essere utilizzate direttamente come ROI per l'analisi dei dati Raster. Una volta aggiunto un nuovo contenitore al prodotto dati, viene creata una maschera geometrica associata eseguendo il rendering della geometria sulla griglia di dati Raster intrinsecando del prodotto.

La maschera associata avrà sempre lo stesso nome del contenitore della geometria che l'ha creata e può fungere da possibile maschera per la banda selezionata o la griglia di punti di raccordo senza alcuna interazione aggiuntiva da parte dell'utente. Una volta creata la geometria, ad esempio semplicemente disegnandola, la relativa maschera geometrica può essere utilizzata come ROI nei vari strumenti di analisi, come le finestre degli strumenti Statistics, Histogram e Scatter Plot

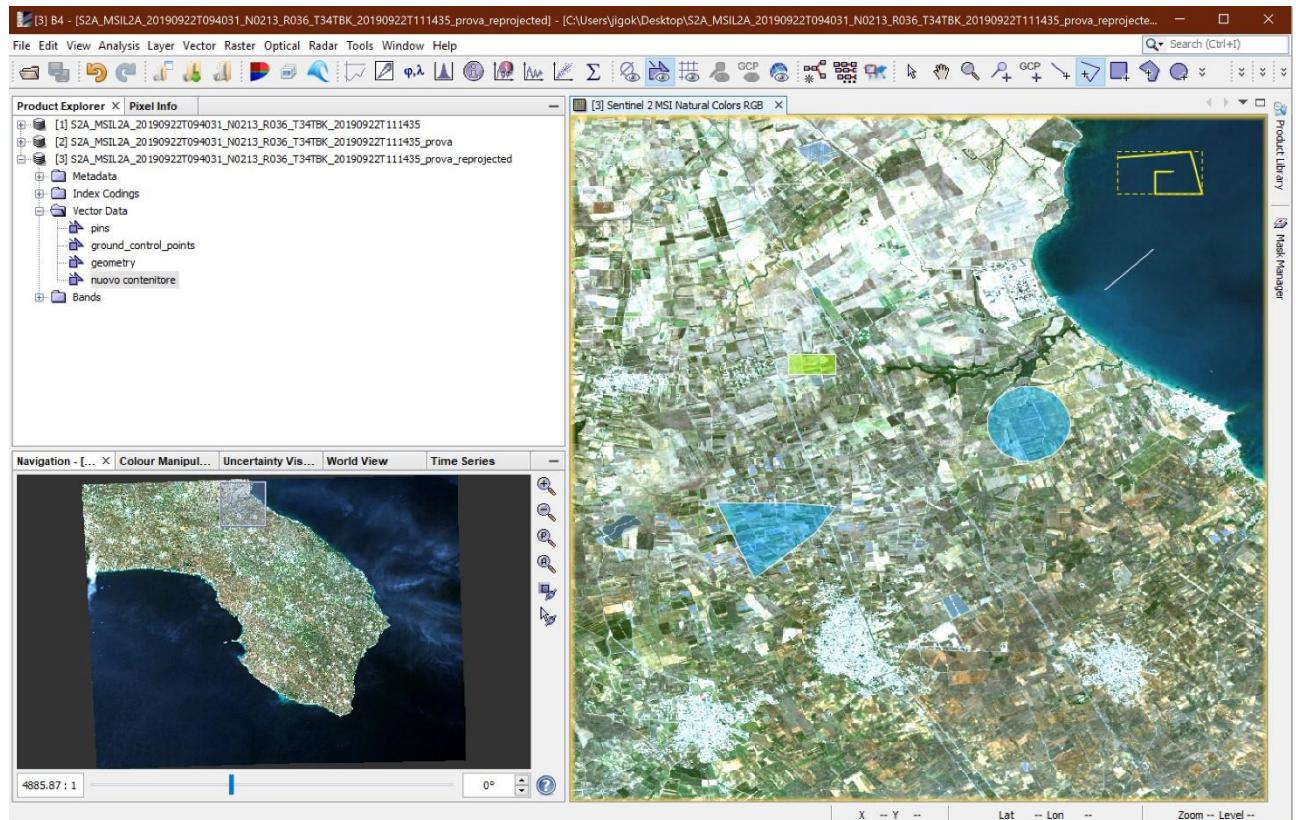
## Creazione della nuova Geometry

Una volta aperta una Image View, vengono create nuove Geometry utilizzando i vari strumenti di disegno forniti dalla Casella degli strumenti Sentinel tramite la Barra delle interazioni:



- Linea: premere il tasto sinistro del mouse per il punto iniziale, trascinare la linea sul punto finale e rilasciare il tasto sinistro del mouse.
- Polilinea: fare clic con il pulsante sinistro del mouse (premere e rilasciare) per il punto iniziale, spostare il segmento di linea e fare clic per aggiungere un punto di vertice, spostarsi al punto finale e fare doppio clic per finalizzare la polilinea.
- Rettangolo: premere il pulsante sinistro del mouse per il punto iniziale, trascinare la linea sul punto finale e rilasciare il pulsante sinistro del mouse.
- Poligono: simile alla polilinea; Fare clic con il pulsante sinistro del mouse (premere e rilasciare) per il punto iniziale, spostare il segmento di linea e fare clic per aggiungere un punto di vertice, passare al punto finale e fare doppio clic per chiudere il poligono.
- Ellisse: simile al rettangolo; Premere il pulsante sinistro del mouse per il punto iniziale, trascinare la linea sul punto finale e rilasciare il pulsante sinistro del mouse.

Qui vi è un esempio di nuove geometrie create su una mappa:



Una volta create le nuove geometrie è bene salvare il nostro lavoro in modo da evitare errori.

## **5 Descrizione dei dati sperimentali**

I dati sperimentali riportati in questo capitolo consistono nelle immagini Sentinel-2 e i dati di verità a terra, che in questo caso sono gli esempi delle classi delle colture maggiormente presenti nel territorio e che sono considerate per effettuare la classificazione. Queste classi includono sia tipologie di colture stagionali che classi di copertura del suolo (foreste, urbano, ecc.).

### **5.1 Immagini Sentinel-2**

Le immagini Sentinel-2 scelte in questo studio sono quelle acquisite ad Aprile e Luglio del 2018, per poter ottenere una classificazione delle colture sia primaverili (ad esempio il grano) che estivi (ad esempio il pomodoro). La scelta della data esatta tra quelle disponibili viene fatta in base all'assenza di nuvole, che diversamente offuscano la scena acquisita dal satellite.

#### **Immagine del 13 Aprile 2018**

L'immagine Sentinel-2 utilizzata per la prima classificazione è stata quella acquisita il 13 Aprile 2018. Per coprire l'area del Tavoliere, sono state utilizzate due Tiles di Sentinel-2 di 100km x 100km, ovvero quella identificata con la label T33TWG (che copre Foggia nord) e quella T33TWF (che copre Foggia sud).

Per avere un'unica immagine su tutta l'area nella classificazione, ovvero il Tavoliere Pugliese, è necessario effettuare con SNAP una operazione di mosaicatura e di ritaglio, il cui risultato è mostrato nella Figura 8.

#### **Immagine del 02 Luglio 2018**

L'immagine Sentinel-2 per la seconda classificazione è stata acquisita il 02 Luglio 2018, ed anche in questo caso sono state utilizzate le due Tiles T33TWG (che copre Foggia nord) e quella T33TWF (che copre Foggia sud).

Analogamente, le immagini sono state mosaicate e ritagliate per ottenere un'unica immagine sul Tavoliere Pugliese, come mostrato in Figura 9.

Come si può notare confrontando le due immagini, il colore verde che rappresenta la vegetazione, è più diffuso nell'immagine di Aprile rispetto a quella di Luglio, poiché nell'area del tavoliere c'è una considerevole presenza di grano, che in Aprile è nel pieno della suo sviluppo.



**Figura 8: Immagine RGB Sentinel-2 del 13 Aprile 2018**



**Figura 9: Immagine RGB Sentinel-2 del 02 Luglio 2018**

## 5.2 Classi delle colture

### Dati di Aprile 2018

Come verità a terra delle classi presenti in Aprile 2018, sono stati utilizzati degli esempi di campi coltivati con diverse tipologie di colture o altri tipi di copertura del suolo su delle aree geografiche note. Queste informazioni di classi e corrispondenti posizioni geografiche sono disponibili grazie ai poligoni georeferenziati (shapefile) estratti dal database del CREA. Questi poligoni sono stati sovrapposti sull'immagine mosaicata di Aprile.

Le classi individuate in Aprile sono 19 e sono elencate nella seguente tabella, insieme alla numerazione associata.

Classe	S_2018_LU_F1_cal_D
Carciofo	1
Asparago	2
Broccoli/Cavolfiore	3
Foresta decidua	4
Grano duro	5
Foresta sempreverde	6
Prateria	7
Olivo	8
Pesca	9
Pannelli fotovoltaici	10
Arbusti	11
Urbano	12
Vite	13
Pomodoro	14
Cece	15
Suolo nudo	16
Orzo	17
Favetta	18
Spinacio	19

## Dati di Luglio 2018

Analogamente, per le classi presenti a Luglio 2018, sono stati utilizzati degli esempi di campi coltivati o altre tipologie di copertura del suolo su delle aree geografiche note, anch'essi estratti dal database del CREA. Questi poligoni sono stati sovrapposti sull'immagine mosaicata di Luglio.

A Luglio, le classi individuate per le colture di aprile sono 15 e sono elencate nella seguente tabella, insieme alla numerazione associata.

Classe	S_2018_LU_F2_cal_D
Carciofo	1
Asparago	2
Broccoli/Cavolfiore	3
Foresta decidua	4
Grano duro raccolto	5
Foresta sempreverde	6
Prateria	7
Olivo	8
Pesca	9
Pannelli fotovoltaici	10
Arbusti	11
Urbano	12
Vite	13
Pomodoro	14
Finocchio	15

# 6 Esperimenti di classificazione

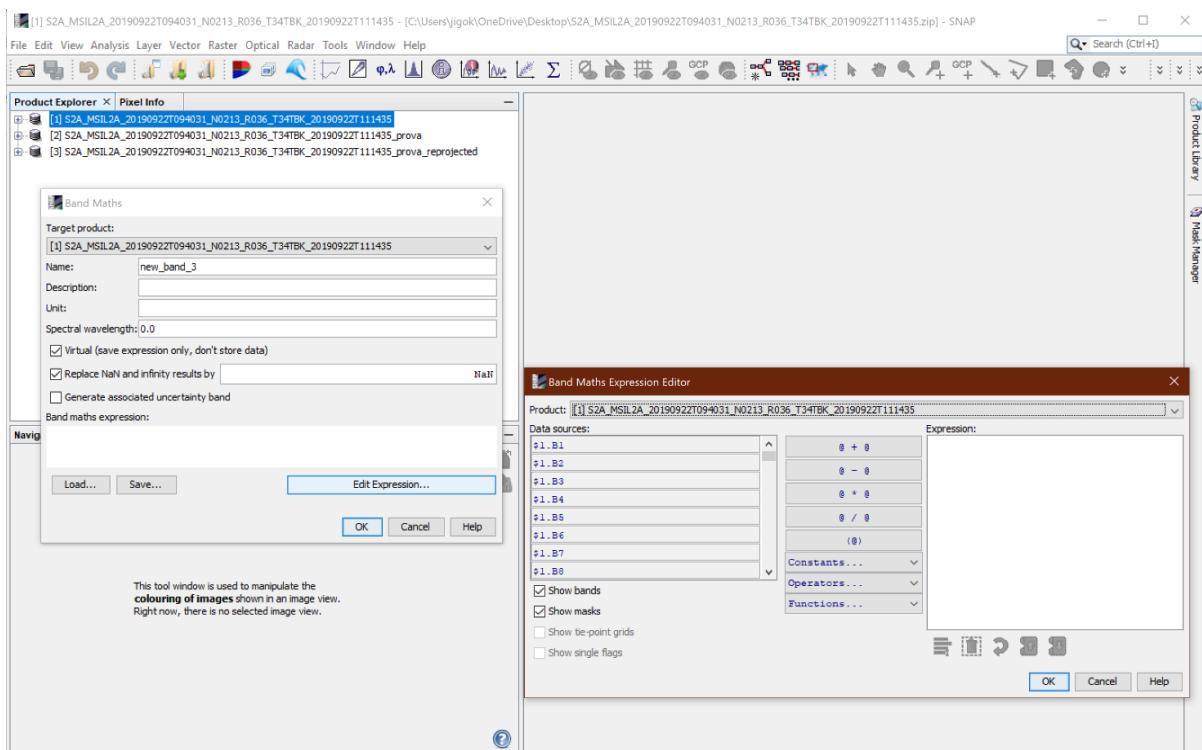
Gli esperimenti di classificazione di Sentinel-2 per ottenere delle mappe delle colture richiedono: un pre-processing dei dati, l'applicazione dell'algoritmo di classificazione sui dati pre-processati ed infine, anche un post-processing per migliorare la qualità della mappa di output. Nella sotto-sezione che descrive la classificazione, dapprima viene descritta la procedura, e poi vengono riportati gli esperimenti di classificazione effettuati con i dati disponibili, ovvero le classificazioni delle immagini Sentinel-2 di Aprile 2018 e Luglio 2018.

## 6.1 Pre-processing dei dati

Il pre-processing dei dati consiste nell'applicare le n operazioni di Resampling, Mosaicing, Reprojection e Masking. Lo scopo è ottenere una pila di immagini multi-banda tutte tra loro co-registrate.

### Band Maths

Lo strumento Band Maths viene utilizzato per creare nuovi valori di esempio di immagini derivati da bande, griglie e flag di punti esistenti. I dati di origine possono provenire da tutti i prodotti di input attualmente aperti e spazialmente compatibili. I dati di input sono combinati da espressioni matematiche arbitrarie per generare i dati di output. Per impostazione predefinita, viene automaticamente aperta una nuova immagine per i nuovi valori di esempio. È possibile disabilitare questo comportamento nella finestra di dialogo delle preferenze. Fare riferimento alla documentazione di Maths Expression Editor per la sintassi e le capacità delle espressioni nella sezione Band Maths. Dopo che la nuova band è stata creata (o una esistente è stata sovrascritta), è possibile passare a Product Explorer per aprire una vista immagine al fine di ispezionare i campioni risultanti.



Una volta completata questa operazione è bene salvare il progetto in modo da evitare eventuali errori.

## Resampling

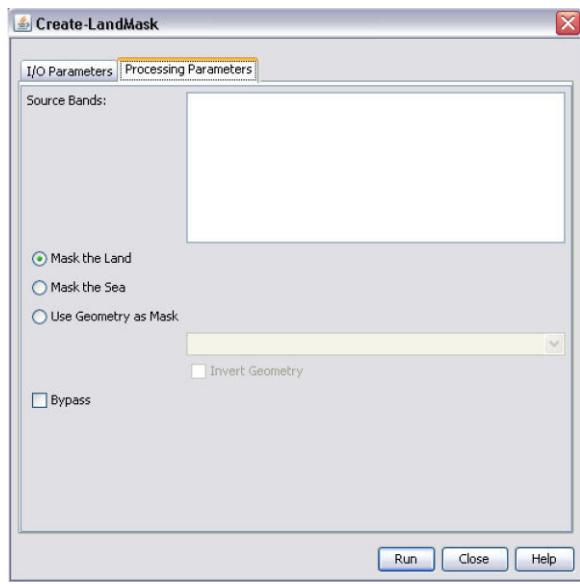
Con il Resampling un prodotto Raster multi-banda con dimensioni multiple (ovvero avente bande con risoluzioni di pixel differenti) viene ricampionato per ottenere un prodotto Raster avente tutte le bande con dimensioni omogenee (ovvero tutte le bande con la stessa risoluzione di pixel). Ciò è necessario per poter applicare funzioni SNAP che non supportano un prodotto di dimensioni multiple. Dopo aver creato il nuovo prodotto, è possibile passare a Product Explorer per aprire una vista immagine per una banda del nuovo prodotto.

Con il ricampionamento, i centri dei vari pixel del prodotto target generalmente non corrispondono ai centri dei pixel del prodotto di input.

Il Resampling dà luogo al processo di determinazione e interpolazione dei pixel dal prodotto sorgente al prodotto target. Gli effetti del Resampling sono particolarmente visibili se i pixel nel prodotto target sono più grandi dei pixel di origine. SNAP fornisce tre diversi metodi di Resampling: Nearest Neighbour, Bi-linear Interpolation, Cubic Convolution.

## Masking

L'operazione di Masking permette di applicare una maschera alle immagini. Se di terra, trasformerà qualsiasi pixel a terra in valori nulli. Se la casella di controllo "preserva terra" è impostata su true, tutti i pixel terrestri saranno conservati e tutti i pixel oceanici saranno impostati su valore nullo. Questa viene utilizzata per determinare rapidamente se un pixel è a terra o nell'oceano. In alternativa, è possibile utilizzare anche una geometria fornita del prodotto. Inoltre un utente potrebbe creare ROI o importare uno Shape file.

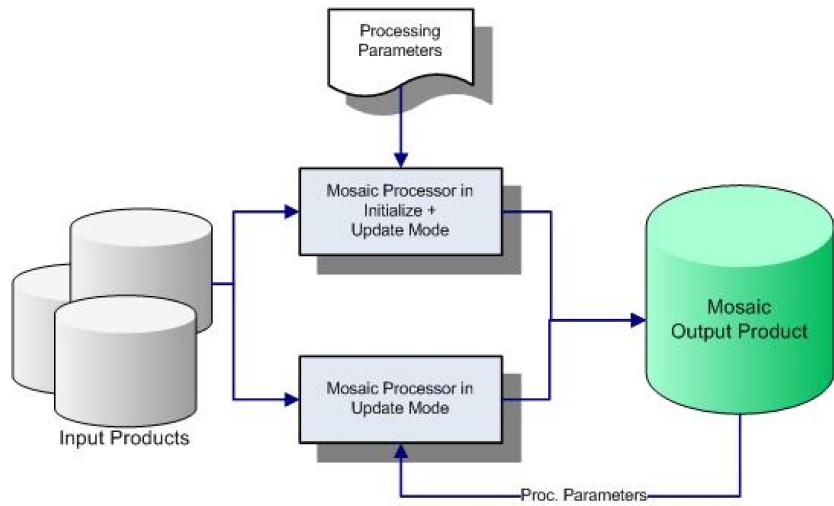


L'operatore utilizza i seguenti parametri:

- Banda di origine: tutte le bande (reali o virtuali) del prodotto di origine. L'utente può selezionare una o più bande.
- Maschera la terra: casella di controllo che indica che i pixel della terra diventeranno valore di nodata.
- Maschera mare: casella di controllo che indica che i pixel del mare diventeranno valore di nodata.
- Usa Geometria come maschera: selezionare una geometria o una ROI dal prodotto da utilizzare come maschera. Qualunque cosa al di fuori dell'area avrà valore nodata.
- Inverti geometria: qualsiasi cosa all'interno della ROI o della geometria avrà un valore nodata.

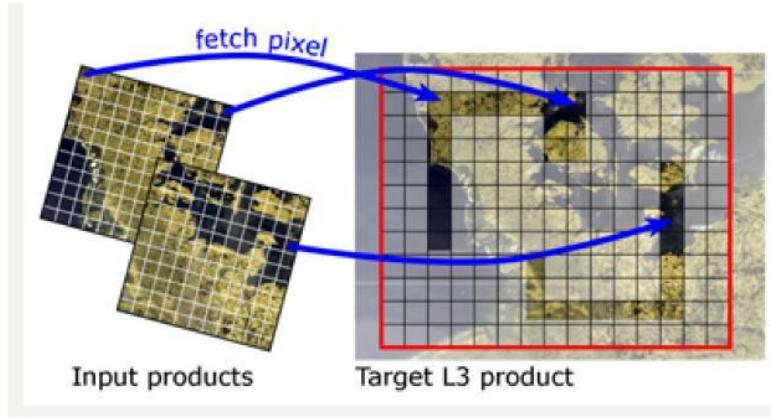
## Mosaicing

La seguente figura mostra il flusso di dati all'interno del processore Mosaic:



In una prima fase, il prodotto di output viene dimensionato utilizzando le informazioni di Riproiezione calcolate dai metadati stimando i parametri quali la dimensione dei pixel, il numero dei pixels e loro coordinate. Nella seconda fase, il prodotto di output viene “riempito” con i valori dei prodotti di input.

In pratica il processore Mosaic esegue il loop su tutte le celle della griglia di destinazione e determina quale pixel di un prodotto sorgente può essere usato per “riempirlo”. Questo viene fatto cercando il pixel vicino più vicino nel prodotto di input che contiene la coordinata del centro geografico del pixel di output corrente. Come promemoria: il processore di binning di livello 3 prende tutti i pixel di input e li ordina in contenitori definiti dal prodotto di output. Il processo è rappresentato dalla seguente immagine:



Il rettangolo rosso segna il confine geografico del prodotto di livello 3; le linee nere sono celle della griglia target. Si noti che i prodotti di input possono avere dimensioni Raster diverse dal prodotto di livello 3 (linee bianche).

## Output

Le bande dei prodotti di output sono definite come variabili di output. I valori dei pixel vengono accumulati utilizzando la media aritmetica di tutti i pixel di input validi di tutti i prodotti di input

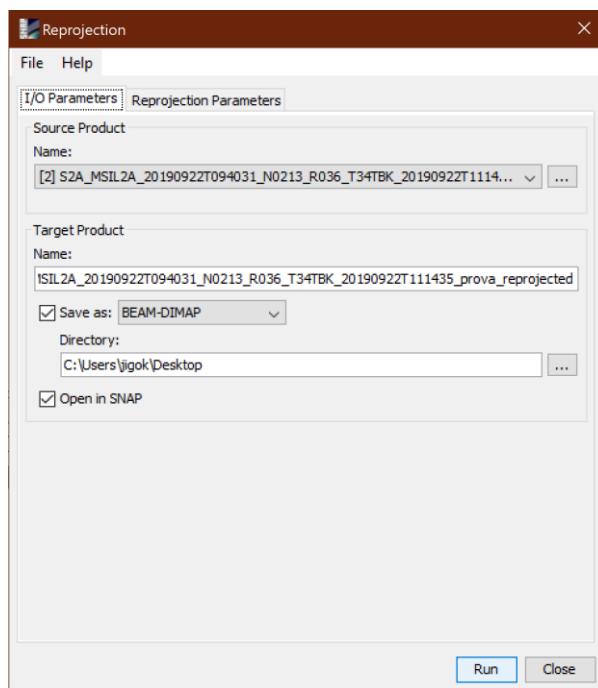
trovati per un determinato pixel di output. Una variabile di output ha un nome e un'espressione che comprende operatori matematici, funzioni e simboli che fanno riferimento ai flag e alle bande definiti nei prodotti di input.

La validità o meno di un valore di pixel nel prodotto di output è determinata da una serie di criteri di pixel validi. Un criterio pixel valido è un'espressione booleana che comprende operatori matematici, funzioni e simboli che fanno riferimento ai flag e alle bande definiti nei prodotti di input. Il processore del mosaico può accettare qualsiasi numero di criteri pixel validi che possono essere combinati AND o OR. Se non viene fornito alcun criterio di pixel valido, si presuppone che tutti i pixel siano validi.

## Reprojection

È anche possibile creare un nuovo prodotto con un sistema di riferimento di coordinate (CRS) progettato in precedenza. Se si desidera che anche il prodotto creato sia orto rettificato, bisogna utilizzare il comando Ortorectificazione.

Dopo aver creato il nuovo prodotto, è possibile passare a Product Explorer per aprire una vista immagine per una banda del nuovo prodotto.



## 6.2 Classificazione

La classificazione supervisionata è una tecnica che ha come obiettivo la classificazione dei pixel di un'immagine in diverse classi in base alle loro caratteristiche.

Esistono due fasi: fase di addestramento, ovvero fase di Training, e fase di classificazione. La fase di addestramento consiste nell'utilizzare una serie di campioni di addestramento, costituiti dai vettori di dati estratti dalle immagini (ogni vettore è costituito dal pixel multi-dimensionale di valori nelle varie bande spettrali) e i corrispondenti valori di classe a cui appartengono, per determinare i parametri del classificatore. La fase di classificazione consiste nell'applicare i parametri del classificatore a tutti i pixel dell'immagine per ottenere la mappa classificata.

L'utente in SNAP può scegliere di addestrare (e salvare) un classificatore; o caricare un classificatore precedentemente salvato per eseguire la classificazione.

### Addestramento e creazione di un classificatore in SNAP

L'utente è tenuto a fornire lo shapefile delle aree ove sono definite le classi ed il numero di campioni di addestramento da utilizzare. Oltre alla formazione, viene eseguita anche la valutazione del classificatore. Ad esempio, se il numero di campioni di addestramento da utilizzare è 5000, verranno estratti  $5000 \times 2 = 10.000$  campioni. I primi 5000 verranno utilizzati come campioni di addestramento e i restanti 5000 verranno utilizzati come campioni per la validazione. Vengono creati due file separati: uno con estensione .class e uno con estensione .xml. I risultati della valutazione sono in un file con estensione .txt. L'utente può scegliere di fare il l'addestramento su dati Raster o vettoriali.

### Addestramento su un dato Raster

L'utente può scegliere una banda (dal primo prodotto elencato nel ProductSet-Reader) come banda di addestramento. Se non ne viene scelto nessuna, verrà utilizzata la prima banda. L'utente può anche scegliere le bande da tutti i prodotti di origine come bande di caratteristiche. Se non viene selezionato nessuno, tutte le bande (ad eccezione della banda di addestramento) verranno utilizzate come bande di caratteristiche. C'è un'opzione per quantizzare i valori di classe se i valori della fascia di addestramento scelta non sono già discreti. Se la banda di addestramento è costituita da dati che sono etichetti discreti come le classi di landcover, non è necessario quantizzarli. Tuttavia, se i dati della banda di addestramento sono continui come la biomassa, allora ci saranno tante classi quanti sono i valori della biomassa nel set di addestramento. Si consiglia di quantizzare i valori in tali casi. Ad esempio, se l'intervallo di valori nella banda di addestramento è [0.0, 1.0], l'utente può impostare il valore della classe minima su 0.0, la dimensione del passo del valore della classe su 0.1 e i livelli della classe su 10 per quantizzare i valori su 10 livelli: 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9.

### Addestramento sui vettori

L'utente può scegliere un numero di vettori di addestramento (dal primo prodotto elencato nel ProductSet-Reader) come classi. Ad esempio, i vettori di addestramento potrebbero essere regioni (poligoni) che rappresentano ciascuna una classe separata come l'acqua, la città o la foresta. Un vettore di addestramento chiamato "acqua" diventerà un'etichetta di classe chiamata "acqua". Le aree possono essere create utilizzando lo strumento "Nuovo contenitore dati vettoriali" e altri strumenti di disegno come "Strumento di disegno rettangolo". Un pixel all'interno di una regione vettoriale di addestramento avrà il nome della regione come classe anziché il suo valore di dati. Le bande caratteristiche vengono scelte allo stesso modo sui dati Raster. L'operatore cercherà di estrarre lo

stesso numero di campioni per ogni classe durante la costruzione del set di campioni di addestramento o test.

## Caricamento e utilizzo di un classificatore salvato

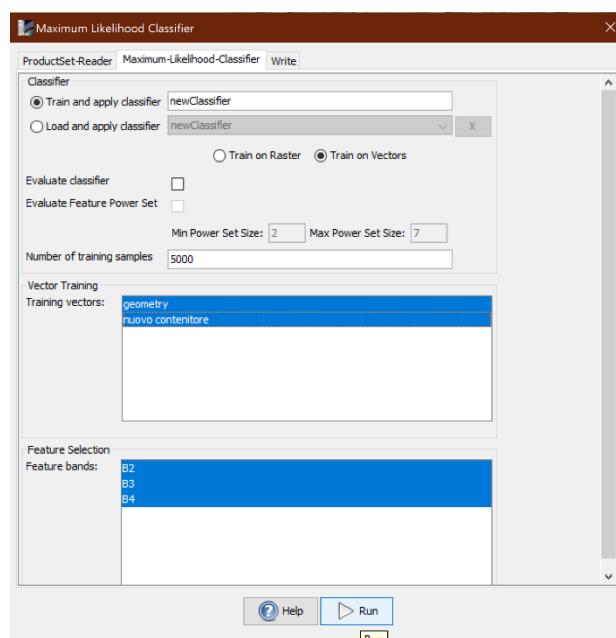
Le informazioni minime che l'utente deve conoscere per utilizzare un classificatore salvato è l'elenco delle funzionalità contenute nel file XML tra le altre informazioni utili. L'utente può specificare più di un prodotto con funzionalità di origine. Per ogni nome in "featureNames" nel file XML, l'operatore cercherà una banda nei prodotti di funzionalità il cui nome lo contiene. Passa in rassegna i prodotti nell'ordine in cui sono elencati in "ProductSet-Reader" e utilizza la prima banda che trova che contiene il nome della funzione. Ad esempio, se il nome è "g0" e ci sono due prodotti di funzionalità ed entrambi contengono una banda denominata "g0", verrà utilizzata la banda del primo prodotto di funzionalità.

## Metodo della massima verosimiglianza con SNAP

Il classificatore di massima verosimiglianza è uno dei metodi di classificazione più popolari, in cui un pixel con la massima verosimiglianza viene classificato nella classe corrispondente. La probabilità è definita come la probabilità a posteriori di un pixel appartenente a una determinata classe.

In SNAP si trova nella voce Raster, Classification, Supervised Classification,

Una volta selezionato il metodo di classificazione va trascinato nella prima casella "ProductSet-Reader" il prodotto che vogliamo classificare. Si passa poi alla casella "Maximum-Likelihood Classifier" e si andranno ad inserire le nostre preferenze: se vogliamo caricare un nuovo classificatore o se vogliamo crearne uno nuovo, in questo caso andrà dato un nome a quest'ultimo, se vogliamo lavorare con i Raster o con i vettori, in questo caso sono stati selezionati i vettori e poi evidenziate le bande su cui agire e i vettori di addestramento creati in precedenza. Una volta selezionate le proprie impostazioni si va su Run e si creerà un nuovo progetto con la sola mappa di classificazione ottenuta.

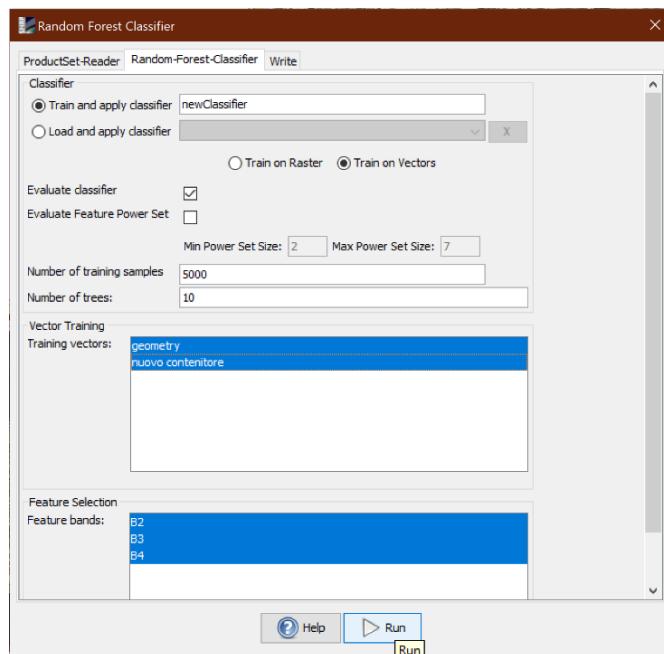


## Random Forest con SNAP

La Random Forest (RF) è una tecnica di classificazione, che è basata sulla costruzione di alberi di regressione, inventata da Breiman. Una RF campiona casualmente e iterativamente i dati e le variabili per generare un grande gruppo, o foresta, di alberi di classificazione e regressione. L'output di classificazione da RF rappresenta la modalità composta da molti alberi decisionali che raggiungono un modello più robusto di un singolo albero di classificazione prodotto da un singolo modello. L'output di regressione da RF rappresenta la media di tutti gli alberi di regressione coltivati in parallelo senza potatura. Tre proprietà utili della RF sono le stime di errore interne, la capacità di stimare l'importanza delle variabili e la capacità di gestire variabili esplicative deboli. La natura iterativa della RF offre un netto vantaggio rispetto agli altri metodi in quanto questo avvia in modo efficace (alimentando sottoinsiemi casuali di dati di allenamento) i dati per previsioni più solide. Questo aiuta a ridurre la correlazione tra alberi. Sottoinsiemi casuali di variabili predittive consentono derivazioni di misure di importanza variabile e prevengono problemi associati a variabili correlate e eccesso di adattamento.

In SNAP si trova nella voce Raster, Classification, Supervised Classification.

Una volta selezionato il metodo di classificazione va trascinato nella prima casella “ProductSet-Reader” il prodotto che vogliamo classificare. Si passa poi alla casella “Random-Forest-Classifier” e si andranno ad inserire le nostre preferenze: se vogliamo caricare un nuovo classificatore o se vogliamo crearne uno nuovo, in questo caso andrà dato un nome a quest’ultimo, se vogliamo lavorare con i Raster o con i vettori, in questo caso sono stati selezionati i vettori e poi evidenziate le bande su cui agire e i vettori di addestramento creati in precedenza. Una volta selezionate le proprie impostazioni si va su Run e si creerà un nuovo progetto con la sola mappa di classificazione ottenuta.



## Accuratezza della classificazione: Confusion Matrix

Una matrice di confusione è una tabella che viene spesso utilizzata per misurare l'accuratezza di un modello di classificazione (o "classificatore") su una serie di dati di test per i quali sono note le classi reali (tabella di contingenza). Viene applicata principalmente ai classificatori basati sull'apprendimento supervisionato. Ogni riga della matrice rappresenta il numero di occorrenze di elementi assegnati ad una certa classe (classe prevista), mentre ogni colonna rappresenta il numero di occorrenze di elementi appartenenti ad una classe effettiva (o viceversa). Sulla diagonale vi sono i numeri degli elementi correttamente classificati, ovvero il numero di elementi assegnati ad un certa classe e che effettivamente appartengono a quella classe. Il nome deriva dal fatto che rende facile vedere se il sistema confonde gli elementi delle diverse classi, vale a dire se etichetta erroneamente gli elementi di una classe usando l'etichetta di un'altra classe.

Di seguito viene presentata la matrice di confusione nel caso di problemi a due classi.

Una matrice di confusione è una tabella con due righe e due colonne. Gli elementi della matrice vengono indicati come numero di falsi positivi, di falsi negativi, di veri positivi e di veri negativi. La matrice riepiloga quindi i risultati della previsione su un problema di classificazione. Il numero di previsioni corrette e non corrette viene riepilogato con i valori di conteggio e suddiviso per ciascuna classe. Questa mostra i modi in cui il modello di classificazione è confuso quando fa previsioni e dà un'idea non solo degli errori commessi da un classificatore, ma soprattutto dei tipi di errori che vengono commessi.

Esempio:

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

- Classe 1: positiva
- Classe 2: negativa

Definizione dei termini:

- Positivo (P): l'osservazione è positiva (ad esempio: è una mela).
- Negativo (N): l'osservazione non è positiva (ad esempio: non è una mela).
- True Positive (TP): l'osservazione è positiva e si prevede che sia positiva.
- Falso negativo (FN): l'osservazione è positiva, ma è prevista negativa.
- Vero negativo (TN): l'osservazione è negativa e si prevede che sia negativa.
- Falso positivo (FP): l'osservazione è negativa, ma è prevista positiva.

L'accuratezza o precisione della classificazione (OA, Overall Accuracy) viene data dalla relazione:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

L'accuratezza OA non è una metrica esaustiva per misurare le prestazioni reali di un classificatore, perché produrrà risultati fuorvianti se il set di dati è sbilanciato, cioè quando il numero di osservazioni nelle diverse classi varia notevolmente.

### **Producer's Accuracy o Recall:**

La Producer's Accuracy (PA, accuratezza del produttore) o Recall (richiamo) è definito come il rapporto tra il numero totale di esempi positivi correttamente classificati diviso per il numero totale di esempi positivi. Una PA alta indica che la classe è stata riconosciuta correttamente (piccolo numero di FN). E' data dalla relazione:

$$\text{Recall} = \frac{TP}{TP + FN}$$

### **User's Accuracy o Precision:**

La User's Accuracy (UA, accuratezza dell'utente) o Precision (precisione) si ottiene dividendo il numero totale di esempi positivi correttamente classificati per il numero totale di esempi positivi previsti. L'alta UA indica che un esempio etichettato come positivo è effettivamente positivo (piccolo numero di FP). La precisione è data dalla relazione:

$$\text{Precision} = \frac{TP}{TP + FP}$$

**Elevata PA, bassa UA:** ciò significa che la maggior parte degli esempi positivi sono riconosciuti correttamente (basso FN) ma ci sono molti falsi positivi.

**Bassa PA, alta UA:** questo dimostra che ci mancano molti esempi positivi (alto FN) ma quelli che ritieniamo positivi sono effettivamente positivi (basso FP).

### **F-Measure:**

Dal momento che abbiamo due misure (Richiamo (PA) e Precisione (UA)), avere una misura che rappresenta entrambe è di aiuto. Per questo, la F-Measure utilizza la media armonica al posto della media aritmetica in quanto punisce maggiormente i valori estremi. L'F-Measure sarà sempre più vicino al valore più piccolo di Precisione o Richiamo.

$$F\text{-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

## Metriche di accuratezza

Per riassumere, la matrice di confusione consente di calcolare le seguenti metriche di accuratezza:

- Producer's Accuracy (PA)
- User's Accuracy (UA)
- Overall Accuracy (OA)

Nel caso a due classi, la matrice è una tabella a due righe e due colonne, come mostrato nella sottosezione precedente. Nel caso ad N classi, la matrice è una tabella a N righe e N colonne. Il calcolo della PA, UA viene quindi effettuato classe per classe. L'accuratezza complessiva è data dalla OA.

### Producer's Accuracy

L'accuratezza del produttore (PA) è l'accuratezza della mappa dal punto di vista del creatore della mappa (il produttore). Questa rappresenta quanto spesso le caratteristiche reali sul terreno sono mostrate correttamente sulla mappa classificata o la probabilità che una certa copertura del suolo di un'area sul terreno sia classificata come tale. È anche il numero di elementi di riferimento (pixel) classificati accuratamente divisi per il numero totale di elementi di riferimento per quella classe (colonna).

### User's accuracy

La precisione dell'utente (UA) è l'accuratezza dal punto di vista di un utente della mappa, non del creatore della mappa. Indica essenzialmente quanto spesso la classe sulla mappa sarà effettivamente presente sul terreno e viene chiamata affidabilità di un classificatore. La precisione dell'utente è calcolata prendendo il numero totale di classificazioni corrette per una determinata classe e dividendola per il totale degli elementi che sono stati previsti per quella classe (riga).

### Overall Accuracy

L'accuratezza complessiva (OA) ci dice essenzialmente quanti elementi di riferimento (pixel) sono stati mappati correttamente. OA è generalmente espressa in percentuale, ad esempio una OA del 100% rappresenta una classificazione perfetta in cui tutti gli elementi di riferimento (pixel) sono stati classificati correttamente. Gli elementi sulla diagonale della matrice di confusione rappresentano le aree che sono state correttamente classificate. Per calcolare OA si fa la sommatoria del numero di elementi correttamente classificati (diagonale) e si divide per la sommatoria del numero totale degli elementi di riferimento.

## Classificazione della Sentinel-2 del 13 Aprile 2018

Per la classificazione dell’immagine Sentinel-2 del 13 Aprile 2018, questa è stata inizialmente pre-processata, come già spiegato nel capitolo 5, effettuando le seguenti operazioni:

- Resampling per rendere le bande delle due mappe tutte della stessa dimensione di pixel (a 20m);
- Masking tramite la Geometry fornita dalla banda “quality\_scene\_classification”, nominata “scl\_water” per eliminare le zone identificate come acquose (mare e laghi) da entrambe le mappe.

Successivamente è stato effettuato il Mosaicing per unire le due mappe di S2 (cioè l’immagine T33TWG che copre Foggia nord e quella T33TWF relativa a Foggia Sud) e di ritaglio dell’immagine, per averne un’unica immagine centrata sul Tavoliere Pugliese su cui sarà effettuata la classificazione.

Ed infine, come ultima operazione di pre-processing del dato, il Reprojection ha associato alla nuova mappa le coordinate geografiche reali.

Riassumendo, applicando le operazioni di Resampling, Masking, Mosaicing, Ritaglio e Reprojection, le immagini ottenute hanno un pixel di 20m sul sistema cartografico con stessa dimensione e sulla posizione geografica effettiva.

Dopo aver effettuato queste operazioni, vengono creati i Data Vector Container contenenti i poligoni riguardanti ognuno una tipologia di coltura diversa su cui verrà effettuata la classificazione come descritto in precedenza.

Nella tabella che segue sono riportate le classi e l’indice di classe, per un totale di 19 classi.

### **Classi ed indice di classe relativi ad Aprile 2018**

<b>Classe</b>	<b>Indice di classe</b>
Pomodoro	1
Broccoli/Cavolfiore	2
Favetta	3
Suolo nudo	4
Asparago	5
Pesca	6
Orzo	7
Cece	8
Foresta sempreverde	9
Pannelli fotovoltaici	10
Foresta decidua	11
Vite	12
Spinacio	13
Carciofo	14
Prateria	15
Grano duro	16
Arbusti	17
Olivo	18
Urbano	19

Nella prossima tabella possiamo osservare i colori associati alle classi e il valore ad esse attribuito da SNAP. Per far partire la numerazione da 1 invece che da 0, è stata effettuata l'operazione di Band Maths in cui veniva sommato 1 ad ogni valore della tabella. Di seguito ritroviamo entrambe le tabelle e le relative numerazioni.

Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Tomato	Red	0	1.367%	
Broccoli/Cau...	Green	1	0.769%	
Horse_beans	Light Green	2	5.377%	
Bare_soil	Brown	3	7.587%	
Asparagus	Dark Green	4	2.763%	
Peach	Light Red	5	3.170%	
Barley	Yellow	6	10.078%	
Chickpea	Light Green	7	2.345%	
Evergreen_f...	Dark Teal	8	1.811%	
Photovoltaic...	Grey	9	13.531%	
Deciduous_f...	Dark Green	10	2.132%	
Vine	Light Red	11	4.074%	
Spinach	Light Green	12	2.308%	
Artichoke	Magenta	13	0.904%	
Grassland	Brown	14	0.807%	
Durum_Wheat	Yellow	15	14.510%	
Shrublands	Yellow	16	6.960%	
Olive	Dark Green	17	12.210%	
Urban		18	7.300%	

Tabella con numerazione attribuita da Snap a partire da 0

Colour	Value
	0
Red	1
Green	2
Light Green	3
Brown	4
Dark Green	5
Light Red	6
Yellow	7
Light Green	8
Dark Teal	9
Grey	10
Dark Green	11
Light Red	12
Light Green	13
Magenta	14
Brown	15
Yellow	16
Yellow	17
Dark Green	18
	19

Tabella dopo l'operazione di Band Maths

I metodi di classificazione sono stati applicati su delle bande selezionate di Sentinel-2. Sulla base della letteratura le bande selezionate sono le 10 bande spettrali da VNIR a SWIR originariamente a 10m e 20m, escludendo quelle originariamente a 60m (ovvero le bande 1, 9, 10) che sono utilizzate principalmente per lo screening delle nuvole e la correzione atmosferica.

Dopo aver definito i Data Vector Container e le Bande, sono stati applicati i due metodi di classificazione, il metodo della Massima Verosimiglianza e il Random Forest. Per entrambi i classificatori sono stati introdotti 50000 esempi di Training.

Successivamente, per la classificazione dell'immagine di Aprile 2018, vengono riportati i valori di PA, UA e OA ottenute dalle corrispondenti Matrici di Confusione. Per il metodo della Massima Verosimiglianza vengono riportati i valori ottenuti in training ed in testing. Per il Random Forest solo i risultati ottenuti con i dati di testing.

## **Accuratezza della classificazione della Sentinel-2 del 13 Aprile 2018**

Le accuratezze ottenute tramite la Matrice di Confusione vengono di seguito riportate:

### **Massima Verosimiglianza**

Di seguito vengono riportati i valori di PA, UA e OA ottenuti in training ed in testing, oltre che il numero dei punti per classe. Il data set non è bilanciato in termini di numeri di punti, per alcune classi il numero è dell'ordine delle centinaia e per altre è dell'ordine di qualche migliaio.

Dai dati di training si può notare che i valori di accuratezze più alte (PA maggiori del 90% ed anche per UA) si ottengono per alcune classi tipo Foresta, Urbano, Prateria.

Altre classi hanno invece valori bassi (PA minori del 40%), come ad esempio Pomodoro, Suolo nudo, Vite ed Olivo. Questo perché ad esempio il Pomodoro in Aprile è stato appena piantato e si confonde con suolo nudo e viceversa. Anche la Vite ad Aprile è poco vegetato e viene riconosciuto parzialmente, mentre l'Olivo, che è sempreverde, in realtà è una classe mista perché in mezzo agli alberi c'è suolo nudo o anche vegetato), per cui presenta confusione.

Per le altre classi, i valori di accuratezza PA o anche UA sono comunque buoni (spesso superiori al 75%) come ad esempio per il grano o l'orzo, che sono due colture abbastanza simili sia per tipologia che stadio di crescita. Ciò dimostra la ricchezza di informazioni di Sentinel-2 che permettono di ottenere questi risultati usando una sola data.

In totale la OA raggiunta in training è del 63%.

Dai dati di testing si può notare come l'OA, che è pari al 50%, è più basso rispetto a quelli di training. Questo è un risultato prevedibile dato che è stato usato un data set indipendente differente rispetto a quello usato per il training. Inoltre il numero di punti per ogni classe è anche più alto di quelli usati in training, ed essendoci una naturale variabilità al suolo anche tra gli elementi nell'ambito di una stessa classe, è prevedibile quindi trovare una accuratezza inferiore. Infatti, la diminuzione della PA e UA si verifica principalmente per le colture, piuttosto che per le classi Foresta, Urbano o carciofo, che sono le classi più stabili.

**Tabella delle accuratezze in fase di Training**

Classe	Indice	Numero dei punti	Producer's Accuracy	User's Accuracy
Pomodoro	1	2632	32%	75%
Broccoli/Cavolfiore	2	555	56%	70%
Favetta	3	2632	55%	71%
Suolo nudo	4	1205	32%	32%
Asparago	5	1649	82%	28%
Pesca	6	103	78%	17%
Orzo	7	2632	79%	80%
Cece	8	2632	80%	85%
Foresta sempreverde	9	129	97%	100%
Pannelli fotovoltaici	10	2632	86%	91%
Foresta decidua	11	234	94%	97%
Vite	12	2632	39%	51%
Spinacio	13	346	79%	47%
Carciofo	14	1848	73%	67%
Prateria	15	166	98%	93%
Grano duro	16	2632	82%	72&
Arbusti	17	156	87%	40%
Olivo	18	2632	39%	69%
Urbano	19	205	94%	47%

**Overall Accuracy in Training: 17456/27652 = 63%**

**Tabella delle accuratezze in fase di Testing**

Classe	Indice	Numero dei punti	Producer's Accuracy	User's Accuracy
Pomodoro	1	4543	29%	78%
Broccoli/Cavolfiore	2	696	48%	40%
Favetta	3	3805	33%	25%
Suolo nudo	4	3989	8%	10%
Asparago	5	4217	53%	22%
Pesca	6	91	48%	2%
Orzo	7	2919	43%	11%
Cece	8	9965	55%	41%
Foresta sempreverde	9	203	83%	97%
Pannelli fotovoltaici	10	5067	85%	82%
Foresta decidua	11	194	68%	100%
Vite	12	5439	24%	34%
Spinacio	13	488	43%	14%
Carciofo	14	2911	78%	57%
Prateria	15	342	0.5%	14%
Grano duro	16	44925	54%	93%
Arbusti	17	342	19%	6%
Olivo	18	2264	32%	24%
Urbano	19	274	95%	23%

**Overall Accuracy in Testing: 46321/91817 = 50%**

## Random Forest

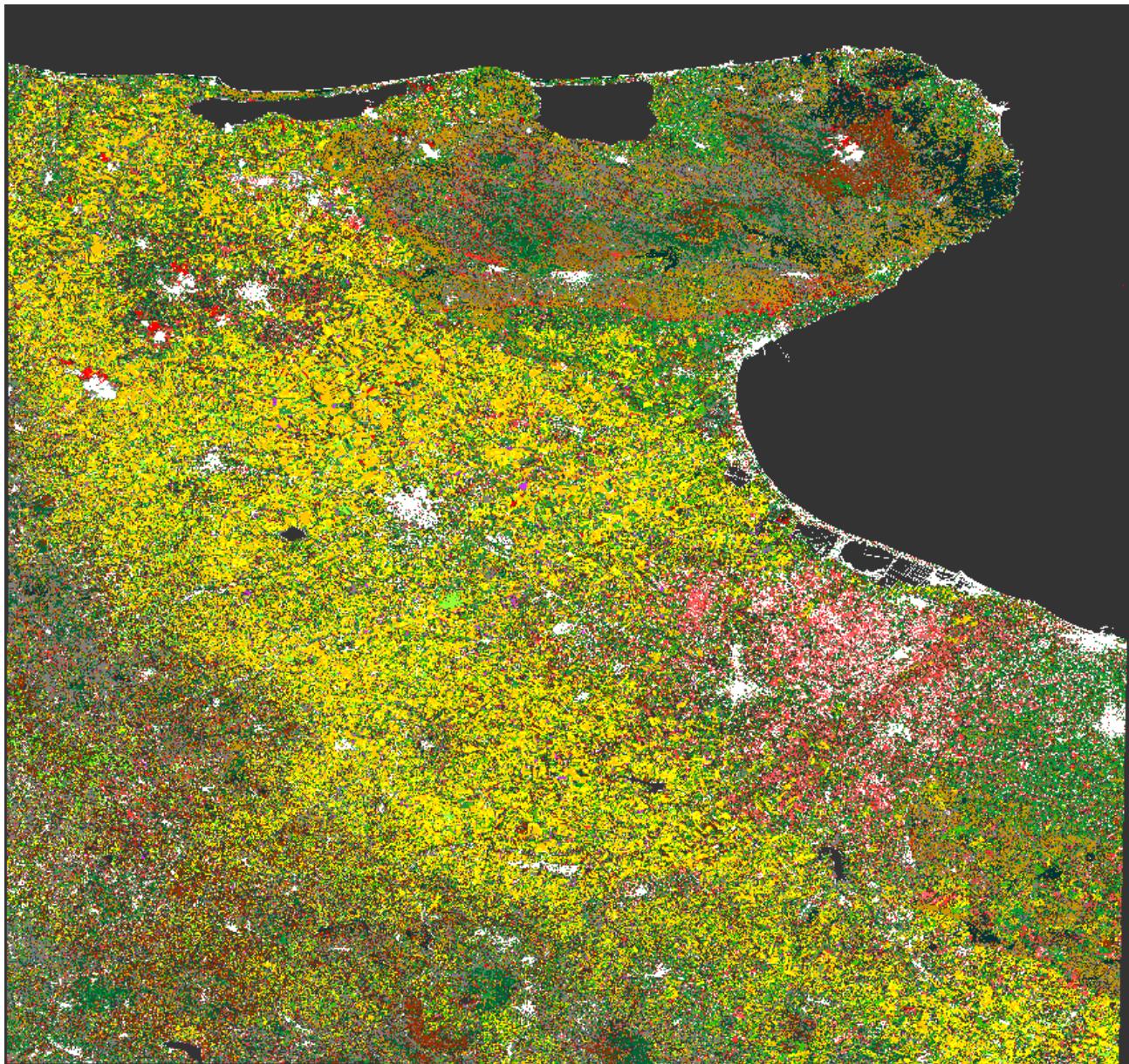
Nella classificata con il Random Forest, l'OA in testing è pari al 42%, un valore più basso rispetto a quello ottenuto con il metodo della Massima Verosimiglianza. Questo accade anche per la maggior parte delle PA e UA delle varie classi, tranne che per Pomodoro, Vite ed Olivo dove si osservano dei valori più alti rispetto all'altro metodo. Almeno in questo esperimento della classificazione della mappa di Aprile 2018, si evince che il Random Forest è meno accurato rispetto al metodo della Massima Verosimiglianza.

**Tabella delle accuratezze in fase di Testing**

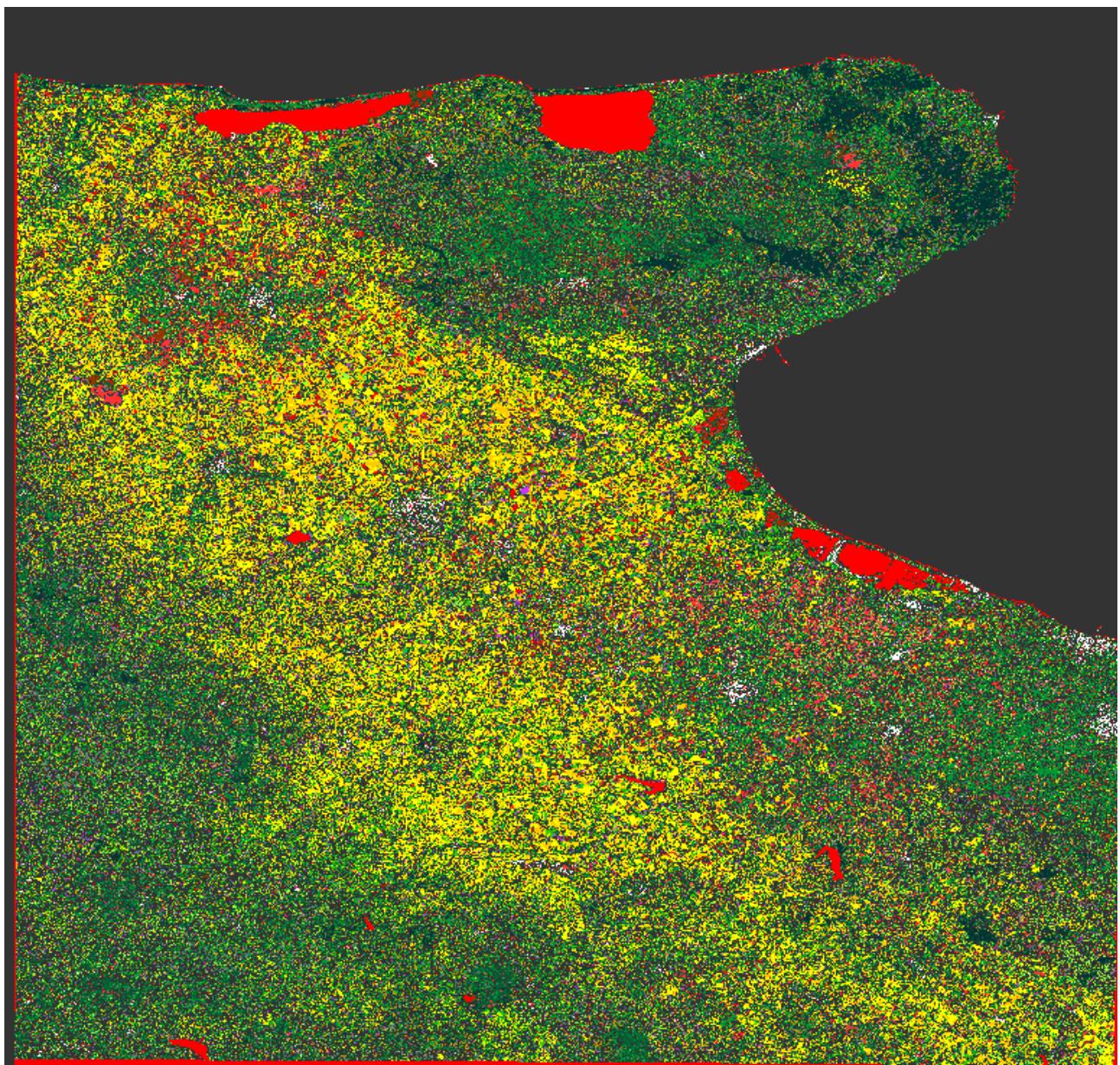
Classe	Indice	Numero dei punti	Producer's Accuracy	User's Accuracy
Pomodoro	1	4543	52%	37%
Broccoli/Cavolfiore	2	696	16%	8%
Favetta	3	3805	43%	18%
Suolo nudo	4	3989	5%	15%
Asparago	5	4217	24%	29%
Pesca	6	91	27%	5%
Orzo	7	2919	32%	6%
Cece	8	9965	56%	49%
Foresta sempreverde	9	203	79%	35%
Pannelli fotovoltaici	10	5067	68%	58%
Foresta decidua	11	194	60%	65%
Vite	12	5439	28%	33%
Spinacio	13	488	24%	19%
Carciofo	14	2911	69%	51%
Prateria	15	342	0.5%	5%
Grano duro	16	44925	43%	91%
Arbusti	17	342	0.8%	0.7%
Olivo	18	2264	41%	14%
Urbano	19	274	49%	51%

**Overall Accuracy in Testing: 39229/91817 = 42%**

Successivamente vengono mostrate dapprima la mappa di classificazione ottenuta con il metodo della Massima Verosimiglianza e poi quella ottenuta con il metodo del Random Forest.



**Mappa delle colture di Aprile 2018 classificata con il metodo della Massima Verosimiglianza**



**Mappa delle colture di Aprile 2018 classificata con il metodo del Random Forest**

Come si può notare, le mappe di classificazione con il metodo della Massima Verosimiglianza e Random Forest presentano diverse gradazioni di colore. Per esempio, il bianco della classe Urbano ed il giallo rappresentante le colture grano ed orzo, spiccano maggiormente nella mappa ottenuta dalla Massima Verosimiglianza. Invece le gradazioni di verde, che rappresentano le varie tipologie di boscaglia, Olivo e colture ortive (Broccoli, Favetta, Asparago), sono più evidenti nella mappa ottenuta dal Random Forest.

NOTA: I contorni in grigio rappresentano il background che non viene classificato durante il processamento. Tuttavia, nella mappa ottenuta da Random Forest, i laghi, che erano stati preventivamente mascherati, sono comunque erroneamente colorati in rosso, a differenza dell'altra mappa, dove sono rimasti colorati in grigio.

## Classificazione della Sentinel-2 del 02 Luglio 2018

Per la classificazione dell'immagine Sentinel-2 del 2 Luglio 2018, sono state effettuate le medesime operazioni come per quella di Aprile. Riassumendo, applicando le operazioni di Resampling, Masking, Mosaicing, Ritaglio e Reprojection, le immagini ottenute hanno un pixel di 20m sul sistema cartografico con stessa dimensione e sulla posizione geografica effettiva.

Dopo aver effettuato queste operazioni, vengono creati i Data Vector Container contenenti i poligoni riguardanti ognuno una classe o tipologia di coltura diversa su cui verrà effettuata la classificazione come descritto in precedenza. Nella tabella che segue sono mostrate le classi e gli indici corrispondenti, per un totale di 15 classi.

**Classi ed indice di classe relativi a Luglio 2018**

Classe	Indice di classe
Pomodoro	1
Prateria	2
Broccoli/Cavolfiore	3
Olivo	4
Asparago	5
Carciofo	6
Pesca	7
Vite	8
Grano duro raccolto	9
Finocchio	10
Foresta sempreverde	11
Pannelli fotovoltaici	12
Foresta decidua	13
Urbano	14
Arbusti	15

Nella prossima tabella possiamo osservare i colori associati alle classi e il valore ad esse attribuito da SNAP. Anche in questo caso, per far partire la numerazione da 1 invece che da 0, è stata effettuata l'operazione di Band Maths in cui veniva sommato 1 ad ogni valore della tabella. Di seguito ritroviamo entrambe le tabelle e le relative numerazioni.

Label	Colour	Value	Frequency	Description
no data		-1	0.000%	no data
Tomato	Red	0	3.369%	
Grassland	Brown	1	1.213%	
Broccoli/Cau...	Green	2	11.867%	
Olive	Dark Green	3	10.945%	
Asparagus	Medium Green	4	3.157%	
Artichoke	Purple	5	2.579%	
Peach	Light Red	6	2.527%	
Vine	Red	7	9.722%	
Durum_Whe...	Yellow	8	9.836%	
Fennel	Light Green	9	13.662%	
Evergreen_f...	Dark Teal	10	4.918%	
Photovoltaic...	Grey	11	9.039%	
Deciduous_f...	Dark Green	12	4.391%	
Urban	White	13	7.951%	
Shrublands	Brown	14	4.824%	

**Tabella con numerazione attribuita da Snap a partire da 0**

Colour	Value
	0
Red	1
Brown	2
Green	3
Dark Green	4
Purple	5
Light Red	6
Red	7
Yellow	8
Light Green	9
Dark Teal	10
Grey	11
Dark Green	12
White	13
Brown	14
Brown	15

**Tabella dopo l'operazione di Band Maths**

Anche in questo caso, per la classificazione sono state selezionate 10 bande spettrali da VNIR a SWIR originariamente a 10m e 20m, escludendo quelle originariamente a 60m (ovvero le bande 1, 9, 10) che sono utilizzate principalmente per lo screening delle nuvole e la correzione atmosferica.

Dopo aver definito i Data Vector Container e le Bande, sono stati applicati il metodo della Massima Verosimiglianza ed il Random Forest. Per entrambi i classificatori sono stati introdotti 50000 esempi di Training.

Successivamente, per la classificazione dell'immagine di Luglio 2018 e per entrambi i metodi, vengono riportati i valori di PA, UA e OA ottenuti in testing dalle corrispondenti Matrici di Confusione.

## Accuratezza della classificazione della Sentinel-2 del 2 Luglio 2018

Le accuratezze ottenute tramite la Matrice di Confusione vengono di seguito riportate:

### Massima Verosimiglianza

Di seguito vengono riportati i valori di PA, UA e OA ottenuti in testing, oltre che il numero dei punti per classe. Anche in questo caso i numeri di punti per classe differiscono da alcune centinaia a qualche migliaio.

In questa classificazione, alcune classi come Pomodoro, Grano (raccolto), Foreste ed urbano, hanno valori di PA alti (PA maggiori dell'80%). Si deve notare che a Luglio il Grano è stato raccolto, quindi quello che è stato classificato in realtà un suolo nudo con presenza di stoppie e steli secchi.

Altre classi hanno invece valori bassi (PA minori del 40%), come ad esempio Prateria, Broccoli, Asparago, Carciofo e Finocchi. Questo perché d'estate, parte di queste colture possono essere state raccolte e possono avere una densità di vegetazione molto variabile. Le praterie possono essere in parte secche.

In totale la OA raggiunta in testing è del 58%. Da considerare che in training, il valore di OA è probabilmente più alto di almeno il 10%. Rispetto ad Aprile, questo valore è più alto anche perché il numero di classi in questo caso è inferiore (15 rispetto a 19).

**Tabella delle accuratezze in fase di Testing**

Classe	Indice	Numero dei punti	Producer's Accuracy	User's Accuracy
Pomodoro	1	7440	79%	84%
Prateria	2	341	10%	26%
Broccoli/Cavolfiore	3	2929	21%	30%
Oliveto	4	2298	58%	43%
Asparago	5	1383	31%	28%
Carciofo	6	532	13%	8%
Pesca	7	97	27%	2%
Vite	8	5429	40%	78%
Grano duro raccolto	9	4168	81%	50%
Finocchio	10	1326	7%	9%
Foresta sempreverde	11	200	88%	97%
Pannelli fotovoltaici	12	5029	74%	94%
Foresta decidua	13	195	100%	100%
Urbano	14	278	96%	37%
Arbusti	15	124	15%	5%

**Overall Accuracy in Testing: 18349/31618 = 58.0%**

## Random Forest

Nella classificata con il Random Forest, l'OA sale leggermente a 58.2%. Tuttavia a livello di singole classi, si può notare che la PA in alcuni casi è più alta di almeno 10% rispetto al metodo della Massima Verosimiglianza (ad esempio per Broccoli, Olivo, Vite), mentre per altri casi è più bassa di almeno il 10% (ad esempio per Carciofo, Pesca, Grano).

Pertanto, in questo esperimento, il Random Forest ha una prestazione complessiva simile al Massima Verosimiglianza.

**Tabella di accuratezza in fase di Testing**

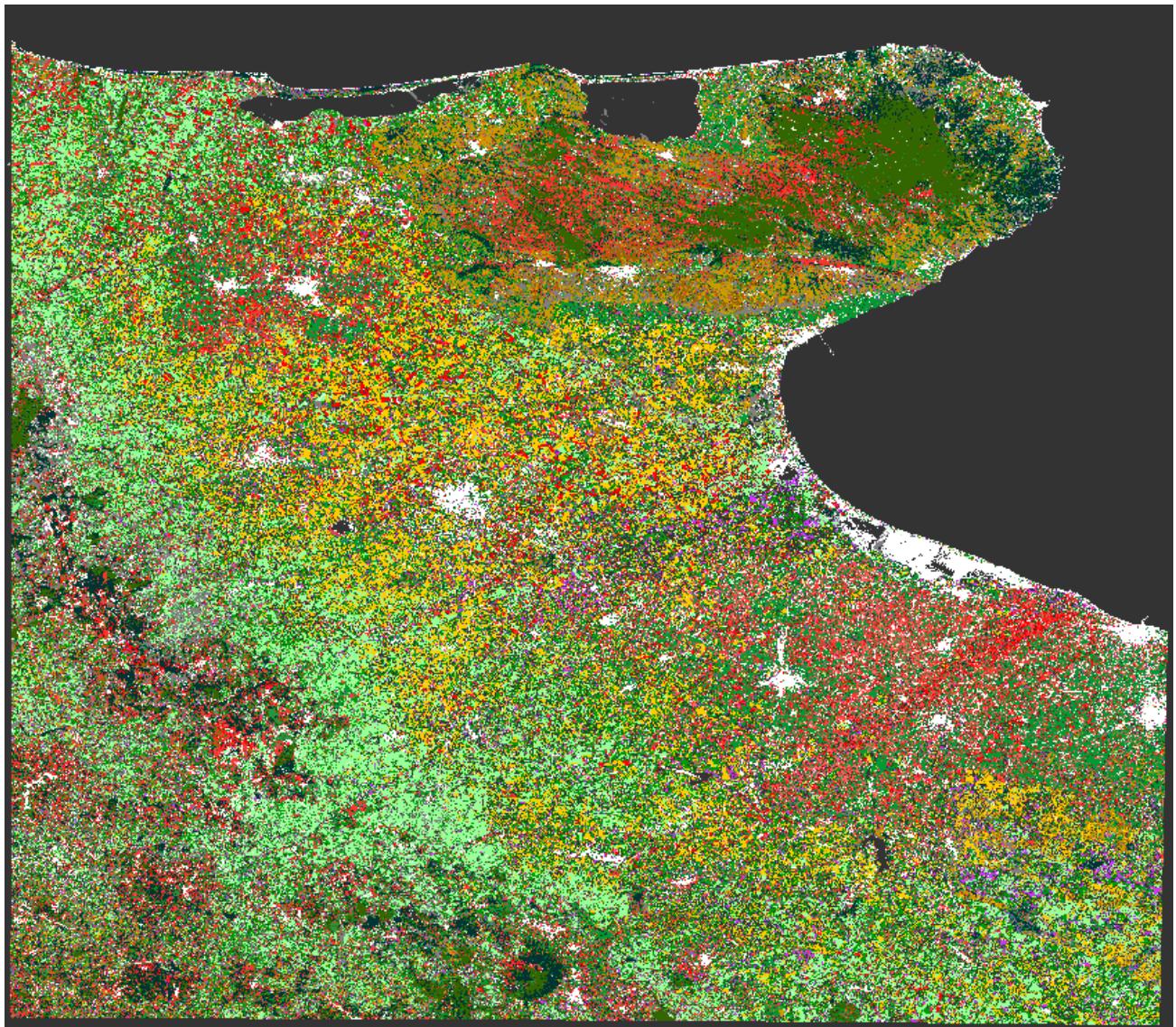
Classe	Indice	Numero dei punti	Producer's Accuracy	User's Accuracy
Pomodoro	1	7440	75%	76%
Prateria	2	341	7%	32%
Broccoli/Cavolfiore	3	2929	46%	30%
Olivo	4	2298	79%	46%
Asparago	5	1383	28%	25%
Carciofo	6	532	3%	15%
Pesca	7	97	1%	0.5%
Vite	8	5429	49%	79%
Grano duro raccolto	9	4168	50%	49%
Finocchio	10	1326	5%	9%
Foresta sempreverde	11	200	96%	43%
Pannelli fotovoltaici	12	5029	78%	90%
Foresta decidua	13	195	100%	94%
Urbano	14	278	45%	72%
Arbusti	15	124	4%	2%

**Overall Accuracy in Testing:  $18415/31618 = 58.2\%$**

Successivamente vengono mostrate dapprima la mappa di classificazione ottenuta con il metodo della Massima Verosimiglianza e poi quella ottenuta con il metodo del Random Forest.

Come si può notare dalle due mappe di classificazione, quella ottenuta dal metodo della Massima Verosimiglianza rispetto a quella ottenuta con il Random Forest, presenta una maggiore presenza di bianco, che rappresenta l'urbano, e di giallo, che rappresenta le aree di grano (già raccolto). In teoria queste aree mostrano anche delle differenze rispetto a quelle ottenute in Aprile.

In entrambe le mappe, le zone in rosso all'interno del Tavoliere Pugliese rappresentano il Pomodoro e questa classe, avendo una alta PA e UA maggiori del 75% dovrebbero essere affidabili, anche più del risultato ottenuto in Aprile, quando il pomodoro era ancora poco sviluppato. Tuttavia nella mappa Random Forest appaiono in rosso anche altre zone (a parte i laghi interni che sono stati comunque preventivamente mascherati)



**Mappa delle colture di Luglio 2018 classificata con il metodo della Massima Verosimiglianza**



**Mappa delle colture di Luglio 2018 classificata con il metodo del Random Forest**

Inoltre, nella prima mappa è presente una gradazione di verde più chiara rispetto a quella ottenuta con Random Forest (ovvero maggiore presenza di classi ortive tipo Finocchio, Broccoli). Tuttavia, per queste classi la PA è alquanto bassa quindi meno affidabile.

NOTA: I contorni in grigio rappresentano il background che non viene classificato durante il processamento. Tuttavia, nella mappa ottenuta da Random Forest, i laghi, che erano stati preventivamente mascherati, sono stati comunque erroneamente colorati in rosso, a differenza dell'altra mappa, dove sono rimasti colorati in grigio.

## 6.3 Post-Processing

### Closing

Il closing ovvero la chiusura è un operatore importante nel campo della morfologia matematica. Come la sua doppia apertura dell'operatore, può essere derivato dalle operazioni fondamentali di erosione e dilatazione. Come questi ultimi operatori, viene normalmente applicata alle immagini binarie.

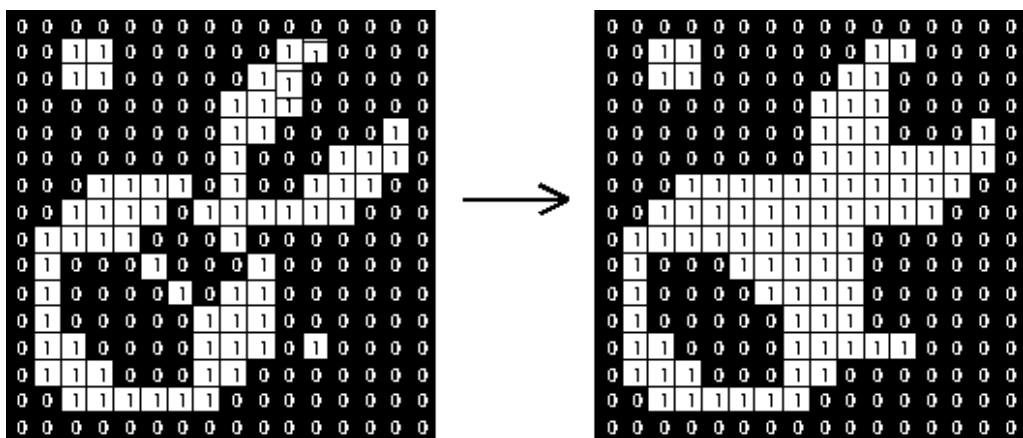
Il closing è simile in qualche modo alla dilatazione in quanto tende ad allargare i confini delle regioni di primo piano (luminose) in un'immagine (e restringe i buchi di colore di sfondo in tali regioni), ma è meno distruttiva della forma del contorno originale.

Come con altri operatori morfologici, l'operazione esatta è determinata da un elemento strutturante. L'effetto dell'operatore è di preservare le aree di sfondo che hanno una forma simile a questo elemento di strutturazione o che possono contenere completamente l'elemento di strutturazione, eliminando tutte le altre regioni di pixel sullo sfondo.

Il closing viene definito semplicemente come una dilatazione seguita da un'erosione usando lo stesso elemento strutturante per entrambe le operazioni. L'operatore di closing richiede quindi due ingressi: un'immagine da chiudere e un elemento strutturante.

Uno degli usi della dilatazione consiste nel riempire piccoli buchi di colore dello sfondo nelle immagini, ad es. "rumore di pepe". Uno dei problemi nel fare ciò, tuttavia, è che la dilatazione distorcerà anche indiscriminatamente tutte le regioni di pixel. Eseguendo un'erosione sull'immagine dopo la dilatazione, ovvero una closing, riduciamo parte di questo effetto. L'effetto della closing può essere facilmente visualizzato.

Dopo che il closing è stata eseguito, la regione di sfondo sarà tale che l'elemento di strutturazione può essere fatto per coprire qualsiasi punto dello sfondo senza che una parte di esso copra anche un punto di primo piano, e quindi ulteriori closing non avranno alcun effetto. Questa proprietà è nota come idempotenza. L'effetto di un closing su un'immagine binaria usando un elemento di struttura quadrata  $3 \times 3$  è illustrato nella Figura seguente:



## Applicazione del Closing

L'operazione di Closing è stata applicata a tutte le classificazioni ottenute, in modo da ottenere mappe con aree più compatte ed omogenee all'interno dei campi agricoli e rendendo le immagini più nitide eliminando punti isolati quanto più possibile.

Questo processo aumenta in parte anche l'accuratezza complessiva. Infatti come esempio, nel caso dell'immagine classificata del 13 Aprile 2018 con il Metodo della Massimo Verosimiglianza, vengono riportati i nuovi valori di accuratezza ottenuti con l'operazione di closing.

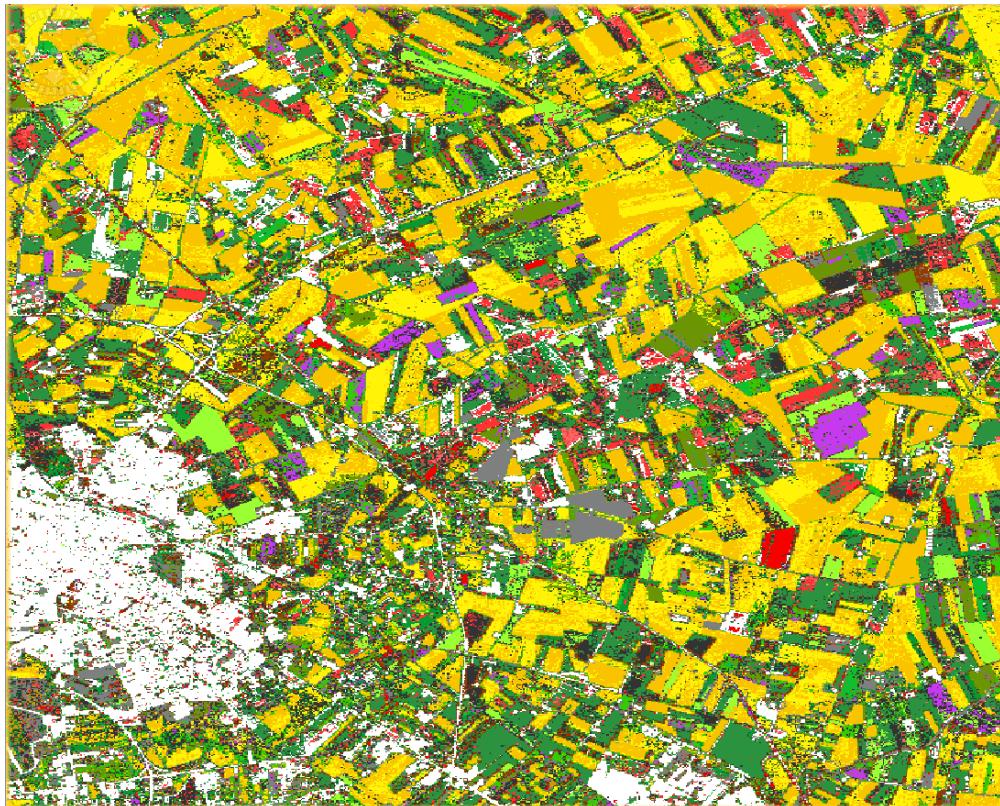
L'OA ottenuta è del 56 % , vi è un miglioramento del 6% rispetto alla classificazione senza il Closing.

Classe	Indice	Numero dei punti	Producer's Accuracy
Pomodoro	1	4543	29.41%
Broccoli/Cavolfiore	2	696	48.56%
Favetta	3	3805	26.32%
Suolo nudo	4	3989	8.28%
Asparago	5	4217	54.52%
Pesca	6	91	24.81%
Orzo	7	2919	33.95%
Cece	8	9965	59.27%
Foresta sempreverde	9	203	84.73%
Pannelli fotovoltaici	10	5067	86.85%
Foresta decidua	11	194	69.59%
Vite	12	5439	26.90%
Spinacio	13	488	51.64%
Carciofo	14	2911	86.22%
Prateria	15	342	0.5%
Grano duro	16	44925	66.51%
Arbusti	17	342	23.33%
Olivo	18	2264	33.13%
Urbano	19	274	100%

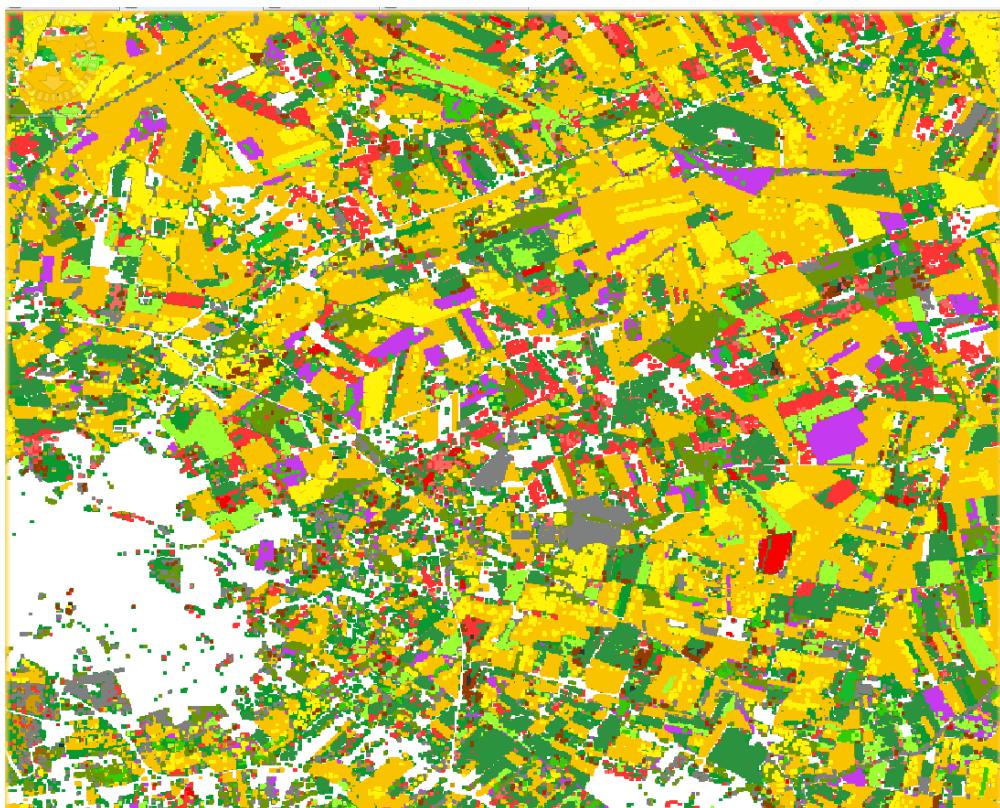
**Overall Accuracy in Testing: 56.43 %**

Di seguito verranno mostrate porzioni di mappe classificate prima e dopo l'applicazione del Closing per entrambi i metodi.

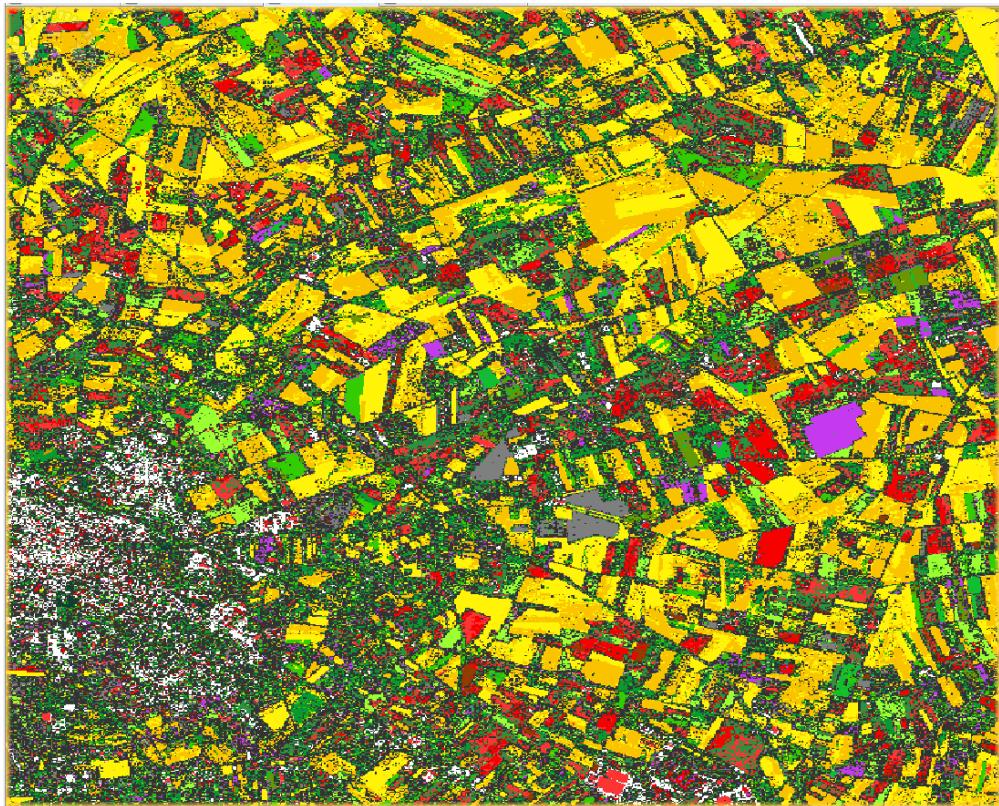
## Mappe del 13 Aprile 2018



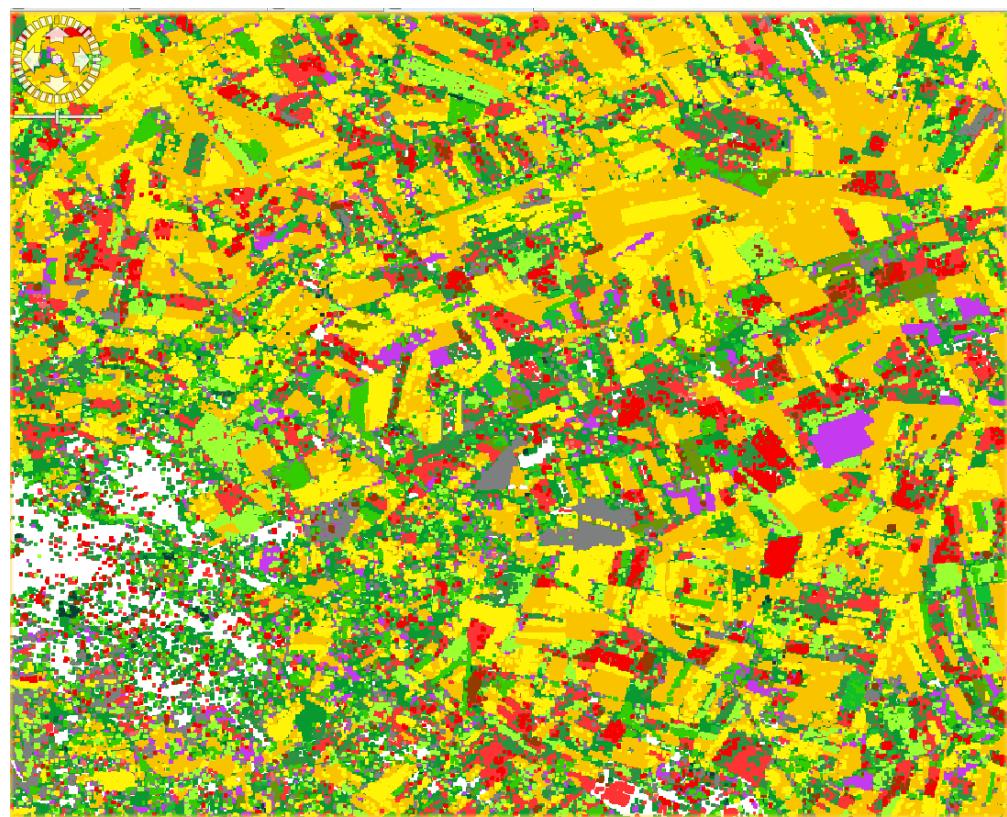
Mappa di Aprile 2018 ottenuta con il metodo della Massima Verosimiglianza senza Closing



Mappa di Aprile 2018 ottenuta con il metodo della Massima Verosimiglianza con Closing



**Mappa di Aprile 2018 ottenuta con il metodo del Random Forest senza Closing**

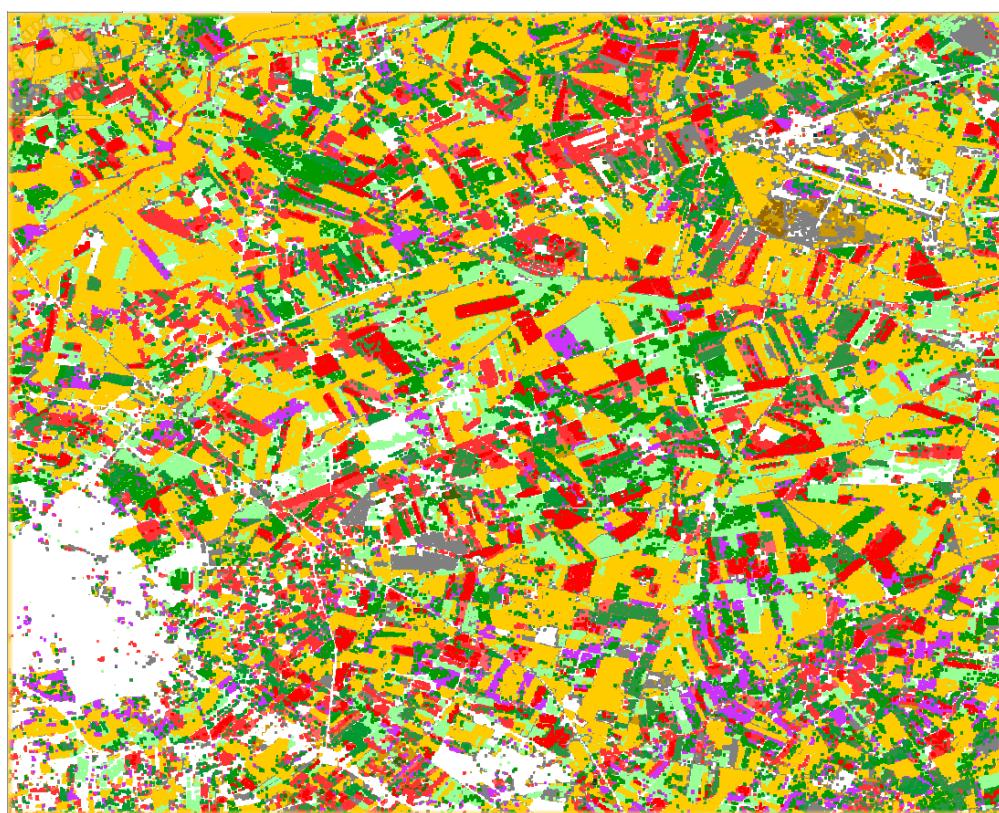


**Mappa di Aprile 2018 ottenuta con il metodo del Random Forest con Closing**

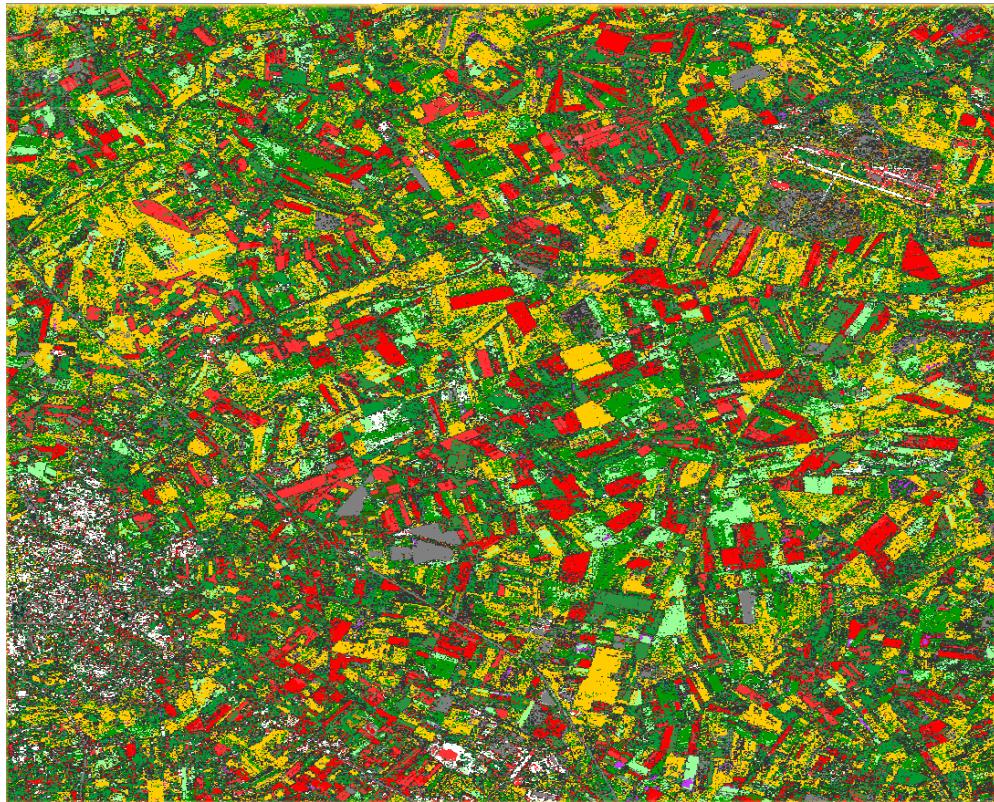
## Mappe del 2 Luglio 2018



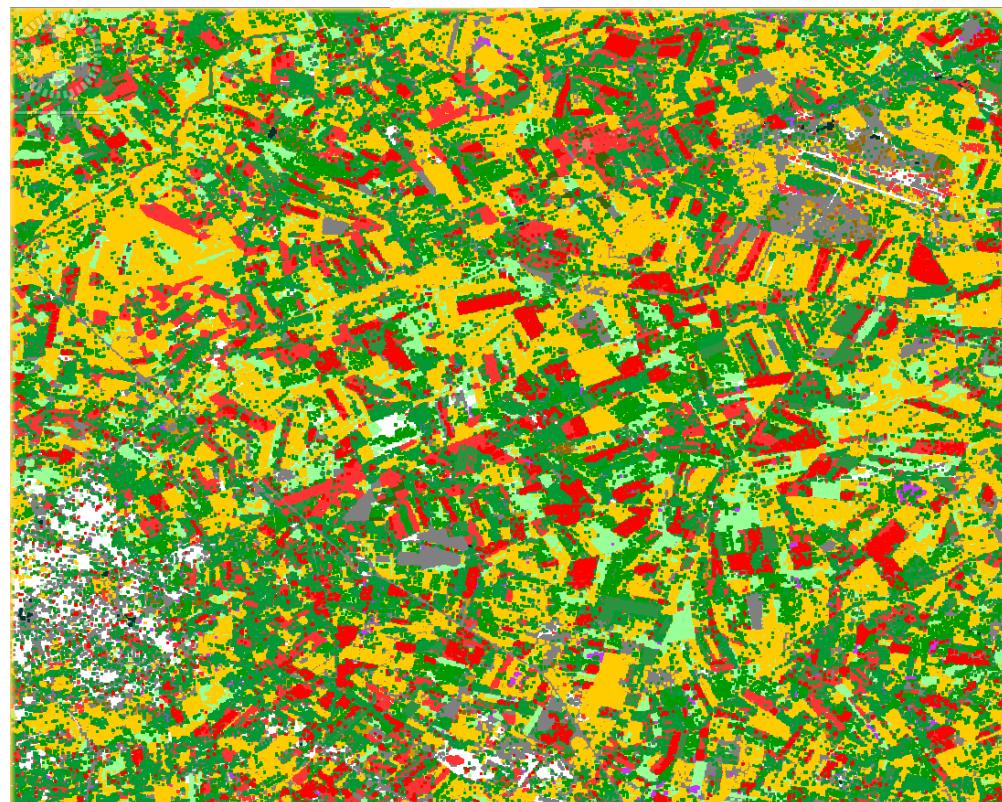
Mappa di Luglio 2018 ottenuta con il metodo della Massima Verosimiglianza senza Closing



Mappa di Luglio 2018 ottenuta con il metodo della Massima Verosimiglianza con Closing



**Mappa di Luglio 2018 ottenuta con il metodo del Random Forest senza Closing**



**Mappa di Luglio 2018 ottenuta con il metodo del Random Forest con Closing**

## 7 Conclusioni

L’obiettivo di questo studio è stato quello di effettuare una classificazione dei dati telerilevati Sentinel-2 per ottenere delle mappe delle colture stagionali presenti nell’area agricola del Tavoliere Pugliese.

I dati telerilevati utilizzati sono stati i dati multi-banda Sentinel-2 acquisiti in due date ovvero il 13 Aprile 2018 e il 2 Luglio 2018. Essi sono stati pre-processati usando come software il Toolbox SNAP.

Per la classificazione sono stati applicati due metodi di classificazione supervisionata, ovvero il metodo della Massima Verosimiglianza e il metodo del Random Forest. Per lo scopo sono stati utilizzati degli esempi di poligoni di verità a terra di colture della zona agricola di studio che fanno parte di un data base raccolto dal Consiglio per la Ricerca in Agricoltura e l’Analisi Economica (CREA) nell’ambito del progetto di ricerca Europeo “SENSAGRI”. Per Aprile 2018 i poligoni rappresentavano 19 classi, mentre per Luglio 2018 15 classi.

La classificazione con i dati di Aprile 2018 effettuata con il metodo della Massima Verosimiglianza, è stata valutata calcolando la Producer’s Accuracy (PA), la User’s Accuracy (UA) e la Overall Accuracy (OA). Per riassumere, l’OA in fase di Training è pari al 63%, mentre in fase di Testing è più bassa, come ci si aspettava dalla letteratura, cioè pari al 50%. Usando invece il metodo Random Forest, la OA in fase di Testing ottenuta è stata più bassa, ovvero il 42%. Quindi per questa sperimentazione, è emerso è il metodo della Massima Verosimiglianza risulta avere una accuratezza maggiore rispetto al metodo Random Forest.

Analogamente, la classificazione con i dati Luglio 2018 è stata effettuata con il metodo della Massima Verosimiglianza e con il metodo del Random Forest. La OA ottenuta in fase di Testing con il metodo della Massima Verosimiglianza è stata pari al 58%, un valore molto simile a quello ottenuto con il Random Forest, che è stato pari al 58.2% . Tuttavia tra i due metodi sono presenti alcune differenze di accuratezze a livello di singole classi, ovvero per alcune classi le accuratezze sono più alte per il primo metodo e più bassi nel secondo e viceversa.

Comparando i risultati di Aprile e Luglio, si è ottenuto che a Luglio i valori di accuratezza in Testing sono più alti e questo è principalmente dovuto al più basso numero di classi presenti a Luglio e quindi alla minore confusione tra classi.

Infine, sono state prodotte delle mappe, che sono state anche rifinite con l’operatore di Closing, applicato come operazione di post-processing. Le mappe ottenute in questo modo mostrano risultati migliori (circa il 6% in più) poiché il Closing elimina punti isolati di rumore dalle immagini rendendole le aree più compatte ed omogenee.

## 8 Bibliografia

- [1] «Wikipedia,» 8 Novembre 2019. [Online]. "Random Forest". Available: [https://en.wikipedia.org/wiki/Random\\_forest](https://en.wikipedia.org/wiki/Random_forest)
- [2] «Wikipedia,» 16 Novembre 2019. [Online]. "Spettro elettromagnetico". Available: [https://it.wikipedia.org/wiki/Spettro\\_elettromagnetico](https://it.wikipedia.org/wiki/Spettro_elettromagnetico)
- [3] «Wikipedia, » 11 February 2020, [Online], "Confusion matrix". Available: [https://en.wikipedia.org/wiki/Confusion\\_matrix](https://en.wikipedia.org/wiki/Confusion_matrix)
- [4] «Wikipedia,» 25 Ottobre 2019. [Online]. "Spettro Visibile". Available: [https://it.wikipedia.org/wiki/Spettro\\_visibile](https://it.wikipedia.org/wiki/Spettro_visibile)
- [5] «Wikipedia,» 4 Giugno 2019. [Online]. "Metodo della massima verosimiglianza". Available: [https://it.wikipedia.org/wiki/Metodo\\_della\\_massima\\_verosimiglianza](https://it.wikipedia.org/wiki/Metodo_della_massima_verosimiglianza)
- [6] Breiman L. "Machine Learning", (2001) 45: 5. [Online] Available: <https://doi.org/10.1023/A:1010933404324>
- [7] Breiman L., Machine Learning, 45, 5–32, 2001,2001 Kluwer Academic Publishers. Manufactured in The Netherlands, "Random Forest", p.28. cap.1-2-3.
- [8] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Sentinel-2". Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>
- [9] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Access to Sentinel data via download ".Available: <https://sentinel.esa.int/web/sentinel/sentinel-data-access>
- [10] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Data Products". Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products>
- [11] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Ground Segment". Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/ground-segment>
- [12] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Mission Objectives". Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/mission-objectives>
- [13] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Overview". Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/overview>
- [14] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Radiometric Resolution". Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric>
- [15] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Resolution". Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions>
- [16] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Satellite Description". Available: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/satellite-description>
- [17] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Sentinel-2 MSI Technical Guide". Available: <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi>
- [18] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "SNAP". Available: <http://step.esa.int/main/toolboxes/snap/>
- [19] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "Spatial Resolution". Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial>
- [20] European Space Agency - ESA, ESA 2000 - 2019, [Online]. "The Sentinel-2 Toolbox ". Available: <https://sentinel.esa.int/web/sentinel/toolboxes/sentinel-2;jsessionid=CB551BD468730CFFF09FE0B785ADDB23.jvm2>
- [21] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Sentinel Toolbox Overview". Available: <nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html>

- [22] European Space Agency - ESA, ESA 2000 - 2019, Snap, from Sentinel Toolbox help. "The BEAM-DIMAP Data Format". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [23] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Application Overview". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [24] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Product Explorer". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [25] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Image View". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [26] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Color Manipulation". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [27] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Import ESRI Shapefile". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [28] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Product Library". Available: nbdocs:/org/esa/snap/productlibrary/docs/general/Overview.html
- [29] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Scatter Plot". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [30] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Geo-Coding Information". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [31] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Histogram". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [32] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Product/Band Information". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [33] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Profile Plot". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [34] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Correlative Plot". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [35] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Statistics Plot". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [36] European Space Agency - ESA, ESA 2000 - 2019, SNAP, from Sentinel Toolbox help. "Metadata Plot". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html

- [37] European Space Agency - ESA, ESA 2000 - 2019, SNAP,from Sentinel Toolbox help. "Vector Data Management". Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [38] European Space Agency - ESA, ESA 2000 - 2019, SNAP,from Sentinel Toolbox help. "Import ESRI Shapefile".Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [39] European Space Agency - ESA, ESA 2000 - 2019, SNAP,from Sentinel Toolbox help. "Band Maths".Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [40] European Space Agency - ESA, ESA 2000 - 2019, SNAP,from Sentinel Toolbox help. "Mosaic".Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [41] European Space Agency - ESA, ESA 2000 - 2019, SNAP,from Sentinel Toolbox help. "Reprojection".Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [42] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Resampling".Available: nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [43] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Create Land Mask".Available : nbdocs:/org/esa/snap/snap/help/docs/general/overview/SnapOverview.html
- [44] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Multi-size Mosaic".Available: nbdocs:/org/esa/s2tbx/dataio/mosaic/ui/docs/general/overview/SnapOverview.html
- [45] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Collocation".Available: nbdocs://org.esa.snap.snap.help/org/esa/snap/snap/help/docs/general/overview/SnapOvervie w.html
- [46] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Decision Tree Classification". Available: nbdocs:/org/esa/s1tbx/fex/docs/general/Overview.html
- [47] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help." Supervised Classification ".Available: nbdocs:/org/esa/snap/classification/docs/general/Overview.html
- [48] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Random Forest Classification ".Available: nbdocs:/org/esa/snap/classification/docs/general/Overview.html
- [49] European Space Agency - ESA, ESA 2000 - 2019, Snap,from Sentinel Toolbox help. "Maximum Likelihood Classification ".Available: nbdocs:/org/esa/snap/classification/docs/general/Overview.html
- [50] Leo Breiman,"SpingerLink",Ottobre 2001, Volume 45, Issue 1, pp 5–32.[Online]. "Random Forest". Available: <https://link.springer.com/article/10.1023/A:1010933404324>
- [51] M. Pal (2005) Random forest classifier for remote sensing classification, International Journal of Remote Sensing, 26:1, 217-222, DOI: 10.1080/01431160412331269698.
- [52] Petroni N. C.,"Lezioni di Calcolo delle Probabilità e Statistica",,(2015/2016),Università degli Studi di Bari Aldo Moro Dipartimento di Matematica,p.269.cap.8.

- [53] R. Fisher, S. Perkins, A. Walker and E. Wolfart.(2003),[Online]. "Closing". Available: <https://homepages.inf.ed.ac.uk/rbf/HIPR2/close.htm>
- [54] Shixin Gu and Qingtong Wu , "Synced", 24 Ottobre 2017, [Online]. "How Random Forest Algorithm Works in Machine Learning". Available: <https://syncedreview.com/2017/10/24/how-random-forest-algorithm-works-in-machine-learning/>
- [55] Will Koehrsen,"Towards Data Science, Sharing concepts, ideas, and codes.",30 Agosto 2018, [Online]. "An Implementation and Explanation of the Random Forest in Python". Available: <https://towardsdatascience.com/an-implementation-and-explanation-of-the-random-forest-in-python-77bf308a9b76>
- [56] Abhishek Sharma. "Confusion Matrix in Machine Learning",2019, [Online]. Available : <https://www.geeksforgeeks.org/confusion-matrix-machine-learning>
- [57] Humbold State University-Geospatial Online-GSP 216-Introducing to remote sensing, "Accuracy Metrics",2014,  
[Online] Available: [http://gis.humboldt.edu/OLM/Courses/GSP\\_216\\_Online/lesson6-2/metrics.html](http://gis.humboldt.edu/OLM/Courses/GSP_216_Online/lesson6-2/metrics.html)

## **9 Ringraziamenti**

*Mi è doveroso dedicare questo spazio del mio elaborato alle persone che hanno contribuito, con il loro supporto, alla realizzazione dello stesso.*

*Vorrei innanzitutto ringraziare il Professor Donato Impedovo per avermi fatto conoscere e approfondire questo argomento ma soprattutto per la sua fiducia e il suo sostegno nell'intraprendere questo percorso.*

*Un ringraziamento speciale va al Dott. Giuseppe Satalino per il suo sincero ed indispensabile aiuto in ogni singolo momento di progettazione e stesura del mio lavoro e per la sua gentilezza e disponibilità, oltre che per i suoi preziosi consigli.*

*Si ringraziano il Dr. Sergio Ruggieri del Consiglio per la Ricerca in Agricoltura e l'Analisi Economica (CREA), Centro di ricerca Agricoltura e Ambiente (AA) di Bari ed il Dr. Michele Rinaldi Centro di ricerca Cerealicoltura e Colture Industriali (CI) di Foggia, per aver messo a disposizione i dati a terra relativi alle colture osservate nell'area del Tavoliere Pugliese e che sono stati raccolti durante il progetto Europeo H2020 SENSAgRI: "SeNtinel Synergy for Agriculture", 2016-2019. <http://www.sensagri.eu/>.*

*Ringrazio tutto lo staff dell'IREA, in cui ho svolto un tirocinio formativo della durata di 3 mesi e complementare alla redazione della tesi, per l'ospitalità e per le skills acquisite sul campo.*

*Ringrazio di cuore i miei genitori e mia sorella. Grazie per avermi sempre sostenuto e per avermi permesso di portare a termine questo capitolo della mia vita sopportandomi e motivandomi ogni giorno. Grazie per essere qui oggi, perché*

*Ringrazio i miei adorati nonni per essere stati come dei secondi genitori per me in questi anni. Grazie per avermi dedicato le più dolci parole nella gioia e nel dolore. Grazie per non aver mai smesso di credere in me. Ma soprattutto grazie a te Nonno, so che saresti fiero di me oggi.*

*Ringrazio il mio fidanzato Francesco per avermi trasmesso la sua immensa forza e il suo coraggio. Grazie per tutto il tempo che mi hai dedicato. Grazie perché ci sei sempre stato.*

*Grazie a tutti i miei colleghi di corso e in particolare a Gianni, Ciccio e Daniele per avermi sempre incoraggiato fin dall'inizio del percorso universitario. Grazie per aver condiviso insieme tutti i bei momenti che racchiudono questi tre anni della mia vita.*

*Grazie ai miei amici di sempre Loris, Alessia, Giovanni, Lilliana, Cosimo, Gianluca, Tiziana, Alessio, Maria Teresa e Anna Paola, per essere stati sempre dalla mia parte pronti a motivarmi e a sostenermi. Grazie per aver ascoltato tutti i miei sfoghi e per tutti i momenti di spensieratezza trascorsi insieme.*

*Infine, dedico questa tesi a me stessa, ai miei sacrifici, alla mia forza di volontà e alla mia tenacia che mi hanno permesso di arrivare fin qui.*

*Grazie infinite a tutti voi.*