



DIPARTIMENTO DI SCIENZE AGRARIE ALIMENTARI E AMBIENTALI

CORSO DI LAUREA IN: SCIENZE FORESTALI, DEI SUOLI E DEL PAESAGGIO

**PREPROCESSAMENTO DELLE SERIE TEMPORALI
TELERILEVATE PER L'EFFICIENTAMENTO DELLA MAPPATURA
DELLA VEGETAZIONE**

**REMOTELY SENSED TIME SERIES PREPROCESSING FOR EFFICIENT
VEGETATION MAPPING**

TIPO TESI: sperimentale

Studente:
LORENZO CESARETTI

Relatore:
PROF. ROBERTO PIERDICCA

Correlatori:
PROF. SIMONE PESARESI
PROF. SIMONA CASAVECCHIA

ANNO ACCADEMICO 2020-2021

SOMMARIO

ELENCO DELLE TABELLE.....	5
ELENCO DELLE FIGURE	9
ACRONIMI E ABBREVIAZIONI	14
CAPITOLO 1 INTRODUZIONE E SCOPO DELLA TESI	15
CAPITOLO 2 MATERIALI E METODI	22
2.1 Inquadramento territoriale.....	22
2.1.1 Gola di Frasassi.....	22
2.1.2 Monte Conero	27
2.2 Strumenti software	32
2.2.1 R.....	32
2.2.2 JupyterLab	32
2.3 Preprocessamento delle immagini	33
2.3.1 Indici di vegetazione	36
2.4 Costruzione delle serie temporali.....	37
2.4.1 Individuazione dei valori anomali	38
2.4.2 Aggregazione e Smoothing.....	40
2.5 Analisi funzionale	44
2.6 Classificazione	47
CAPITOLO 3 RISULTATI.....	53
3.1 Variazioni stagionali descritte dall'analisi funzionale	53
3.1.1 Variazioni stagionali derivate dalle serie temporali RI.....	53
3.1.2 Variazioni stagionali derivate dalle serie temporali MCARI.....	54
3.1.3 Variazioni stagionali derivate dalle serie temporali GNDVI.....	55
3.1.4 Variazioni stagionali derivate dalle serie temporali NDVI.....	56
3.1.5 Variazioni stagionali derivate dalle serie temporali NDRE.....	57
3.1.6 Variazioni stagionali derivate dalle serie temporali MNDWI	58
3.1.7 Variazioni stagionali derivate dalle serie temporali NDMI	59

3.2 Mappatura della vegetazione	60
3.2.1 Mappatura dei predittori derivati dal RI	60
3.2.2 Mappatura dei predittori derivati dal MCARI	61
3.2.3 Mappatura dei predittori derivati del GNDVI.....	62
3.2.4 Mappatura dei predittori derivati dal NDVI	63
3.2.5 Mappatura dei predittori derivati dal NDRE.....	64
3.2.6 Mappatura dei predittori derivati dal MNDWI.....	65
3.2.7 Mappatura dei predittori derivati dal NDMI.....	66
3.2.8 Mappatura dei predittori derivati dal totale degli indici	67
3.3 Confronto tra i metodi.....	69
3.4 Confronto tra i predittori.....	72
CAPITOLO 4 DISCUSSIONI E CONCLUSIONI	74
CAPITOLO 5 APPENDICE	77
BIBLIOGRAFIA	110

ELENCO DELLE TABELLE

Tabella 1-1: Comuni piattaforme di telerilevamento con specifiche generali di sensore .	18
Tabella 2-1: Specifiche del sensore MSI Sentinel-2 per ciascuna banda di acquisizione.	34
Tabella 3-1: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	53
Tabella 3-2: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	54
Tabella 3-3: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	54
Tabella 3-4: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	55
Tabella 3-5: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	55
Tabella 3-6: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	56
Tabella 3-7: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	56
Tabella 3-8: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	57

Tabella 3-9: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	57
Tabella 3-10: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	58
Tabella 3-11: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	58
Tabella 3-12: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	59
Tabella 3-13: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	59
Tabella 3-14: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.	60
Tabella 3-15: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	61
Tabella 3-16: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	61
Tabella 3-17: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	62
Tabella 3-18: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	62
Tabella 3-19: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	63

Tabella 3-20: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	63
Tabella 3-21: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	64
Tabella 3-22: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	64
Tabella 3-23: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	65
Tabella 3-24: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	65
Tabella 3-25: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	66
Tabella 3-26: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	66
Tabella 3-27: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	67
Tabella 3-28: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	67
Tabella 3-29: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	68
Tabella 3-30: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.	69

Tabella 5-1: Matrice di confusione prodotta dal modello M01-SCR, area di studio Monte Conero, con i valori percentuale di producer accuracy e user accuracy.	105
Tabella 5-2: Matrice di confusione prodotta dal modello M03-SCT, area di studio Monte Conero, con i valori percentuale di producer accuracy e user accuracy.	105
Tabella 5-3: Matrice di confusione prodotta dal modello M03-SCR, area di studio Gola di Frasassi, con i valori percentuale di producer accuracy e user accuracy.	107
Tabella 5-4: Matrice di confusione prodotta dal modello M04-SCR, area di studio Gola di Frasassi, con i valori percentuale di producer accuracy e user accuracy.	107

ELENCO DELLE FIGURE

Figura 1-1: Schema di rappresentazione del workflow generale.	21
Figura 2-1: Localizzazione dell'area di studio Gola di Frasassi (foto di Giacomo Quattrini)	23
Figura 2-2: Dettaglio sulla geologia per l'area di studio Gola di Frasassi, estratta dalla banca dati geologica della Regione Marche (scala 1:10000)	24
Figura 2-3: Diagramma semplificato di Bagnolus e Gaussien per la stazione Valmontagnana.	25
Figura 2-4: Localizzazione dell'area di studio Monte Conero.....	28
Figura 2-5: Dettaglio sulla geologia per l'area di studio Monte Conero, estratta dalla banca dati geologica della Regione Marche (scala 1:10000)	29
Figura 2-6: Diagrammi di Walter e Lieth delle stazioni termo-pluviometriche M. Cappuccini e Ancona Torette.....	30
Figura 2-7: Configurazione orbitale dei satelliti Sentinel-2.....	34
Figura 2-8: Esempio di scree-plot degli eigenvalues per ogni componente principale estratta dalla FPCA. Nelle ascisse il numero di componenti e nelle ordinate la frazione di varianza spiegata, assoluta e cumulativa	46
Figura 2-9: Distribuzione spaziale dei training data per l'area di studio Monte Conero ..	49
Figura 2-10: Distribuzione spaziale dei training data per l'area di studio Gola di Frasassi.	50
Figura 2-11: Schema semplificato di funzionamento della cross validation.	51
Figura 3-1: Grafico boxplot relativo ai valori di overall accuracy (OA) per metodo (M01, M02, M03, M04) e area di studio (CO, VA). M03 si dimostra il metodo più performante... 70	
Figura 3-2: Grafico boxplot relativo ai valori di overall accuracy (OA) per indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI) e metodo di pulizia (bp, tsc). La funzione tsclean si dimostra il metodo più performante per gli indici GNDVI, MCARI, NDMI, NDRE, NDVI e RI.....	71
Figura 3-3: Grafico boxplot relativo ai valori di overall accuracy (OA) per indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI) e metodo di applicazione della GAM e	

delle aggregazioni (ag, doy). La modalità di applicazione della GAM sui dati giornalieri precedente alle aggregazioni si dimostra il metodo più performante per gli indici GNDVI, MCARI, MNDWI, NDMI, NDRE e RI.	71
Figura 3-4: Grafico boxplot relativo ai valori di overall accuracy (OA) per predittori (SCR, SCT, TS, TSR) e area di studio (CO, VA). SCT e SCR si dimostrano i predittori più performanti.....	72
Figura 3-5: Grafico boxplot relativo ai valori di overall accuracy (OA) per indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI).....	73
Figura 5-1: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice RI.....	77
Figura 5-2: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice RI.....	78
Figura 5-3:Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice RI.....	78
Figura 5-4: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice RI.....	79
Figura 5-5: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice RI.	79
Figura 5-6: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice RI.	80
Figura 5-7: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice RI.	80
Figura 5-8: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice RI.	81
Figura 5-9: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice MCARI.	81
Figura 5-10: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice MCARI.	82
Figura 5-11: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice MCARI.	82
Figura 5-12: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice MCARI.	83
Figura 5-13: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice MCARI.....	83

Figura 5-14: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice MCARI.....	84
Figura 5-15: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice MCARI.....	84
Figura 5-16: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice MCARI.....	85
Figura 5-17: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice GNDVI.....	85
Figura 5-18: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice GNDVI.....	86
Figura 5-19: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice GNDVI.....	86
Figura 5-20: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice GNDVI.....	87
Figura 5-21: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice GNDVI.	87
Figura 5-22: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice GNDVI.	88
Figura 5-23: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice GNDVI.	88
Figura 5-24: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice GNDVI.	89
Figura 5-25: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice NDVI.	89
Figura 5-26: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice NDVI.	90
Figura 5-27: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice NDVI.	90
Figura 5-28: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice NDVI.	91
Figura 5-29: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice NDVI.....	91
Figura 5-30: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice NDVI.....	92

Figura 5-31: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice NDVI.....	92
Figura 5-32: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice NDVI.....	93
Figura 5-33: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice NDRE.....	93
Figura 5-34: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice NDRE.....	94
Figura 5-35: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice NDRE.....	94
Figura 5-36: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice NDRE.....	95
Figura 5-37: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice NDRE.....	95
Figura 5-38: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice NDRE.....	96
Figura 5-39: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice NDRE.....	96
Figura 5-40: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice NDRE.....	97
Figura 5-41: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice MNDWI.....	97
Figura 5-42: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice MNDWI.....	98
Figura 5-43: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice MNDWI.....	98
Figura 5-44: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice MNDWI.....	99
Figura 5-45: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice MNDWI.....	99
Figura 5-46: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice MNDWI.....	100
Figura 5-47: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice MNDWI.....	100

Figura 5-48: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice MNDWI.	101
Figura 5-49: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice NDMI.	101
Figura 5-50: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice NDMI.	102
Figura 5-51: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice NDMI.	102
Figura 5-52: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice NDMI.	103
Figura 5-53: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice NDMI.	103
Figura 5-54: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice NDMI.	104
Figura 5-55: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice NDMI.	104
Figura 5-56: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice NDMI.	105
Figura 5-57: Mappa della vegetazione prodotta dal modello M01-SCR per l'area di studio Monte Conero	106
Figura 5-58: Mappa della vegetazione prodotta dal modello M03-SCT per l'area di studio Monte Conero	106
Figura 5-59: Mappa della vegetazione prodotta dal modello M03-SCR per l'area di studio Gola di Frasassi.....	108
Figura 5-60: Mappa della vegetazione prodotta dal modello M04-SCR per l'area di studio Gola di Frasassi.....	109

ACRONIMI E ABBREVIAZIONI

LSP	LAND SURFACE PHENOLOGY
ESA	EUROPEAN SPACE AGENCY
MSI	MULTI SPECTRAL INSTRUMENT
STL	SEASONAL AND TREND DECOMPOSITION USING LOESS
NA	NO-DATA
GAM	GENERALIZED ADDITIVE MODEL
PCA	PRINCIPAL COMPONENT ANALYSIS
FPCA	FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS
RF	RANDOM FOREST
RI	REDNESS INDEX
TS	SERIE TEMPORALI
TSR	SERIE TEMPORALI RICOSTRUITE
SCT	SCORES TOTALI
SCR	SCORES RIDOTTI
OA	OVERALL ACCURACY

Capitolo 1

INTRODUZIONE E SCOPO DELLA TESI

La parola “fenologia” deriva dal greco “phaino” che significa mostrare o apparire. Viene definita come la parte dell’ecologia che studia i rapporti tra i fattori climatici (temperatura, umidità, fotoperiodo) e la manifestazione stagionale di alcuni fenomeni della vita vegetale e animale (Lieth, 1974). La fenologia vegetale in particolare si occupa della definizione delle fasi di sviluppo delle piante (o fasi fenologiche), ovvero i cicli biologici che si verificano durante la vita annuale delle diverse specie vegetali, la cui dinamica è in gran parte guidata dalle condizioni ambientali locali e globali. Germinazione, germogliamento, comparsa delle foglie, fioritura e fruttificazione sono fasi tipicamente primaverili o estive; cambiamenti di colorazione e caduta delle foglie sono legate al periodo autunnale (Lieth, 1973). Il clima è il principale fattore di controllo dei diversi cicli biologici delle piante (Walther *et al.*, 2002; Badeck *et al.*, 2004; Wolkovich *et al.*, 2012). Le componenti fisiche e biologiche (proprietà del suolo, condizioni topografiche, competizione intraspecifica o interspecifica e diversità genotipica) altresì contribuiscono (Caparros-Santiago, Rodriguez-Galiano and Dash, 2021). A loro volta, molte importanti risposte del sistema climatico sono influenzate dalla stagionalità della vegetazione (Richardson *et al.*, 2013). Cambiamenti nel ciclo di vita delle piante, associati a processi importanti quali fotosintesi, assorbimento di CO₂ e traspirazione, possono avere impatti significativi sui fenomeni atmosferici (ripartizione dell'energia sulla superficie terrestre), sui cicli di acqua e nutrienti e i flussi di carbonio (Garrity *et al.*, 2011; Mizunuma *et al.*, 2013; Caldararu, Purves and Palmer, 2014; Estiarte and Peñuelas, 2015; Forkel *et al.*, 2016).

Natura 2000 è il principale strumento della politica dell'Unione Europea per la conservazione della biodiversità, nonché la più grande rete coordinata di aree protette al mondo. Estesa sul territorio comunitario per oltre il 18% delle terre emerse e l'8% delle superfici marine (Commissione Europea, 2013), si tratta di una rete ecologica istituita ai sensi della Direttiva (CEE) n. 43/92 del Consiglio del 21 maggio 1992 “Habitat”, relativa alla conservazione degli habitat naturali e seminaturali e della flora e della fauna selvatiche. Insieme alla Direttiva (CE) n. 147/2009 del Parlamento Europeo e del Consiglio del 30

novembre 2009 “Uccelli”, concernente la conservazione degli uccelli selvatici (in precedenza 1979/409/CEE), costituiscono il cuore della strategia comunitaria in materia di conservazione. Secondo quanto previsto dall'articolo 11 della Direttiva Habitat, gli Stati Membri sono tenuti a garantire la sorveglianza dello stato di conservazione degli habitat (elencati nell'Allegato I) e delle specie (elencate negli Allegati II, IV e V) di interesse comunitario su tutto il territorio nazionale. Ogni sei anni i risultati del monitoraggio devono essere trasmessi alla Commissione Europea attraverso l'elaborazione di un rapporto nazionale sullo stato di attuazione delle disposizioni, in accordo all'articolo 17. I dati devono essere riportati, per ogni habitat e specie, a livello biogeografico e cartografati nelle modalità definite dalla Commissione Europea. Il monitoraggio dello stato di conservazione è un'attività indispensabile anche per valutare il raggiungimento di quanto previsto all'articolo 2 della Direttiva Uccelli, ovvero il conseguimento per tutte le specie di avifauna di un livello adeguato di conservazione. Sebbene l'articolo 12 della Direttiva Uccelli non preveda esplicitamente di includere nei rapporti nazionali sull'attuazione della Direttiva la valutazione dello stato di conservazione delle specie inserite negli Allegati, la Commissione Europea e Stati Membri hanno concordato sull'opportunità di inserire tali valutazioni nei rapporti, al fine di allineare i processi di monitoraggio e rendicontazione delle Direttive Habitat e Uccelli. Con l'approvazione del Regolamento (UE) n. 1010/2019 del Parlamento europeo e del Consiglio, del 5 giugno 2019, recante norme finalizzate all'armonizzazione degli obblighi di comunicazione nella normativa ambientale europea, tale allineamento è stato formalizzato. In questo scenario, in cui si rende necessario al fine del raggiungimento dell'obiettivo di conservazione, un'osservazione periodica costante delle dinamiche evolutive degli ambienti naturali e semi-naturali, la fitosociologia si dimostra essere uno strumento fondamentale a disposizione degli Stati Membri e di tutta la comunità scientifica (Evans, 2006; Biondi *et al.*, 2012; Pesaresi *et al.*, 2020). Le mappe fitosociologiche della vegetazione consentono di utilizzare le comunità vegetali, descritte dalla distribuzione topografica e dall'organizzazione in fitocenosi, come indicatori ambientali e strumenti di indagine ecologica. La classificazione fitosociologica coglie l'esistenza di raggruppamenti di piante non casuali che ripetendosi con una certa costanza, pur essendo variabili, sono ben definiti ecologicamente. In questo senso le mappe fitosociologiche possono rappresentare adeguatamente la distribuzione degli habitat e, se ripetute nel tempo, risulterebbero utili a valutarne lo stato di conservazione (Gigante *et al.*, 2016; Viciani *et al.*, 2016). La mappatura “tradizionale” della vegetazione prevede il rilevamento in campagna delle associazioni vegetali. Richiede l'utilizzo di risorse e tempo, tali da rendere particolarmente oneroso l'aggiornamento delle mappe degli habitat. È

necessario, infatti, considerare le difficoltà legate all'accessibilità dei siti, all'estensione delle aree da rilevare, alla soggettività nelle operazioni di spazializzazione e fotointerpretazione, nonché alla mancanza di un vero e proprio parametro di accuratezza nella produzione delle carte (Pesaresi *et al.*, 2020). Ichter *et al.* (2017) riportano che, dopo l'adozione della Direttiva Habitat, al fronte di un aumento nella produzione di mappe con approccio fitosociologico, si osservano un numero decisamente esiguo di cartografie aggiornate.

Negli ultimi quarant'anni, il telerilevamento satellitare ha acquisito un ruolo fondamentale nello studio delle dinamiche evolutive della vegetazione in funzione di una migliore comprensione della diversità degli habitat (Tucker, 1979; Polgar, Gallinat and Primack, 2014; Corbane *et al.*, 2015; Zlinszky *et al.*, 2015; Dash and Ogutu, 2016; Pesaresi *et al.*, 2020). Le informazioni fornite dai dati tele rilevati si dimostrano un importante strumento complementare alle indagini sul campo, con molteplici vantaggi rispetto alla mappatura tradizionale, in termini di velocità di produzione delle cartografie, ripetibilità del monitoraggio (anche in aree inaccessibili) e copertura globale (vanden Borre *et al.*, 2011). Le piante assorbono, trasmettono e riflettono la radiazione solare descrivendo un determinato comportamento fenologico. La finestra spettrale compresa tra 400 e 700 nm (luce visibile) viene definita intervallo attivo fotosintetico ed è sfruttata all'interno dei pigmenti nei tessuti fogliari, per il funzionamento delle reazioni biochimiche legate alla fotosintesi. L'energia assorbita nelle altre regioni dello spettro elettromagnetico è generalmente inferiore poiché poco utile per i processi fotochimici e spesso dannosa. Il telerilevamento passivo si basa sull'acquisizione dell'informazione spettrale dalla frazione di radiazione solare riflessa dalla superficie terrestre. Questa viene intercettata dai sensori montati sulle piattaforme satellitari e registrata in forma di dato discreto in specifiche bande spettrali in funzione della lunghezza d'onda del segnale. Le bande descrivono le diverse proprietà di riflettanza associate alle tipologie di copertura del suolo, compresa quindi la vegetazione. Dalle modalità con le quali le bande possono essere combinate algebricamente, vengono conformati utili indicatori fenologici, chiamati indici di vegetazione, che esprimono i cambiamenti nella risposta spettrale della copertura vegetale. Un numero limitato di studi ha esplorato l'utilizzo di variabili biofisiche derivate direttamente da dati satellitari, cioè proprietà della vegetazione che possono essere quantificate (pigmenti, componenti chimiche o strutturali) e correlate all'attività fotosintetica, alla produttività o a condizioni di deperimento (Verrelst *et al.*, 2019). La crescente offerta di dati satellitari gratuiti, ad alta risoluzione spaziale e temporale, in particolare l'archivio delle missioni Landsat (NASA) e le più recenti missioni Sentinel (ESA), consente di estrapolare il comportamento spettrale della vegetazione durante tutto l'arco

dell'anno. A partire dalla disponibilità di un dato multi-temporale è possibile costruire delle serie storiche che permettano di quantificare i cambiamenti stagionali e inter-annuali nella fenologia. Gli indici, calcolati da combinazioni tra bande o derivati da variabili biofisiche fungono da indicatori di vigore e abbondanza della vegetazione (Caparros-Santiago, Rodriguez-Galiano and Dash, 2021). Questi cambiamenti sono utili a distinguere differenti tipologie di habitat, per mezzo di un processo di classificazione finalizzato alla produzione di mappe (Feret, Corbane and Alleaume, 2015; Franklin *et al.*, 2015; Marzialesi *et al.*, 2019; Rapinel *et al.*, 2019). Una serie storica è un insieme di dati raccolti in un periodo di tempo a una data frequenza temporale. Le fonti più comuni di dati satellitari, comprese le già citate Landsat e Sentinel 2, sono riassunte in Tabella 1-1 con i relativi parametri di risoluzione spaziale, spettrale e temporale. Lo studio della fenologia vegetale su scala regionale e globale, da serie storiche di dati acquisiti attraverso sensori ottici satellitari viene definita *Land Surface Phenology* (LSP).

Tabella 1-1: Comuni piattaforme di telerilevamento con specifiche generali di sensore

Satellite	Sensore	Risoluzione spaziale (m)	Bande spettrali	Frequenza di acquisizione
NOAA	AVHRR	1100	5	12 h
SPOT	Vegetation	1150	4	24 h
Envisat	MERIS	300	15	3 giorni
Terra, Aqua	MODIS	250 - 1000	36	24 h
Terra	ASTER	15 - 90	15	16 giorni
Suomi NPP	VIIRS	375	22	12 h
Landsat 7, 8	ETM+, OLI, TIRS	15 - 100	8 - 11	16 giorni
Sentinel 2	MSI	10 - 60	13	5 giorni

I valori di riflettanza spettrale sono soggetti ad impurità, spesso causate da condizioni atmosferiche avverse e copertura nuvolosa, ma anche disturbi indotti dall'effetto del suolo, dall'ombreggiamento o malfunzionamenti del sensore. Questi artefatti possono alterare l'andamento temporale degli indici di vegetazione, causando una riduzione della precisione delle stime fenologiche. Pertanto, ai fini del corretto processamento del dataset, è necessario operare una fase di *data cleaning*, quindi individuazione ed eliminazione dei valori anomali (*outliers*). In generale, è anomalo un valore che si discosta in modo significativo dal suo corrispettivo atteso. Nelle serie temporali uni-variate, nelle quali il fenomeno è descritto da una sola variabile, possono essere considerati anomali valori puntuali e porzioni di curve. I metodi di individuazione dei valori anomali puntuali sono i più citati in letteratura e tra questi,

le tecniche cosiddette *model-based*. Un punto al tempo t può essere dichiarato anomalo se la distanza dal suo valore atteso è maggiore di una soglia predefinita. Ogni tecnica si basa sull'adattamento di un modello e il calcolo del valore atteso avviene in modo differente. Le tecniche si differenziano nei metodi di stima (*estimation model-based*) e di predizione (*prediction model-based*) in funzione del tipo di osservazioni utilizzate (precedenti e/o successive i valori correnti). Gli *estimation model-based* sfruttano la statistica di base, gli algoritmi di *smoothing* e l'analisi dei residui e vengono usati per lo studio delle serie temporali *non-streaming* (non aggiornate in tempo reale). I *prediction model-based* consistono nella costruzione di un modello e nel suo progressivo aggiornamento man mano che nuove osservazioni subentrano; pertanto, sono usati nelle serie temporali *streaming*. Gli altri metodi di individuazione dei valori anomali nelle serie temporali sono chiamati *density-based* e *histogramming*. Le tecniche comprese all'interno del primo gruppo considerano anomali i punti in funzione del numero di valori vicini precedenti e successivi. Il rilevamento dei valori anomali *density-based* è stato ampiamente usato per i dati non temporali, in quanto il concetto di valore vicino è più complesso in dati ordinati temporalmente. Il metodo *histogramming* si basa sulla suddivisione delle serie temporali in blocchi ordinati nei quali viene calcolata la media dei valori. Sono anomali quei punti la cui rimozione determina una riduzione del numero di blocchi (Blázquez-García *et al.*, 2020).

Nella LSP, le curve fenologiche sono costruite per mezzo di un procedimento di *smoothing*, cioè di lisciamento del dato discreto. In statistica, il lisciamento consiste nell'applicazione di una funzione di filtro il cui scopo è evidenziare i *pattern* significativi, attenuando il rumore generato da artefatti ambientali, informatici o fisiologici. Il processo consiste nel mediare i dati di una serie con valori contigui in modo da descriverne un andamento. Sono stati sviluppati diversi algoritmi matematici e in questa sede viene utilizzato un *Generalized Additive Model* (GAM) basato sulla *Cyclic Cubic Spline Regression*. Il lisciamento può precedere o essere seguito da una fase di aggregazione dei dati, effettuata per riempire quelle lacune temporali nelle quali non sono presenti immagini. Il prodotto della fase di aggregazione sono le immagini composite, nelle quali ogni pixel è descritto da un andamento fenologico distribuito in uno specifico intervallo temporale. I dati possono essere aggregati per annualità (giorni dell'anno), mensilità (semestre, trimestre, bimestre) o settimane (settimane, bisettimane). Entrambe queste fasi sono importanti per la definizione delle singole curve e i processamenti ai quali possono essere sottoposte.

Questo lavoro si pone l'obiettivo di valutare gli effetti sulla classificazione della vegetazione mediante procedimenti di apprendimento supervisionato che prevedono differenti

metodologie di pre-processamento delle serie temporali (Figura 1-1). Vengono testate 2 modalità di individuazione dei valori anomali e 2 modalità di applicazione dell'algoritmo di *smoothing*. Da queste si estraggono 4 blocchi di serie temporali da utilizzare direttamente come variabili di predizione e, in un secondo momento, da assoggettare a *Functional Principal Component Analysis* (FPCA). La FPCA è una tecnica statistica che permette di riassumere l'informazione legata alla variabilità temporale delle curve in un certo numero di componenti principali. I coefficienti che quantificano il valore di ogni singola componente sono chiamati *scores*, questi incarnano l'ordine cronologico delle variazioni funzionali e possono essere utilizzati in analisi successive consentendo il confronto e l'interpretazione dei risultati da diversi punti di vista per lo stesso habitat (Pesaresi et al., 2020). Gli *scores* estratti vengono utilizzati, nel totale e in una loro frazione, come seconde e terze variabili di predizione per la classificazione. Infine, da queste ultime si ricostruiscono le serie temporali originali, da classificare come quarte e ultime variabili di predizione. Le immagini, acquisite dalla piattaforma satellitare Sentinel-2, si riferiscono a un periodo di quattro anni (2017 – 2020). Le variabili fenologiche utilizzate sono sette indici di vegetazione: *Normalized Difference Vegetation Index* (NDVI), *Modified Chlorophyll Absorption in Reflectance Index* (MCARI), *Green Normalized Difference Vegetation Index* (GNDVI), *Normalized Difference Red/Green Redness Index* (RI), *Normalized Difference Red-Edge* (NDRE), *Normalized Difference Moisture Index* (NDMI), *Modified Normalized Difference Water Index* (MNDWI). L'algoritmo di classificazione proposto è *Random Forest* e i dati per l'addestramento vengono forniti dai rilievi fitosociologici precedentemente effettuati nelle aree di studio. Il lavoro si inserisce in una filiera di studi svolti in collaborazione tra il Dipartimento di Scienze Agrarie, Alimentari e Ambientali e il Dipartimento di Ingegneria Informatica dell'Università Politecnica delle Marche.

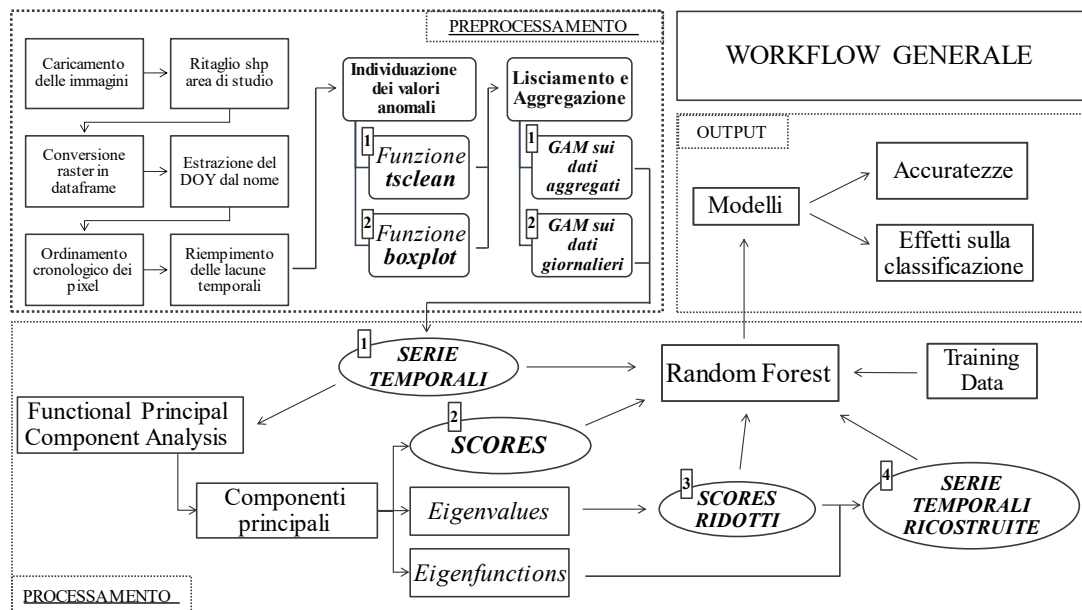


Figura 1-1: Schema di rappresentazione del workflow generale. Vengono evidenziate le fasi di pre-processamento del dataset (costruzione delle serie temporali, metodi di individuazione dei valori anomali e tecniche di lisciamiento e aggregazione) poiché focali ai fini del raggiungimento degli obiettivi del lavoro.

Capitolo 2

MATERIALI E METODI

Nella sezione seguente vengono forniti gli strumenti per comprendere il contesto e le elaborazioni effettuate. Sono introdotte le due aree di studio, il Monte Conero e la Gola di Frasassi, le specifiche delle immagini utilizzate nonché le fasi principali di costruzione e processamento delle serie temporali. Per ogni passaggio topico sono riportati i codici implementati nel software di programmazione. Il presente lavoro pone un accento marcato sulla componente informatica, legata alla stesura e al miglioramento degli *script*.

2.1 Inquadramento territoriale

2.1.1 Gola di Frasassi

L'area di studio coincide con la Zona Speciale di Conservazione "Gola di Frasassi IT5320003", si estende per un'area di 728 ha all'interno del Parco Naturale Regionale della Gola della Rossa e di Frasassi tra i comuni di Genga e Fabriano in provincia di Ancona (Figura 2-1). Il territorio è compreso nella fascia montana preappenninica, comunemente chiamata Dorsale Marchigiana Esterna. Il limite dell'area a nord è costituito dai rilievi di Monte la Croce e Monte Ginguno e prosegue ad est lungo la strada Provinciale che da Pierosara si dirige verso San Vittore. Il confine sud è identificabile nell'allineamento abitato di Valtreara – Monte di Valmontagnana - Monte Civitella. Verso ovest la ZSC è delimitata dal nucleo abitato di Spineto e l'area industriale di Pianello. La quota massima del sito è di 931.2 m.s.l.m. del Monte di Valmontagnana, mentre la quota minima rilevata è di 200 m.s.l.m. al limite con la sponda sinistra del fiume Esino (Mini *et al.*, 2015).



Figura 2-1: Localizzazione dell'area di studio Gola di Frasassi (foto di Giacomo Quattrini)

La Gola di Frasassi si colloca all'interno dell'anticlinale di Monte Valmontagnana – Monte Frasassi. Dal punto di vista geologico il Calcare Massiccio costituisce l'ossatura di tutto il sistema montuoso locale. Il Calcare Massiccio è una formazione calcarea compatta, originata nel Giurassico inferiore in ambiente di piattaforma carbonatica. Si tratta di facies calcaree assai pure, molto permeabili, affioranti per uno spessore di circa 500 m senza che sia raggiunto il letto. Il Calcare Massiccio è sormontato da formazioni calcaree e calcareo marnose di ambiente pelagico (formazioni del Bugarone, Maiolica, Scaglia bianca e rossa) databili a cavallo tra il Giurassico medio e l'Eocene, con spessore complessivo di poche centinaia di metri. All'interno di questa successione è presente un orizzonte marnoso poco permeabile (Marne a Furoidi) dallo spessore di circa 50 m, che riveste un importante ruolo nella circolazione idrica sotterranea. Alla base della successione carbonatica è presente un importante orizzonte di rocce evaporitiche note come Anidridi di Burano, di età triassica. Di più recente formazione sono invece i depositi arenaceo-pelitici riferibili al Miocene e i depositi detritici di falda, di composizione variabile, spesso costituiti da elementi a spigolo vivo, riferibili all'Olocene (Figura 2-2).

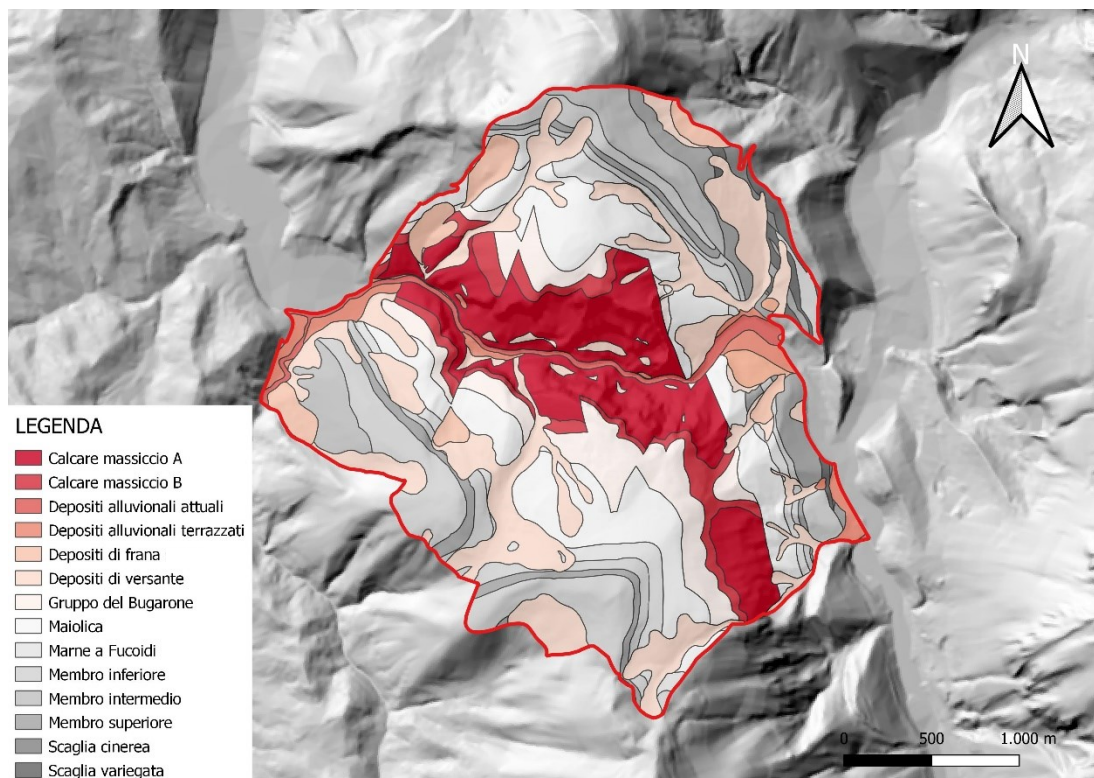


Figura 2-2: Dettaglio sulla geologia per l'area di studio Gola di Frasassi, estratta dalla banca dati geologica della Regione Marche (scala 1:10000)

L'assetto topografico dell'area è da ricondurre alla paleo-geografia regionale. I fenomeni di regressione e avanzamento della linea di costa, combinati al perdurare delle spinte orogeniche, hanno incrementato l'effetto erosivo dei corsi d'acqua sui rilievi carbonatici. Da qui hanno avuto origine le tipiche morfologie che si osservano nelle zone di gola diffuse sul territorio regionale e, nel caso specifico, della Gola di Frasassi. Il reticolo idrografico attuale è costituito da piccoli corsi d'acqua soggetti a portate variabili e, spesso, secche stagionali. L'elemento principale è rappresentato dal fiume Sentino che attraversa la gola e si immette nelle acque del fiume Esino in prossimità di San Vittore.

I dati presi in esame per la caratterizzazione climatica dell'area si riferiscono alla stazione di Valmontagnana, di cui sono disponibili 17 anni di osservazione. Il diagramma termopluviometrico di Bagnouls e Gaussen, mostra che le precipitazioni, dal minimo estivo, crescono rapidamente nel periodo autunnale fino al massimo del mese di ottobre, decrescono fino al mese di gennaio per poi mantenersi sui valori intermedi nel periodo primaverile. A un periodo autunnale e primaverile con piovosità relativamente elevata fa quindi riscontro una marcata diminuzione nella stagione estivo. Per quanto riguarda le temperature, il mese più caldo in assoluto è luglio, condizione ricorrente per le stazioni continentali marchigiane. Il

Il mese più freddo è gennaio, seguito da dicembre e febbraio; l'andamento delle temperature è regolare, con aumento da gennaio ad agosto e poi un graduale decremento. Il pattern termopluviometrico è tipico di stazioni a clima temperato con caratteri di mediterraneità, in cui l'effetto del mare prolunga la stagione estiva verso l'inverno. Il raffronto tra temperature e precipitazioni evidenzia una limitata aridità estiva, che si verifica verosimilmente nelle annate più siccitose (Figura 2-3, Mini et al., 2015).

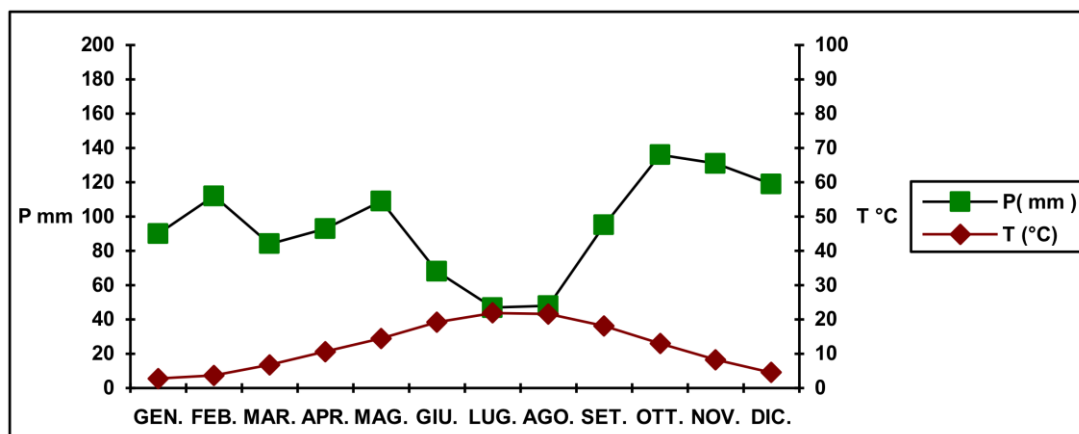


Figura 2-3: Diagramma semplificato di Bagnolus e Gaussen per la stazione Valmontagnana.

L'attuale assetto di uso del suolo è il risultato dell'interazione delle attività dell'uomo con le caratteristiche strutturali (clima, orografia e geologia) del territorio. Lo sfruttamento delle aree naturali seguiti dallo sviluppo delle attività industriali e dalla realizzazione di infrastrutture sono stati i fattori che più di tutti hanno modellato l'attuale paesaggio. Osservando le principali categorie d'uso del suolo è possibile notare che le aree boscate sono la componente che domina il paesaggio. Le aree agricole costituiscono circa il 30% della superficie, a queste si aggiunge circa il 7-8% di zone contraddistinte da vegetazione erbacea o arbustiva, non prettamente agricole (praterie, aree arbustive e boschive in evoluzione). Le aree urbanizzate sono rappresentate da piccoli nuclei abitati e edifici rurali isolati ma soprattutto dalle aree industriali e dalle infrastrutture di fondovalle (Mini *et al.*, 2015).

2.1.1.1 Vegetazione

L'area è caratterizzata da un elevata eterogeneità vegetazionale dovuta alla variabilità topografica e litologica e all'interazione continua tra uomo e ambiente. Lo sfruttamento delle aree naturali per le attività agrosilvopastorali sono stati importanti fattori di modellamento del paesaggio. L'area è coperta da boschi per gran parte della sua superficie. Un'estesa zona pascoliva copre la parte sommitale del monte Valmontagnana. I versanti meno acclivi della zona orientale sono ricoperti da cedui spesso invecchiati o irregolari di leccio che si alternano

a boschi di carpino. Nel dettaglio, si nota che la specie più diffusa è il leccio spesso in popolamenti quasi puri che occupano il 25% dell'area, mentre le carpinete sono più frequenti nei versanti esposti a nord. Le formazioni a prevalenza di roverella sono localizzate dove si presentano situazioni migliori per giacitura e costituiscono complessivamente il 15% della superficie totale. Sono presenti rimboschimenti di pino nero soprattutto nell'area occidentale del sito che guarda la valle del torrente Sentino. L'area di studio, oltre ad una considerevole ricchezza floristica, presenta una importante diversità fitocenotica e degli habitat. La loro organizzazione spaziale è rappresentata dalle mappe della vegetazione e degli habitat di. Le principali associazioni vegetali presenti si riportano di seguito (Biondi, Casavecchia, Galassi, *et al.*, 2009).

Cytiso sessilifolii – *Quercetum pubescentis*: boschi a dominanza di *Quercus pubescens*, con una canopy piuttosto aperta (85%) ed una altezza media di 7,5 m. Nello strato arboreo è costantemente presente *Fraxinus ornus*, spesso *Ostrya carpinifolia* mentre è meno frequente *Acer obtusatum*. Lo strato arbustivo (*Juniperus oxycedrus*, *Cotinus coggygra*, *Spartium junceum* e *Cytisus sessilifolius*) è piuttosto abbondante. Nello strato erbaceo sono frequenti *Osyris alba*, *Brachypodium rupestre* e *Teucrium chamaedrys*.

Cephalanthero longifoliae – *Quercetum ilicis*: boschi a dominanza di *Quercus ilex*, con una canopy totale (100%) e un'altezza media di 7 m. Si sviluppa in aree molto pendenti su litologie di calcare massiccio e formazione del bugarone. Nello strato arboreo sono presenti *Fraxinus ornus*, *Acer monspessulanum*, *Ostrya carpinifolia*, *Quercus pubescens* e *Phillyrea latifolia*. Nello strato erbaceo sono frequenti: *Melica uniflora*, *Melittis melissophyllum*, *Rubia peregrina*, *Arabis turrita* e *Cyclamen hederifolium*.

Scutellario columnae – *Ostryetum carpinifoliae*: boschi a dominanza di *Ostrya carpinifolia*, con una canopy piuttosto densa e chiusa (95%) ed una altezza arborea media di 13 m. Nello strato arboreo sono presenti *Acer obtusatum*, *Fraxinus ornus*, *Quercus cerris*, *Corylus avellana*, *Sorbus torminalis* e *Quercus pubescens*. Nello strato arbustivo sono costantemente presenti, *Cotinus coggygria*, *Cornus mas*, *Crataegus monogyna* e *Euonymus latifolius*. Nello strato erbaceo invece sono frequenti *Cyclamen hederifolium*, *Helleborus foetidus*, *Daphne laureola*, *Euphorbia amygdaloides*, *Hepatica nobilis*, *Buglossoides purpureocaerulea*, *Lathyrus venetus*, *Viola reichenbachiana*, *Primula vulgaris*, *Helleborus viridis* ssp. *bocconeii*, *Asparagus acutifolius* e spesso *Ruscus aculeatus*.

Asparago acutifolii – *Ostryetum carpinifoliae*: boschi a dominanza di *Ostrya carpinifolia*, con una canopy piuttosto chiusa (95%) ed una altezza media di circa 10 m. Nello strato arboreo sono frequenti ed abbondanti *Quercus ilex*, *Acer monspessulanum*, *Acer obtusatum*, *Quercus*

pubescens, *Sorbus aria* e *Phillyrea latifolia*. Nello strato arbustivo sono frequenti *Cotinus coggygia*, *Coronilla emerus*, *Juniperus oxycedrus*, *Viburnum tinus* e *Crataegus monogyna*. Nello strato erbaceo, invece sono frequenti *Daphne laureola*, *Euphorbia amygdaloides*, *Asparagus acutifolius*, *Hepatica nobilis*, *Brachypodium rupestre*, *Rubia peregrina*, *Ceterach officinarum*, *Arabis turrata*, *Echinops sicalus* e *Cyclamen hederifolium*.

Spartio juncei – *Cytisetum sessilifolii* var. a *Spartium junceum*: cespuglieti e mantelli di vegetazione caratterizzati in genere da una elevata copertura. Si sviluppa prevalentemente su aree poco pendenti e su litologie calcareo-marnose in seguito al recupero di aree prative abbandonate. Questa variante è legata a suoli più o meno evoluti e profondi ed è caratterizzata dalla rilevante presenza di *Spartium junceum*.

Spartio juncei – *Cytisetum sessilifolii* var. a *Juniperus oxycedrus* e *Juniperus Communis*: si sviluppa nei settori sommitali del Monte di Valmontagnana sui substrati calcarei. Questa variante è legata a suoli poco evoluti e ciottolosi (litosuoli) ed è caratterizzata dalla rilevante presenza di *Juniperus oxycedrus* e *Juniperus Communis*.

Asperulo purpureae – *Brometum erecti*: formazioni erbose secche seminaturali e facies coperte da cespugli su substrato calcareo della classe *Festuco* – *Brometalia*. La fitocenosi occupa perlopiù i settori sommitali di Monte Valmontagnana sui substrati calcarei e calcareo-marnosi.

Boschi ripariali: formazioni forestali attribuite al *Salici albae* – *Populetem nigrae* e al *Rubo* – *Salicetum albae*, occupano i settori planiziali sui depositi alluvionali recenti con elevata umidità.

2.1.2 Monte Conero

L'area di studio comprende una porzione di territorio a cavallo tra le Zone Speciali di Conservazione "Monte Conero IT5320007" e "Portonovo e falesia calcarea a mare IT5320007". Si estende per un totale di 650 ha all'interno del Parco Naturale Regionale del Conero in provincia di Ancona, tra i comuni di Sirolo e Ancona (Figura 2-4). Il Monte Conero è un promontorio di origine calcarea di 582 m.s.l.m, essendo l'unico tratto di costa rocciosa calcarea da Trieste al Gargano, interrompe la continuità del litorale basso e sabbioso tipica della costa Adriatica. Rappresenta un punto di riferimento per gli uccelli migratori ed ha una grande importanza biogeografica per la distribuzione delle specie vegetali (Soriga *et al.*, 2015).



Figura 2-4: Localizzazione dell'area di studio Monte Conero.

Dal punto di vista geologico il Monte Conero è la porzione dell'Appennino umbro-marchigiano che più si spinge verso oriente. Le rocce del Conero appartengono tutte alla Successione Umbro-Marchigiana (Figura 2-5); la formazione più antica affiorante è la Maiolica, del Cretacico inferiore. Dello stesso periodo, si ritrovano le Marne a furoidi, la Scaglia bianca, la Scaglia rossa, la Scaglia variegata, la Scaglia cinerea e la formazione del Bisciario. Considerando l'intero promontorio, sono visibili altre formazioni risalenti al Terziario (lo Schlier, la Formazione gessoso-solfifera, la Formazione a colombacci e l'Orizzonte del Trave)(Soriga *et al.*, 2015).

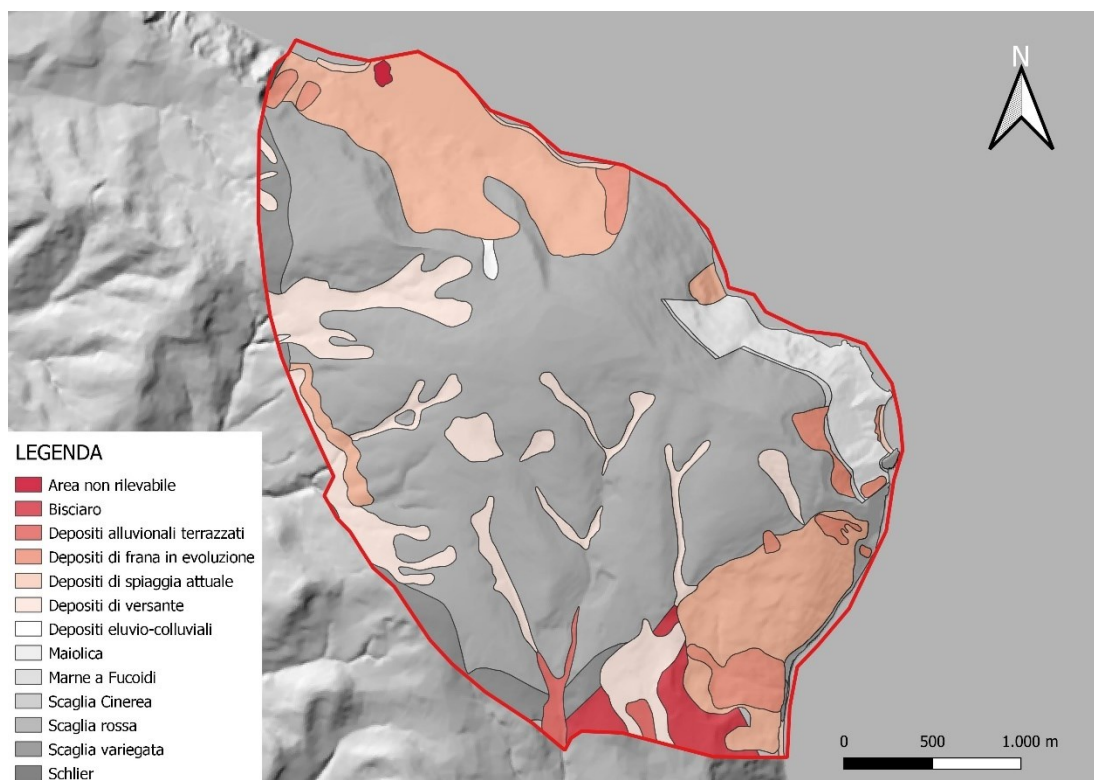


Figura 2-5: Dettaglio sulla geologia per l'area di studio Monte Conero, estratta dalla banca dati geologica della Regione Marche (scala 1:10000)

Il promontorio del Conero ha una morfologia generalmente aspra, generata dalla presenza del nucleo calcareo sul quale si appoggiano tutte le altre geologie. Il versante nord/nord-est si presenta ripido e caratterizzato da falesie a picco sul mare. La parte sommitale è sub-pianeggiante. I restanti versanti degradano dolcemente verso il settore interno, con impluvi e incisioni più o meno marcate. In merito al reticolo idrografico, essendo lo spartiacque vicinissimo al litorale, tutti i corsi d'acqua principali si dirigono verso il settore interno (ovest), contrariamente a quanto accade comunemente sul territorio regionale. I più importanti corsi d'acqua del promontorio del Conero sono i fossi Boranico e Betelico, entrambi affluenti del torrente Aspio. Il primo nasce dal Monte dei Corvi, mentre il secondo nasce nel Conero propriamente detto, dalla zona dei Falcieri. Si ricorda in particolare il Betelico, che solca profondamente il nucleo centrale del Monte che risulta così diviso in due cime: la minore, detta "di Pian Grande" e la maggiore, che è la cima vera e propria del Conero. Spostandosi verso sud, l'altro corso d'acqua che solca il nucleo del Monte è il fosso di San Lorenzo, che, dopo aver ricevuto l'apporto della fonte di Capo d'Acqua, assume il nome di Fosso dei Mulini e poi sfocia nei pressi della spiaggia di Numana bassa.

Per la caratterizzazione climatica dell'area sono stati presi in esame le stazioni termo-pluviometriche di Ancona Torrette (3 m.s.l.m.) e Monte Cappuccini (104 m.s.l.m. Ancona) (Figura 2-6, Soriga et al., 2015). I diagrammi di Walter e Lieth mostrano un andamento delle piogge caratterizzato da due massimi e due minimi relativi e da un massimo ed un minimo assoluto. Si evidenzia un carattere di sub-mediterraneità del clima con aridità estiva limitata ai mesi di luglio e agosto, mentre si può ipotizzare che questa sia particolarmente ridotta o forse assente nei settori più elevati del Conero. Il fenomeno ventoso più rilevante è la bora, un vento asciutto e freddo che interessa i versanti a nord del promontorio durante il periodo invernale. Importanti sono anche il vento di scirocco, che apporta aria calda dai quadranti meridionali, e il maestrale, da ovest.

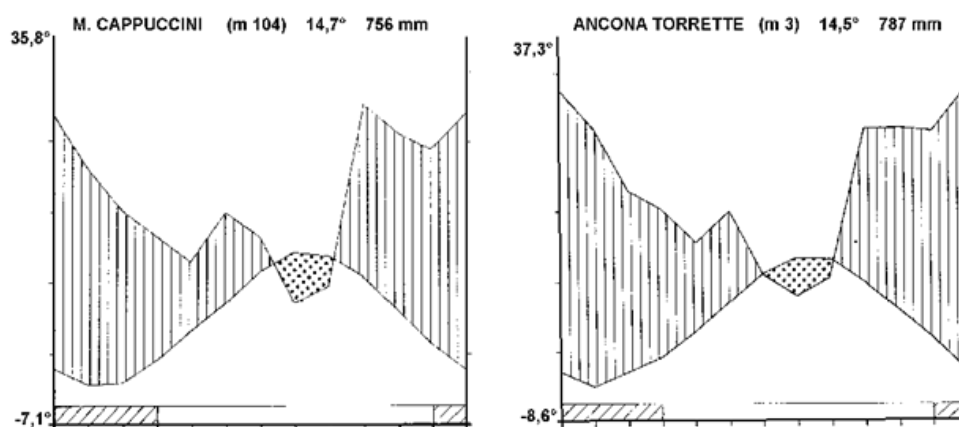


Figura 2-6: Diagrammi di Walter e Lieth delle stazioni termo-pluviometriche M. Cappuccini e Ancona Torrette.

L'uso del suolo prevalente è rappresentato dall'ecosistema naturale e seminaturale. La presenza di aree ad uso insediativo è piuttosto ridotta e legata a edifici isolati. In modo analogo le strutture commerciali. I terreni agricoli sono caratterizzati suoli rocciosi e poco sviluppati, a basse produttività, perciò destinati ad indirizzo cerealicolo-foraggero. Il territorio circostante è costituito da un sistema collinare omogeneo a pendenza variabile tipico del paesaggio marchigiano, in cui si rinvencono forti connotazioni antropiche tipiche dell'agroecosistema.

2.1.2.1 Vegetazione

L'ampia diversità biocenotica che caratterizza l'area è determinata dalla sua posizione centrale rispetto al bacino del Mare Adriatico, dall'altezza del promontorio del Conero che con i suoi 572 m supera notevolmente la media altitudinale della costa Adriatica, e dalla

diversità delle condizioni geologiche, geomorfologiche e climatiche del territorio (Biondi, Bagella and Casavecchia, 2001). L'agricoltura, il taglio e le attività pastorali che erano molto sviluppate in passato hanno contribuito in parte all'aumento della biodiversità del territorio, determinando un'alta diversificazione del paesaggio vegetale. L'abbandono di queste attività in quasi tutto il territorio ha favorito lo sviluppo di processi dinamici naturali di ricolonizzazione della vegetazione. La grande ricchezza floristica dell'area del Conero deriva dalla sua alta diversità di habitat, dalle condizioni climatiche e bioclimatiche così come dagli eventi paleogeografici del bacino adriatico. La vegetazione è composta per la quasi totalità da formazioni forestali di caducifoglie, a dominanza di carpino nero, e sempreverdi, a prevalenza di leccio e rimboschimenti a pino d'Aleppo. La composizione delle leccete si differenzia in funzione dell'esposizione dei versanti e dell'altitudine, fattori che determinano il verificarsi di ambiti microclimatici diversificati. I boschi misti di caducifoglie a prevalenza di carpino nero si ritrovano nelle esposizioni fresche settentrionali. I versanti occidentali e meridionali del promontorio sono occupati dai rimboschimenti, effettuati nel decennio tra il 1930 e 1940 per risanare le pessime condizioni idrogeologiche (disboscamento, sovrasfruttamento dei pascoli e conduzione agraria inappropriata). Le principali associazioni vegetali presenti si riportano di seguito (Biondi, Casavecchia, Angelini, *et al.*, 2009).

Cyclamino hederifolii – *Quercetum ilicis*: boschi termofili di *Quercus ilex*, con prevalenza di specie sempreverdi, principalmente distribuito lungo la costa adriatica della penisola italiana. Nell'area del Conero, l'associazione è ampiamente diffusa nel versante più caldo esposto a sud del rilievo e nelle zone meno ripide del settore settentrionale, in località protette dai venti freddi di bora. Nello strato arboreo si possono rinvenire *Fraxinus ornus*, *Phillyrea media* e *Arbutus unedo*. Altre specie caratterizzanti sono *Pistacia lentiscus*, *Cyclamen hederifolium*, *Rubia peregrina*, *Smilax aspera* e *Rosa sempervirens*.

Cephalanthero longifoliae – *Quercetum ilicis*: boschi mesofili di *Quercus ilex*, misto a *Ostrya carpinifolia*, *Acer opalus* subsp. *obtusatum* e *Fraxinus ornus*. Questi boschi sono diffusi soprattutto lungo il settore settentrionale esposto verso il mare e, sporadicamente, si presentano anche a sud, in condizioni più fresche lungo i valloni. È presente anche una variante a *Laurus nobilis*, legata a situazioni di maggiore umidità aerea. Nelle quote superiori ai 300 m la formazione si arricchisce di elementi mesofili dell'ordine *Quercetalia pubescentis* (*Querco-Fagetea*) come *Ilex aquifolium*, *Viola reichembachiana*, *Mercurialis perennis*, *Euphorbia amygdaloides*, *Lonicera caprifolium* e *Primula vulgaris* considerate specie diagnostiche della subassociazione *ruscetosum hypoglossi*.

Asparago acutifolii – *Ostryetum carpinifoliae*: boschi termofili con *Ostrya carpinifolia* dove *Fraxinus ornus* e *Acer opalus* subsp. *obtusatum* sono frequenti, e a cui si associa talvolta anche il *Quercus ilex*. Si rinvencono nella zona collinare e sporadicamente nei valloni lungo il versante settentrionale. Nello strato arbustivo sono frequenti *Cotinus coggygria*, *Coronilla emerus*, *Juniperus oxycedrus*, *Viburnum tinus* e *Crataegus monogyna*. Nello strato erbaceo, invece sono frequenti *Daphne laureola*, *Euphorbia amygdaloides*, *Asparagus acutifolius*, *Hepatica nobilis*, *Brachypodium rupestre*, *Rubia peregrina*, *Ceterach officinarum*, *Arabis turrata*, *Echinops siculus* e *Cyclamen hederifolium*.

Scutellario columnae – *Ostryetum carpinifoliae*: boschi a *Ostrya carpinifolia* con *Quercus pubescens*, *Fraxinus ornus* e *Acer opalus* subsp. *obtusatum* delle aree appenniniche. Tale formazione è localizzata nell'area più elevata del Conero, esposta a nord-est, caratterizzata da specie mesofile quali *Sorbus aria*, *Ilex aquifolium*, *Scilla bifolia*, *Mercurialis perennis*, *Ruscus hypoglossum*, *Crataegus laevigata* e *Lonicera etrusca*.

2.2 Strumenti software

2.2.1 R

R è un software open-source multiplatforma che permette la computazione statistica, è disponibile per diversi sistemi operativi, quali Linux, Windows e MacOS. R è altresì un linguaggio di programmazione, raccoglie una grande varietà di strumenti per l'analisi statistica e grafica dei dati. R prevede un'interfaccia a linea di comando, il cui vantaggio principale consiste nella tracciabilità delle procedure e dunque nella replicabilità delle analisi. Le funzioni sono contenute all'interno di librerie (o pacchetti) disponibili in un apposito sito (*repository*): il CRAN (*Comprehensive R Archive Network*), vengono caricate all'occorrenza dall'utente nell'ambiente di lavoro. R è un linguaggio di programmazione orientata agli oggetti, in informatica, la programmazione orientata agli oggetti (*object-oriented programming*) è un paradigma di programmazione che permette di definire oggetti software in grado di interagire gli uni con gli altri attraverso lo scambio di messaggi. In R qualsiasi entità è un oggetto, ognuno dei quali appartiene ad una o più classi, che definiscono i metodi e le funzioni a loro applicabili.

2.2.2 JupyterLab

JupyterLab è un'interfaccia utenti *web-based* rilasciata nel febbraio 2018 che permette la creazione di documenti e editor di testo chiamati *notebook*. La creazione di notebook permette di eseguire un codice all'interno di un elaboratore remoto utilizzando una serie di celle che

possono essere eseguite separatamente. Per lanciare un notebook è necessario impostare un kernel, cioè il linguaggio attraverso il quale verranno eseguite le istruzioni all'interno del file. JupyterLab permette anche un modello unificato per la visualizzazione e la manipolazione di diverse tipologie di dati come ad esempio i CSV. In particolare, JupyterLab è stato utilizzato per lavorare all'interno di elaboratori remoti forniti dall'Università Politecnica delle Marche. L'esecuzione degli script in Jupyter ha permesso una maggiore velocità di produzione dei dati permettendo di lavorare parallelamente sui diversi metodi di calcolo proposti.

2.3 Preprocessamento delle immagini

Le immagini satellitari rappresentano l'input per la creazione di molti prodotti a supporto di servizi di informazione e di monitoraggio del territorio, come le mappe di copertura o cambiamento di uso del suolo. Nell'ambito della LSP, il telerilevamento è uno strumento fondamentale per velocizzare e semplificare la mappatura delle associazioni vegetali. La frequenza di osservazione dei satelliti consente di comprendere il comportamento stagionale delle comunità vegetali e di descrivere la fenologia di una determinata superficie. Le caratteristiche del dataset di partenza sono descritte nei sottocapitoli seguenti. Le elaborazioni non sono effettuate nell'ambito di questo studio, tuttavia vengono riportate per fornire gli strumenti necessari a capire i passaggi successivi (Quattrini, Pesaresi and Casavecchia, 2016; Forconi, Mancini and Pesaresi, 2019; Quattrini *et al.*, 2020).

Le immagini sono state acquisite dalle piattaforme satellitari Sentinel-2A e Sentinel-2B, entrambe gestite dalla *European Space Agency* (ESA) nell'ambito del programma europeo *Copernicus* (Copernicus Europe's eyes on Earth, 2021; The European Space Agency, 2021). Il programma utilizza enormi quantità di dati globali provenienti da satelliti e da sistemi di misurazione terrestri per fornire informazioni che aiutino i prestatori di servizi, le autorità pubbliche e altre organizzazioni internazionali a migliorare la qualità della vita dei cittadini europei. I servizi di informazione forniti sono accessibili agli utenti del programma in modo libero e gratuito e rientrano in sei categorie principali: gestione del territorio, ambiente marino, atmosfera, risposta alle emergenze, sicurezza e cambiamenti climatici. I satelliti sono stati lanciati a distanza di due anni, tra giugno 2015 e marzo 2017, attraverso il vettore europeo Vega, sviluppato in collaborazione dell'Agenzia Spaziale Italiana (ASI). I due satelliti operano su una medesima orbita eliosincrona, inclinata di 98.62° rispetto all'equatore in modo da minimizzare l'impatto delle ombre sul terreno, ad una elevazione media di 786 km e sfasati di 180° (Figura 2-7; The European Space Agency, 2021).

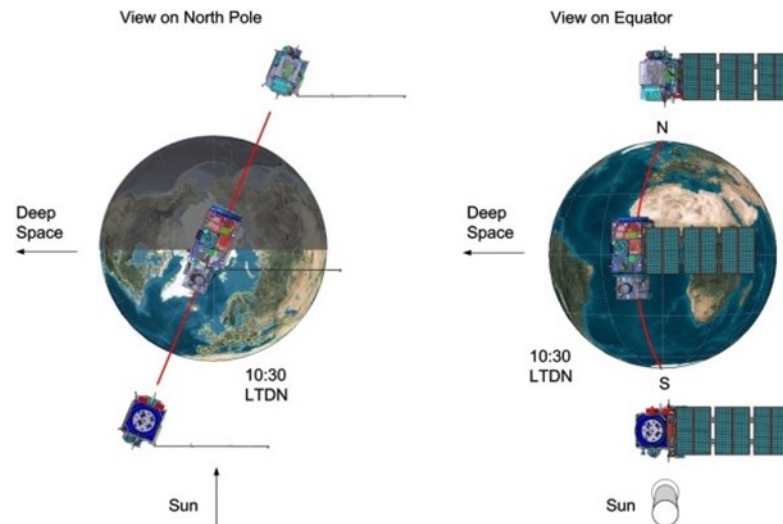


Figura 2-7: Configurazione orbitale dei satelliti Sentinel-2

Questo permette un dimezzamento dei tempi di rivisitazione che sono di 5 giorni all'equatore e di 2-3 giorni alle medie latitudini (The European Space Agency, 2021). La durata di vita stimata è di sette anni ma contengono abbastanza carburante per essere in circolo fino a 12 anni. Per mezzo della strumentazione ottica multispettrale i satelliti riescono a catturare immagini in 13 bande spettrali. Il sensore montato a bordo è chiamato *Multi Spectral Instrument* (MSI) ed è in grado di acquisire immagini con una risoluzione spaziale di 10, 20 e 60 metri nelle diverse bande (Tabella 2-1). La risoluzione radiometrica, cioè la quantità di memoria usata per immagazzinare l'informazione spettrale da ogni pixel, è di 12 bit (da 0 a 2^{12} valori di intensità luminosa).

Tabella 2-1: Specifiche del sensore MSI Sentinel-2 per ciascuna banda di acquisizione

Banda	Regione spettrale	Lunghezza d'onda	Risoluzione spaziale
1	Coastal aerosol	0.433	60
2	Blue	0.490	10
3	Green	0.560	10
4	Red	0.665	10
5	Red Edge	0.705	20
6	Red Edge	0.740	20
7	Red Edge	0.783	20
8	NIR	0.842	10
8A	Red Edge	0.865	20
9	Water vapour	0.945	60
10	SWIR	1.375	60
11	Cirrus	1.610	20

Sentinel-2 fornisce agli utenti diverse tipologie di prodotti che si differenziano per il livello di pre-processamento del dato grezzo. Le immagini di Livello – 1B sono i prodotti di livello più basso disponibili. Subiscono solo una correzione radiometrica e forniscono unicamente i valori di radianza *Top Of Atmosphere* (TOA). L'utilizzo di questi prodotti richiede dunque una conoscenza avanzata delle tecniche di orto-rettificazione. I prodotti di Livello – 1C, oltre ad una correzione radiometrica, sono sottoposti anche a una correzione geometrica che include i processi di georeferenziazione (allineamento delle immagini alla corretta posizione geografica) e di orto-rettificazione (correzione degli effetti dovuti alla direzione visuale), necessari per garantire l'esatto posizionamento delle immagini. Anche queste forniscono valori di riflettanza TOA. Le immagini di Livello – 2A derivano dalla correzione atmosferica dei relativi prodotti di Livello-1C orto-rettificati, cioè si eliminano le interferenze dovute allo strato atmosferico e forniscono valori di riflettanza *Bottom of Atmosphere* (BOA) (The European Space Agency, 2021).

Per ciascuna area di studio sono state collezionate 93 immagini di Livello – 2A riferibili al periodo tra aprile 2017 e marzo 2020. Il pacchetto *sen2r* (Ranghetti *et al.*, 2020) permette di accedere all'archivio dei prodotti Sentinel-2 e di elaborare i dati una volta selezionata l'area di interesse, la finestra temporale e la percentuale di copertura nuvolosa massima accettabile all'interno della scena. Le immagini sono state ritagliate con gli *shapefile* corrispondenti ai limiti delle aree di studio, ossia i confini amministrativi delle ZSC. In un secondo momento è avvenuta l'identificazione dei pixel associabili a nuvole e relative ombre proiettate a terra, che determinano distorsioni nelle risposte spettrali delle superfici. Generalmente le nuvole risultano avere una bassa riflettanza nella regione dell'infrarosso termico e un'alta riflettanza nelle lunghezze d'onda più corte, specialmente nel blu. Infine, è necessario mascherare queste aree sostituendo al valore dei pixel corrispondenti il valore *no-data* (NA). Le parti dell'immagine contenenti nuvole o ombre vengono riconosciute e filtrate in base al valore del NTDCI (*Normalized Difference Thermal Cloud Index*), un particolare NDCI (*Normalized Difference Cloud Index*) (Marshak *et al.*, 2000), calcolato tramite la funzione *cloudMask* del pacchetto *RStoolBox* (Leutner *et al.*, 2019). Questa funzione restituirà un raster in cui ogni pixel ha il valore di NTDCI, calcolato sulle bande del blu e dello SWIR (*Short Wave Infra-Red*), che nel caso di Sentinel-2 sono la 2 e la 11. Successivamente viene generata la maschera filtrando i valori di NDTCI sotto una certa soglia, determinata caso per caso dall'analisi dei valori che ciascun'immagine produce. Solitamente si ottengono buoni risultati fissando la soglia tra -0.10 e 1.15 per le nuvole e tra 0.80 e 0.92 per le ombre. La maschera viene sovrapposta all'immagine di partenza e confrontata pixel per pixel, inserendo il NA nelle zone

dove sono state riconosciute le nuvole. Per farlo viene usata la funzione *mask* del pacchetto *Raster* (Hijmans, 2021).

Prima di procedere al calcolo degli indici è stata effettuata la co-registrazione delle immagini in modo da permetterne una corretta sovrapposizione e ottenere un perfetto allineamento degli strati informativi. La co-registrazione è una fase fondamentale del pre-processamento delle immagini per l'analisi ecologica, dal momento che uno sfasamento dei pixel comporta errori spaziali nelle informazioni contenute (Sundaresan, Varshney and Arora, 2007; Young *et al.*, 2017).

2.3.1 Indici di vegetazione

Gli indici di vegetazione sono le variabili che quantificano l'andamento fenologico delle superfici vegetate. Per ciascun'immagine sono stati forniti sette indici di vegetazione, di seguito riportati.

Normalized Difference Vegetation Index (NDVI)

L'NDVI sfrutta le bande del NIR (ρ_{850}) e del *red* (ρ_{670}) ed è comunemente usato nelle tecniche di telerilevamento per esaltare la componente della riflettanza derivante dalla vegetazione. Minimizza il rumore da ombreggiamento ed effetti topografici, tuttavia tende a saturare rapidamente in contesti a elevata quantità di biomassa (Huete *et al.*, 2002).

$$NDVI = \frac{(\rho_{850} - \rho_{670})}{(\rho_{850} + \rho_{670})}$$

Modified Chlorophyll Absorption in Reflectance Index (MCARI)

L'MCARI prende in considerazione la banda del *red-edge* (ρ_{700}), la prima banda del *red* (ρ_{670}) e la banda del *green* (ρ_{550}) ed è in grado di fornire dei valori di assorbimento della clorofilla senza che questi risultino essere influenzati dalle diverse condizioni di illuminazione, dalla riflettanza di sfondo del suolo o da altri oggetti non fotosintetici osservati (T Daughtry *et al.*, 1999).

$$MCARI = \frac{((\rho_{700} - \rho_{670}) - 0.2(\rho_{700} - \rho_{550}))\rho_{670}}{\rho_{710}}$$

Green Normalized Difference Vegetation Index (GNDVI)

Il GNDVI utilizza le bande del NIR (ρ_{850}) e del *green* (ρ_{570}) ed è un indice di vegetazione comunemente usato per determinare l'assorbimento di acqua e azoto nella chioma della pianta (Gitelson and Merzlyak, 1998).

$$\text{GNDVI} = \frac{(\rho_{850} - \rho_{570})}{(\rho_{850} + \rho_{570})}$$

Normalized Difference Green/Red Redness Index (RI)

Il RI usa le bande del *red* (ρ_{670}) e del *green* (ρ_{570}) ed è un indice di colorazione, cioè corregge le variazioni causate dal colore del suolo a bassi tassi di copertura vegetale (Bannari *et al.*, 1995).

$$\text{RI} = \frac{(\rho_{670} - \rho_{570})}{(\rho_{670} + \rho_{570})}$$

Normalized Difference Red-Edge (NDRE)

Il NDRE utilizza la banda del NIR (ρ_{850}) e del *red-edge* (ρ_{700}) ed è un indice simile all'NDVI con minore tendenza alla saturazione e buona sensibilità ad un range più ampio di concentrazione in clorofilla, azoto, superficie fogliare e biomassa (Sims and Gamon, 2002).

$$\text{NDRE} = \frac{(\rho_{850} - \rho_{700})}{(\rho_{850} + \rho_{700})}$$

Normalized Difference Moisture Index (NDMI)

Il NDMI prende in considerazione le bande del NIR (ρ_{850}) e del SWIR (ρ_{1600}) ed è utilizzato per monitorare i cambiamenti nel contenuto di acqua delle foglie rimuovendo le variazioni indotte dalla struttura interna e dal contenuto di sostanza secca (Gao *et al.*, 1996).

$$\text{NDMI} = \frac{(\rho_{850} - \rho_{1600})}{(\rho_{850} + \rho_{1600})}$$

Modified Normalized Difference Water Index (MNDWI)

Il MNDWI sfrutta le bande del *green* (ρ_{570}) e del SWIR (ρ_{1600}) ed è un indice adatto a estrarre informazioni sulla quantità d'acqua in contesti dominati da aree edificate, ma anche vegetazione e suolo nudo (Xue and Su, 2017).

$$\text{MNDWI} = \frac{(\rho_{570} - \rho_{1600})}{(\rho_{570} + \rho_{1600})}$$

2.4 Costruzione delle serie temporali

Le serie temporali si definiscono come un insieme di misurazioni, solitamente effettuate con cadenza temporale fissa, che esprimono il comportamento di un certo fenomeno rispetto al tempo. Con l'analisi delle serie temporali si possono individuare nel fenomeno componenti di ciclicità o stagionalità e predirne l'andamento futuro. Le immagini vengono impilate e convertite in dati numerici attraverso le funzioni *stack* e *rasterToPoints* del pacchetto *Raster*,

in modo da essere processate sottoforma di matrice. Nelle righe sono riportati i pixel mentre nelle colonne il riferimento all'immagine con la data di presa. L'informazione temporale viene estratta e convertita nel corrispettivo giorno dell'anno. I dati vengono ordinati cronologicamente e trasposti in modo da facilitare le successive fasi di aggregazione e *smoothing*. In funzione del metodo proposto, il dataset viene fuso (i) con un file di testo esterno o (ii) con un vettore contenenti i giorni dell'anno suddivisi per settimane e bisettimane. Questo passaggio è stato implementato usando la funzione *merge* del pacchetto *data.table* (Dowle, 2021) così da velocizzarne il processo computazionale.

```
i)   doy.week.beweek <- read.delim("/Desktop/tesi/doy_week_beweek.txt")
      dataset <- merge (doy.week.beweek, data.ordered)

ii)  dataset <- merge (data.frame(doy = 1:365), data.ordered)
```

A questo punto, la matrice è composta sulle colonne dai pixel ordinati cronologicamente sulle righe. Sono presenti fin ora solo i valori riferiti alle immagini effettivamente acquisite; perciò, distanziati temporalmente da periodi in cui manca l'informazione. Il passo successivo è operare la pulizia del dataset dai valori anomali.

2.4.1 Individuazione dei valori anomali

In questo lavoro vengono confrontate due diverse modalità di individuazione dei valori anomali puntuali: la funzione *tsclean* del pacchetto *forecast* (Hyndman *et al.*, 2021) e la funzione *boxplot* del pacchetto *graphics* (Paul Murrell, 2020). Entrambe rientrano nelle tecniche *model-based*, quindi un punto x al tempo t può essere dichiarato anomalo se la distanza dal suo valore atteso \hat{x}_t è maggiore di una soglia predefinita τ .

$$|x_t - \hat{x}_t| < \tau$$

Nello specifico, si parla di *estimation model-based*, in quanto, la stima del valore atteso si effettua prendono in considerazione osservazioni precedenti e successive i valori correnti.

2.4.1.1 Funzione tsclean

La funzione crea un oggetto di classe *ts*, cioè un vettore in cui i dati sono distanziati temporalmente. Nel caso delle matrici, ogni colonna viene considerata una serie temporale uni-variata e il processamento può avvenire una sola alla volta. La modalità di individuazione dei valori anomali è differente per serie temporali stagionali e non stagionali. Per le serie temporali non stagionali, che non identificano un pattern cronologico ricorrente, viene usato il *Friedman's super smoother* (Friedman, 1984). L'algoritmo è uno stimatore non parametrico

basato su regressione lineare locale a due fattori a blocchi randomizzati. In questa sede l'interesse è focalizzato sulla prima tipologia, dove si identifica una componente stagionale significativa nella variazione del fenomeno. Nello specifico, la funzione utilizza un metodo di decomposizione della serie temporale: *Seasonal and Trend decomposition using LOESS* (STL). Il metodo STL è stato sviluppato da Cleveland et al. (1990), utilizza dei modelli di regressione adattati alla posizione per decomporre una serie temporale in componenti di tendenza, stagionali e residuali. È possibile applicare l'STL a qualsiasi dataset, ma per ottenere dei risultati significativi, è necessario un modello temporale ricorrente (stagionalità). L'algoritmo STL esegue uno smussamento della serie temporale utilizzando il metodo LOESS (*locally estimated scatterplot smoothing*) in due cicli: uno interno e uno esterno. Durante il ciclo interno, vengono calcolate le componenti stagionali e di tendenza. Il residuale si trova sottraendo queste ultime dalla serie temporale.

Poiché il dataset si presenta in forma di matrice, è necessario impostare la funzione con il comando *function* e applicarla a ciascuna colonna attraverso la funzione *sapply* (pacchetto *base*). Il codice viene scritto come segue.

```
clean.function <- function(x) {tsclean(x, replace.missing = FALSE, lambda = NULL)}  
clean.dataset <- sapply(dataset, FUN = clean.function)
```

Per ogni serie temporale vengono individuati i valori anomali e sostituiti da valori NA. Il parametro `replace.missing = FALSE` serve ad evitare che questi vengano rimpiazzati per interpolazione.

2.4.1.2 Funzione boxplot

In statistica il box plot o diagramma a scatola e baffi è una rappresentazione grafica utilizzata per descrivere la distribuzione di un campione tramite semplici indici di dispersione e di posizione. Il box plot è ottenuto a partire dai cinque numeri di sintesi, cioè minimo, primo quartile, mediana, terzo quartile, massimo, e descrive le caratteristiche salienti della distribuzione. Si ottiene riportando su un asse verticale (oppure orizzontale) i cinque numeri di sintesi. La scatola del box plot ha come estremi inferiore e superiore rispettivamente primo quartile e terzo quartile. La mediana divide la scatola in due parti. I baffi si ottengono congiungendo primo quartile al minimo e terzo quartile al massimo. L'altezza della scatola è pari al range interquartile e contiene il 50% centrale delle osservazioni effettuate, quelle comprese tra il primo ed il terzo quartile. Nell'ambito dei metodi di individuazione dei valori anomali, il box plot rientra all'interno delle tecniche che sfruttano la statistica di base. I valori

anomali sono rappresentati come dei punti isolati posizionati al di sopra e/o al di sotto dei baffi della distribuzione. Il posizionamento dei baffi è generalmente calcolato a 1,5 volte il range interquartile, dagli estremi della scatola, cioè primo e terzo quartile.

La funzione *boxplot* viene usata all'interno di un ciclo *for* e ogni colonna è considerata una serie temporale uni-variata di un pixel. Nei linguaggi di programmazione il ciclo *for* è una struttura di controllo iterativa che determina l'esecuzione di una porzione di programma ripetuta per un certo numero noto di volte. Questo permette di settare la funzione affinché il calcolo venga effettuato ciclicamente su ogni colonna del dataset. È necessario impostare un vettore che contenga la sequenza di lunghezza desiderata attraverso il comando *seq* (pacchetto *base*). Si definisce la struttura del ciclo inserendo al suo interno le funzioni da applicare. La funzione *boxplot* viene settata come segue. Il vettore `month` contiene i mesi dell'anno ed è generato in funzione del metodo proposto, attraverso un file di testo esterno o con la funzione *recode* (pacchetto *car*) dalla colonna dei giorni dell'anno. Il parametro `range = 1.5` esprime l'ampiezza dei baffi; quindi, le soglie oltre le quali i valori vengono considerati anomali. Con `$out` i valori anomali vengono estratti e salvati come vettore `outlier`. Infine, l'operatore `%in%` trova la posizione dei valori anomali all'interno della serie temporale così da permetterne la sostituzione con valori NA.

```
c <- seq (from = 2, to = ncol(dataset))
for (item in c) {
  outlier <- boxplot(dataset[,item] ~ month, range = 1.5, plot = FALSE)$out
  dataset[,item][dataset[,item] %in% outlier] <- NA
}
clean.dataset <- dataset
```

2.4.2 Aggregazione e Smoothing

In questa sezione, vengono presentate le fasi di aggregazione e *smoothing* delle serie temporali, nonché i metodi di applicazione proposti. Come anticipato nella parte introduttiva, l'aggregazione del dataset è funzionale al riempimento delle lacune temporali, alla riduzione del rumore nei dati e alla costruzione dei prodotti compositi (immagini che riassumono l'andamento fenologico di periodi differenti). La frequenza di acquisizione delle immagini definisce l'entità dell'aggregazione. Per i prodotti Sentinel-2 si reputa adeguato un tempo di aggregazione settimanale; perciò, le serie temporali saranno costituite da 52 valori. La fase di aggregazione consiste nel mediare i valori dei pixel all'interno di intervalli temporali definiti. Il codice è stato implementato affinché il processo avvenga nel minore tempo possibile. Si utilizzano le regole di sintassi del pacchetto *data.table*, in quanto efficiente in termini di

programmazione e calcolo. La forma generale della sintassi *data.table* è `DT[i, j, by]`, dove `DT` è un oggetto di classe *data.table* (elenco di vettori), `i` esprime quali righe prendere in considerazione, `j` definisce la funzione da applicare (in questo caso la media) e `by` per quale colonna aggregare i risultati. Nella sezione sottostante è riportato il codice e le stringhe usate per l'aggregazione. Come definito nella sezione precedente, il dato di input sono le serie temporali pulite dai valori animali. Vengono aggregate le colonne temporali secondo giorni, settimane o bi-settimane, al fine di creare un *data.table* esterno al ciclo al quale si agganceranno i risultati generati da ogni iterazione. A questo punto si imposta la sequenza di iterazione e si lancia l'aggregazione. La funzione *cbind* è necessaria al termine di ogni blocco di aggregazione per trasporre i risultati al di fuori del ciclo. Infine, l'ultima porzione di codice ripropone lo stesso calcolo per l'ultimo gruppo di pixel da processare.

```
dati <- data.table(clean.dataset)
aggr <- dati[,lapply(.SD, mean, na.rm = TRUE),
               .SDcols = c(colnames(dati[,2:3])),
               by = t.aggr]

dati.aggr <- data.frame(aggr)
c <- seq(from = 4,to = ncol(dati), by = 500)
for (item in c[-length(c)]) {
  aggr <- dati[,lapply(.SD, mean, na.rm=TRUE),
                 .SDcols = c(colnames(dati[,item:(item+500-1)])),
                 by = t.aggr]
  dati.aggr <- cbind(dati.aggr, aggr[,2:ncol(aggr)]) }

aggr <- dati[,lapply(.SD, mean, na.rm = TRUE),
               .SDcols = c(colnames(dati[, (ncol(b)+1):ncol(dati)])),
               by = t.aggr]
dati.aggr <- cbind(dati.aggr, aggr[,2:ncol(aggr)])
```

L'algoritmo di *smoothing* è un *Generalized Additive Model* (GAM) basato su *Cyclic Cubic Spline Regression*. I GAM sono estensioni dei modelli lineari in cui il predittore è dato dalla somma di funzioni regolari più un componente parametrico convenzionale (Wood, 2017). Sono una via di mezzo tra modelli semplici come la regressione lineare e modelli più complessi come l'aumento del gradiente. I modelli lineari, sebbene soffrano di un elevato *bias*, sono di facile interpretazione e consentono di comprendere la relazione lineare tra predittore e risposta. D'altra parte, i modelli di *machine learning* possono effettuare accurate previsioni tra variabili complesse, tuttavia tendono a richiedere molti dati e sono difficili da interpretare. I GAM risolvono questo problema permettendo di configurare relazioni non lineari complesse e

statistiche inferenziali, nonché comprendere e spiegare la struttura sottostante del modello (Adelchi and SCarpa, 2004). La *Cubic Spline Regression*, detta anche *b-spline*, è una forma di *Generalized Linear Model* ampiamente usata nella modellazione dei dati non lineari. Una *spline* è una funzione costituita da più polinomi raccordati tra loro, il cui scopo è interpolare un insieme di punti in intervalli definiti, chiamati nodi. Separando il dataset in sottoinsiemi, si ottiene uno *smoothing* localizzato e vengono risolti i problemi di *overfitting*, dati dall'influenza globale di ogni punto sull'adattamento.

```
tempi <- #tempo di interpolazione
f1 <- function(x) { return(
  predict(
    gam(as.vector(x) ~ s(tempi, bs = c("cc"),
      k = -1),
    na.action = na.exclude),
    data.frame(tempi = seq(1, max(tempi), 1),
      se = FALSE)))
}

dati.tempi.fit <- apply(dati.aggr, MARGIN = 2, f1)
```

In termini di codice, viene definito preliminarmente un vettore `tempi` che contenga l'intervallo temporale prescelto. L'interpolazione è effettuata dalla funzione `gam`, del pacchetto `mgcv` (Wood, 2017). Il parametro `bs = c("cc")` sta per *cyclic cubic*, una modalità di *b-spline* che prevede il raccordo del modello tra il primo e l'ultimo nodo. Questo è necessario a tenere conto della temporalità delle curve, che per loro natura descrivono un andamento continuo tra i mesi di dicembre e gennaio. Il termine `k = -1` serve a impostare il numero di nodi in cui suddividere il dataset secondo un processo automatico di validazione incrociata. La funzione viene applicata con il comando `apply` (pacchetto `base`) che esegue il calcolo su ogni colonna del dataset (`MARGIN = 2`).

2.4.2.1 GAM sui dati aggregati

Il metodo prevede l'applicazione dell'algoritmo di *smoothing* sui dati precedentemente aggregati. Le serie temporali prodotte possono essere direttamente classificate o sottoposte ad analisi funzionale, con un tempo di calcolo contenuto. Viene effettuata una prima aggregazione per giorni dell'anno, al fine di compensare i valori doppi di anni differenti; in un secondo momento si aggrega per settimane.

Sebbene la robustezza del metodo garantisca risultati in termini di accuratezza (Pesaresi *et al.*, 2020), il limite principale è dato dalla capacità del codice di adattarsi a contesti e dati input differenti. Si pensi ad esempio alla differenza di risoluzione temporale tra immagini Sentinel-

2 e Landsat 8. Le prime, avendo una frequenza di acquisizione di 5 giorni, permettono di aggregare per settimane, mentre le seconde, con 16 giorni, richiedono un'aggregazione bisettimanale. Il codice è simile per entrambi i metodi di pulizia proposti. La differenza principale è nel file di testo esterno contenente l'informazione temporale, che nel caso del *boxplot* riporta, oltre giorni dell'anno, settimane e bi-settimane, i mesi.

2.4.2.2 GAM sui giorni dell'anno

Il metodo prevede l'applicazione dell'algoritmo di *smoothing* sui dati ripuliti, non ancora sottoposti ad aggregazione. La compensazione dei giorni doppi viene effettuata direttamente dalla GAM, in modo da ottenere delle serie temporali lisce a 365 valori. A questo punto, è possibile classificare direttamente le serie temporali oppure aggregarle in funzione delle necessità. In questo modo il metodo risulta adattabile a diversi contesti e obiettivi.

Il limite maggiore riguarda il tempo di computazione, che aumenta in funzione di una più grande quantità di dati da elaborare. Sebbene la modalità di aggregazione sia la medesima, è stato predisposto un codice che in funzione dell'esigenza prepara il dataset per aggregare a settimane, bi-settimane o mesi. Di seguito viene riportato.

```
t.aggr <- #uno tra WEEK, BEWEEK e MONTH

if (t.aggr == "WEEK") {
  dati.tempi.fit[, "tempi"] <- c(rep(1:52, each = 7), 52)

} else if (t.aggr == "BEWEEK") {
  dati.tempi.fit[, "tempi"] <- c(rep(1:26, each = 14), 26)

} else if (t.aggr == "MONTH") {
  dati.tempi.fit[1:31, "tempi"] <- 1
  dati.tempi.fit[32:59, "tempi"] <- 2
  dati.tempi.fit[60:90, "tempi"] <- 3
  dati.tempi.fit[91:120, "tempi"] <- 4
  dati.tempi.fit[121:151, "tempi"] <- 5
  dati.tempi.fit[152:181, "tempi"] <- 6
  dati.tempi.fit[182:212, "tempi"] <- 7
  dati.tempi.fit[213:243, "tempi"] <- 8
  dati.tempi.fit[244:273, "tempi"] <- 9
  dati.tempi.fit[274:304, "tempi"] <- 10
  dati.tempi.fit[305:334, "tempi"] <- 11
  dati.tempi.fit[335:365, "tempi"] <- 12
}
```

L'elemento `t.aggr` definisce il tempo di aggregazione ed è impostato a monte del codice affinché sia l'unico parametro variabile. La trasposizione è effettuata con il comando *if* che verifica quale condizione logica si avvera (`TRUE`) e quale no (`FALSE`). Si imposta `t.aggr` uguale ai diversi tempi di aggregazione desiderati. Nel momento in cui viene trovata una corrispondenza si mettono in funzione le stringhe contenute nel blocco o viceversa si passa a quello successivo, e così via. Ogni blocco trasforma la colonna temporale dei giorni dell'anno (`dati.tempi.fit[, "tempi"]`) in settimane, bi-settimane o mesi. È importante che il vettore temporale abbia 365 valori poiché la conversione sfrutta la posizione del giorno all'interno dell'anno. Anche in questo caso il codice è simile per entrambi i metodi di pulizia proposti.

2.5 Analisi funzionale

Conclusesi le fasi di pulizia, *smoothing* e aggregazione, le serie temporali vengono salvate come file di testo con la funzione `write.table` (pacchetto *utils*). Il nome riporta il tipo di indice F03 (formula NDVI), le bande 08_04_00_00 (NIR e *red*), il metodo M01 (GAM sui dati aggregati, pulizia con funzione *tsclean*), la risoluzione delle immagini 10 (Sentinel-2) e la categoria TS (serie temporali). In modo analogo, le coordinate immagine dei pixel vengono salvate con la dicitura `coord` seguita dall'area di studio.

```
write.table(dati.tempi.fit, "F03_08_04_00_00_M01_10_TS.txt", row.names = FALSE)
write.table(dati.coord, "coord_Conero.txt", row.names = FALSE)
ts <- fread(#directory delle serie temporali)
coord <- fread(#directory delle coordinate dell'area di studio)
```

In un secondo momento, le serie temporali vengono ricaricate con la funzione `fread` (pacchetto *data.table*) e sottoposte ad analisi funzionale: *Functional Principal Component Analysis* (FPCA). La FPCA è un metodo statistico per l'analisi di variazione dei dati funzionali. Rientra all'interno delle tecniche di *Functional Data Analysis* poiché a differenza della PCA standard (*Principal Component Analysis*), che elabora dati puntuali, le unità statistiche di base sono funzioni. Ciò si adatta bene all'analisi delle curve temporali, dove le singole osservazioni non sono indipendenti, bensì vincolate dall'ordine cronologico. Il metodo fornisce una valutazione della complessità di un dataset definendo il numero minimo di componenti necessarie a rappresentarne il contenuto, senza perdere informazione. Gli *scores* sono i coefficienti che quantificano le somiglianze tra le serie temporali: contengono informazioni su posizione, forma e variazione di ciascuna curva osservata nello spazio. La maggior parte dell'informazione, espressa in termini di varianza, viene memorizzata nelle

prime componenti. In questo senso la FPCA può essere vista come uno strumento per la compressione e la riduzione di un set di dati (Nicol, 2016). I principali output dell’FPCA oltre gli *scores*, sono gli *eigenvalues* (o autovalori) e le *eigenfunctions* (o autofunzioni). Gli *eigenvalues* rappresentano per ogni serie temporale la variazione spiegata da ogni componente. La loro somma è uguale alla varianza totale dei dati. Le *eigenfunctions* stimano il valore degli *scores* descrivendo le maggiori variazioni funzionali per ogni componente (Hurley *et al.*, 2014). Per mezzo di queste è possibile ricostruire il dato originale (Dt) come sommatoria del prodotto di ogni componente (s) per la sua autofunzione (λ), più il valore medio (μ):

$$Dt = \mu + \sum s \times \lambda$$

L’analisi funzionale viene applicata direttamente con la funzione *FPCA* contenuta nel pacchetto *fdapace* (Wang, Chiou and Müller, 2016). Il procedimento richiede la trasformazione del dataset per mezzo della funzione *MakeFPCAInputs*, che genera una lista contenente le informazioni necessarie alla FPCA. Per ogni pixel, nello specifico: il vettore `dati` contiene i valori dell’indice, il vettore `time` contiene la settimana dell’anno e il vettore `ID` contiene il progressivo. È necessario che l’ordine con cui questi elementi sono inseriti nella matrice `per.fpca` rimanga invariato. Nella funzione i termini `ts.f$Ly` e `ts.f$Lt` fanno riferimento al valore dei pixel e alle settimane corrispondenti. Il parametro `methodXi` si riferisce al metodo usato per stimare gli *scores*, in questo caso `IN` si riferisce a *numeric integration* e viene usato per dati sufficientemente densi e privi di valori mancanti. Il termine `userBwCov = 1` definisce il valore di ampiezza della banda associata alla funzione di covarianza; un’ulteriore operazione di *smoothing* del dato. La funzione genera una lista contenente gli output dell’analisi: la funzione media (`fpcaObjTS$mu`), gli *eigenvalues* (`fpcaObjTS$lambda`), le *eigenfunctions* (`fpcaObjTS$phi`) e gli *scores* (`fpcaObjTS$xiEst`).

```
dati <- as.vector(as.matrix((ts)))
time <- as.numeric(rep(1:tp1, tp2))
ID <- rep(1:tp2, each = tp1)
per.fpca <- as.data.frame(cbind(ID, dati, time))

ts.f <- MakeFPCAInputs(per.fpca$ID, per.fpca$time, per.fpca$dati)
fpcaObjTS <- FPCA (ts.f$Ly,
                  ts.f$Lt,
                  list( methodXi = 'IN',
                        methodMuCovEst = 'smooth',
                        userBwCov = 1,
                        kernel = 'gauss',
                        FVEthreshold = 0.999))
```

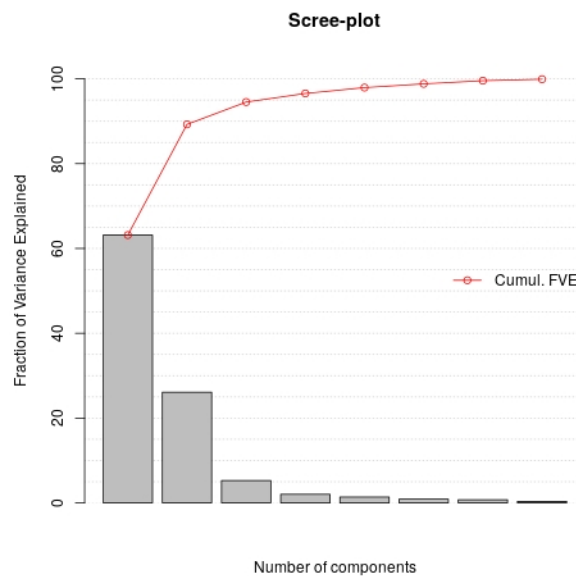


Figura 2-8: Esempio di scree-plot degli eigenvalues per ogni componente principale estratta dalla FPCA. Nelle ascisse il numero di componenti e nelle ordinate la frazione di varianza spiegata, assoluta e cumulativa

Per l'estrazione di un ridotto numero di *scores* si considerano le componenti che spiegano fino al 99% della varianza totale, quindi, quelle in cui il valore di `lambda` (Figura 2-8) è inferiore 0.99, come segue. Il vettore `rid` contiene il numero massimo di componenti da prendere ed è calcolato attraverso la sintassi di R.

```
FPCA.scores <- fpcaObjTS$xiEst
rid <- length(round(cumsum(fpcaObjTS$lambda/sum(fpcaObjTS$lambda))*100,3)
               [round(cumsum(fpcaObjTS$lambda/sum(fpcaObjTS$lambda))*100,3)<99])
FPCA.scores.rid <- fpcaObjTS$xiEst[,1:rid]
```

La ricostruzione delle serie temporali avviene sommando per ogni pixel il prodotto delle componenti per le autofunzioni, più il valore medio. Il calcolo è effettuato con un ciclo *for*, che itera dei blocchi condizionali *if else* sulla base del ridotto numero di componenti estratte (vettore `rid`). Gli *scores*, completi e ridotti, e le serie temporali ricostruite vengono salvati come file di testo (funzione *write.table*) nella cartella contenente le serie temporali originali. A questo punto, il totale dei predittori di classificazione è raggruppato per ogni indice. La matrice degli *scores* è costituita da un inferiore numero di colonne (settimane), poiché il dato temporale viene riassunto dalle componenti. Il numero di pixels, cioè le righe, si equivalgono.

2.6 Classificazione

Per entrambe le aree di studio, la classificazione della vegetazione è stata effettuata addestrando un algoritmo di classificazione supervisionata. L'apprendimento supervisionato permette di istruire un sistema informatico ad effettuare predizioni sulla base dei dati forniti dall'operatore (Benson *et al.*, 2019). L'addestramento di un algoritmo avviene attraverso i *training data*, ossia dati noti verificati per osservazione diretta e/o indiretta. Partendo da questi il sistema costruisce un modello applicabile alle altre istanze di classe incognita. In questo lavoro, per ogni area di studio vengono costruiti 112 modelli, derivati da 4 predittori, 4 metodi e 7 indici. In aggiunta, usando simultaneamente tutti i predittori derivanti dagli indici verranno elaborati ulteriori 32 modelli.

2.6.1.1 Training data

Per effettuare una classificazione supervisionata della vegetazione, è necessario costituire i *training data*, cioè un set di dati di riferimento delle classi target da mappare (le tipologie vegetazionali), omogeneamente distribuito nell'intera area di studio. Sebbene l'acquisizione e il pre-processamento dei *training data* non rientrino tra le attività e gli obiettivi di questo lavoro, se ne riporta una sintesi. In entrambe le aree, le analisi di campo hanno permesso di costruire una matrice fitosociologica dei dati di abbondanza delle specie (Quattrini *et al.*, 2020). Nella Gola di Frasassi le comunità vegetali sono state discriminate mediante partizionamento successivo delle matrici fitosociologiche con *cluster analysis* gerarchico (metodo *Ward*) e fuzzy (indice PEN). In questo modo è stato possibile raggruppare i rilievi fitosociologici in base alla similarità fisionomica e floristica. Nel Monte Conero il procedimento è stato semplificato in un solo processo di *clustering* delle matrici. Il dataset finale di riferimento è stato ottenuto combinando i risultati della *cluster analysis* dei rilievi fitosociologici, con ispezioni speditive in campo e per interpretazione visiva di immagini aeree ad alta risoluzione (classi target non rilevate direttamente). Per ciascuna area di studio, tutte le informazioni relative ai *training data* sono riassunte in un file vettoriale puntuale. Questo contiene nei campi le coordinate di ogni punto, la classe di appartenenza e la fonte di rilevamento. I predittori vengono caricati e spazializzati per mezzo delle coordinate dei pixel con la funzione *rasterFromXYZ* (pacchetto *raster*). I *training data* servono ad estrarre dalle immagini composite i pixel per l'addestramento dell'algoritmo di classificazione. La funzione di riferimento è *extract* (pacchetto *raster*). È necessario impostare lo stesso sistema di riferimento tra i training data e le immagini. Con il parametro `method = 'bilinear'` i valori restituiti sono interpolati dai valori delle quattro celle *raster* più vicine. Infine, i pixel vengono

organizzati in un *data.frame* al quale agganciare il campo classi (`td$CLASS`) dei *training data* di lunghezza equivalente.

```
file.coord <- #directory delle coordinate dell'area di studio
coord <- fread(file.coord)
file.shape <- #directory dei training data
td <- shapefile(file.shape)
file.pred <- #directory del predittore
pred <- as.data.frame(t(fread(file.pred)))

pred.coord <- cbind(coord, pred)
pred.coord.r <- rasterFromXYZ(pred.coord)
crs(pred.coord.r) <- crs(td)

pred.val <- extract(pred.coord.r, td, method = 'bilinear')
dataset <- as.data.frame(pred.val)
dataset$CLASS <- td$CLASS
```

2.6.1.2 Classi descritte

Nell'area di studio Monte Conero, sulla base dei dati vettoriali forniti (Figura 2-9), le categorie vegetazionali descritte sono 4 distribuite in 175 punti: boschi termofili a prevalenza di *Quercus ilex* (associazione *Cyclamino hederifolii* – *Quercetum ilicis*); boschi mesofili a prevalenza di *Quercus ilex* (associazione *Cephalanthero longifoliae* – *Quercetum ilicis* subassociazione *ruscetosum hypoglossi*); boschi a prevalenza di *Ostrya carpinifolia* (associazione *Asparago acutifolii* – *Ostryetum carpinifoliae*); rimboschimenti.

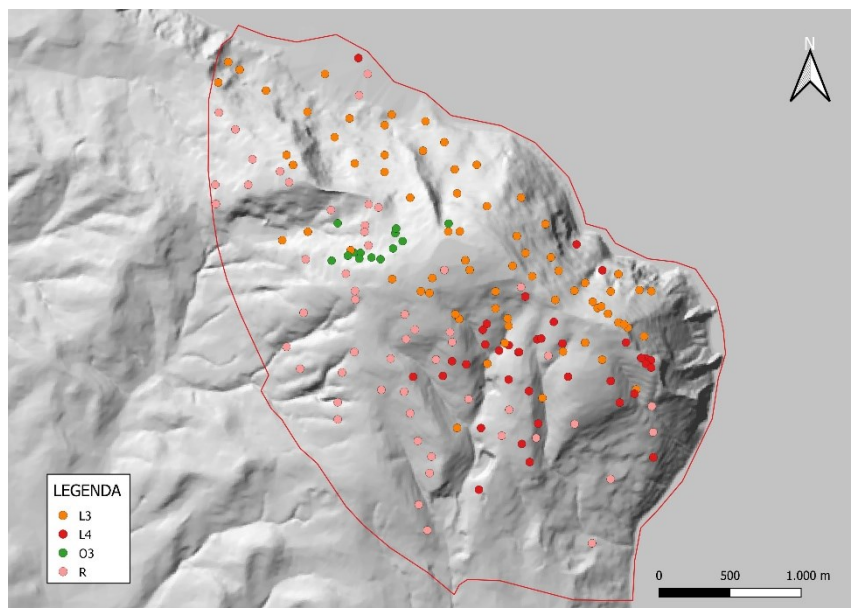


Figura 2-9: Distribuzione spaziale dei training data per l'area di studio Monte Conero. Categorie: L3 – boschi mesofili a prevalenza di *Quercus ilex* (*Cephaanthero longifoliae* – *Quercetum ilicis*); L4 – boschi termofili a prevalenza di *Quercus ilex* (*Cyclamino hederifolii* – *Quercetum ilicis*); O3 – boschi a prevalenza di *Ostrya carpinifolia* (*Asparago acutifolii* – *Ostryetum carpinifoliae*); R – rimboschimenti.

Nell'area di studio Gola di Frasassi, sulla base dei dati vettoriali forniti (Figura 2-10), le categorie vegetazionali descritte sono 8 distribuite in 242 punti: ginestre (associazione *Spartio juncei* – *Cytisetum sessilifolii* variante a *Spartium junceum*); ginepre (associazione *Spartio juncei* – *Cytisetum sessilifolii* variante a *Juniperus oxycedrus* e *Juniperus communis*); boschi a prevalenza di *Quercus ilex* (associazione *Cephalanthero longifoliae* – *Quercetum ilicis*); mosaico di gariga e macchia; boschi a prevalenza di *Ostrya carpinifolia* (associazione *Scutellario columnae*-*Ostryetum carpinifoliae*), praterie (associazione *Asperulo purpureae* – *Brometum erecti*); boschi a prevalenza di *Quercus pubescens* (associazione *Cytiso sessilifolii* – *Quercetum pubescentis*); rimboschimenti.

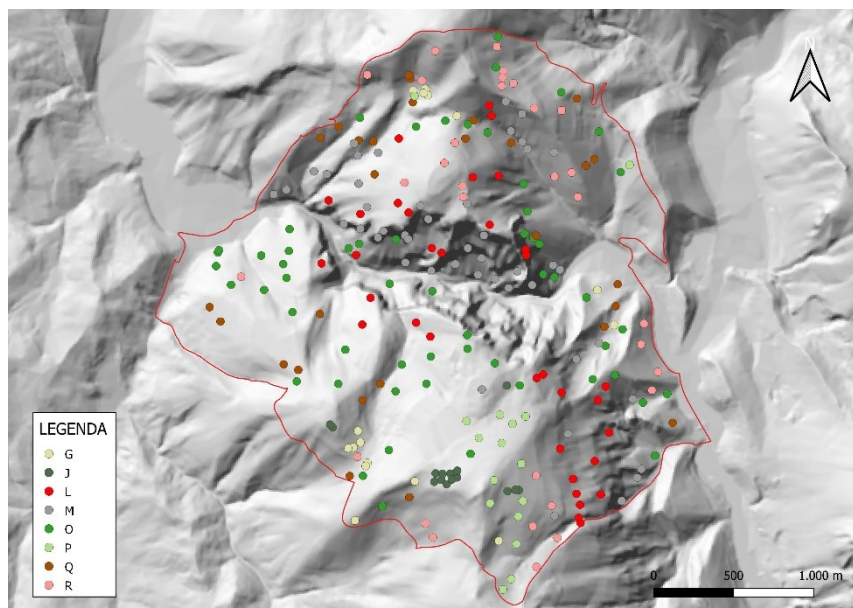


Figura 2-10: Distribuzione spaziale dei training data per l'area di studio Gola di Frasassi.
Categorie: *G* – ginestreti (*Spartio juncei* – *Cytisetum sessilifolii* variante a *Spartium junceum*); *J* – ginepreti (*Spartio juncei* – *Cytisetum sessilifolii* variante a *Juniperus oxycedrus* e *Juniperus communis*); *L* – boschi a prevalenza di *Quercus ilex* (*Cephalanthero longifoliae* – *Quercetum ilicis*); *M* – mosaico a gariga e macchia; *O* – boschi a prevalenza di *Ostrya carpinifolia* (*Scutellario columnae* – *Ostryetum carpinifoliae*); *P* – praterie (*Asperulo purpureae* – *Brometum erecti*); *Q* – boschi a prevalenza di *Quercus pubescens* (*Cytiso sessilifolii* – *Quercetum pubescentis*); *R* – rimboschimenti.

2.6.1.3 Random Forest

Il RF è un classificatore basato su alberi decisionali non correlati, in grado di aggregare i risultati di ognuno, incrementando l'accuratezza generale della classificazione. Questa procedura, definita *Bootstrap Aggregation* (o *bagging*), è una tecnica che combina le previsioni di diversi algoritmi di apprendimento automatico per fare previsioni più accurate rispetto a qualsiasi singolo modello. Individualmente, le previsioni fatte dagli alberi decisionali potrebbero non essere accurate, ma combinate insieme, saranno in media più vicine al risultato. Gli alberi decisionali vengono creati da sottoinsiemi casuali dei *training data*, chiamati *in-bag samples*. I dati che non vengono presi in considerazione nella costruzione di uno specifico albero vengono definiti *out-of-bag samples*. Tramite questi, è possibile stimare il tasso di errore del classificatore e convalidare la previsione del modello.

L'accuratezza generale della classificazione (*overall accuracy*) è definita dal rapporto tra i pixel classificati correttamente dall'algoritmo e il totale dei pixel con i quali il modello viene addestrato. Graficamente è rappresentabile da una matrice, chiamata matrice di confusione,

nella quale le colonne riportano i valori predetti e le righe i valori reali. Il numero di pixel classificati correttamente dall'algoritmo è distribuito sulla diagonale. Due importanti parametri da valutare sono la *producer's accuracy* e la *user's accuracy*. La *producer's accuracy* si ottiene dal rapporto tra il numero dei pixel classificati correttamente in una categoria e il numero dei punti rilevati a terra per quella categoria. La *user's accuracy* si ottiene dal rapporto tra il numero di pixel classificati correttamente in una categoria e il numero totale dei pixel classificati in quella determinata categoria (Congalton, 2001). Il RF classifica i predittori, in accordo alla loro capacità di discriminare le classi obiettivo. L'importanza dei predittori viene valutata sulla base dell'indice normalizzato *Mean Decrease Accuracy*. Più alto sarà il valore dell'indice, maggiore sarà l'importanza del predittore.

Generalmente, la creazione di un modello prevede la divisione del dataset in due parti: il *training set*, necessario all'addestramento dell'algoritmo, e il *test set*, di verifica dell'accuratezza di classificazione. In presenza di una buona numerosità del campione osservato è possibile utilizzare la *cross validation*, una tecnica statistica che permette di sfruttare l'intero dataset come *training* e *test*. Il dataset viene diviso in k cartelle, ognuna delle quali, in modo iterativo, è utilizzata per testare l'algoritmo. In questo modo, il modello si costruisce sfruttando la totalità dei dati per il *training* e una porzione sempre diversa per il *test* (Leonel Jorge, 2018). In questo lavoro, viene utilizzata la *repeated cross validation*, che permette di ripetere ciclicamente il processo un certo numero di volte per aumentare l'accuratezza della classificazione.

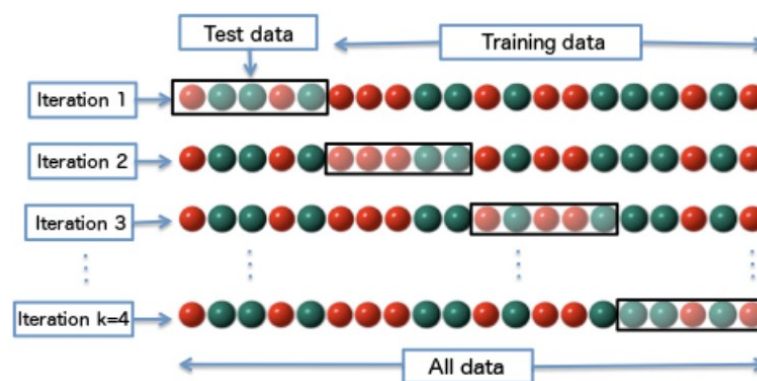


Figura 2-11: Schema semplificato di funzionamento della cross validation.

La funzione *train* viene settata per addestrare l'algoritmo di classificazione. Il vettore `CLASS` contiene le classi da predire associate alla matrice dei *training data*. Il metodo prevede l'uso dell'algoritmo di classificazione *Random Forest* (`method = 'rf'`). Il termine `ntree` indica il numero di alberi decisionali per effettuare la predizione su ogni singola istanza. Il

parametro `tuneGrid` contiene il vettore `mtry` che definisce il numero di *hyperparametri*, cioè le variabili scelte casualmente ad ogni *split* degli alberi decisionali. Con `metric = 'Accuracy'` si setta il criterio di scelta del modello finale, in questo caso il riferimento è l'accuratezza. Infine, `trControl` descrive la modalità di addestramento dell'algoritmo di classificazione. Il metodo viene impostato con il parametro `method = "repeatedcv"` della funzione *trainControl*. I termini `number` e `repeats` indicano rispettivamente il numero di *k* cartelle in cui il dataset viene suddiviso e le ripetizioni necessarie. Tutte le funzioni usate in questa sezione appartengono al pacchetto *caret* (*Classification And REgression Training*), specifico per la costruzione dei modelli di classificazione (Khun, 2021).

```
ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 5)
mtry <- 1:sqrt(ncol(dataset))
tunegrid <- expand.grid(.mtry=mtry)
rfDownsampled <- train(as.factor(CLASS) ~ ., data = dataset,
                        method = "rf",
                        ntree = 1500,
                        tuneGrid = tunegrid,
                        metric = "Accuracy",
                        strata = as.factor(dataset$CLASS),
                        trControl = ctrl,
                        sampsize = campione)
```


Capitolo 3

RISULTATI

3.1 Variazioni stagionali descritte dall'analisi funzionale

In questa sezione vengono riportati i risultati dell'applicazione della FPCA sulle serie temporali settimanali nelle due aree di studio, per ogni indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI) e metodo (M01 GAM sui dati aggregati pulizia con *tsclean*, M02 GAM sui dati aggregati pulizia con *boxplot*, M03 GAM sui dati giornalieri, pulizia con *tsclean*, M04 GAM sui dati giornalieri, pulizia con *boxplot*).

3.1.1 Variazioni stagionali derivate dalle serie temporali RI

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il RI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 94.75% della variazione totale, (ii) 8 componenti dal metodo M02 di cui le prime 3 spiegano il 95.48% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 93.86% della variazione totale, (iiii) 8 componenti dal metodo M04 di cui le prime 3 spiegano il 95.28% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-1. Nelle Figure 5-1, 5-2, 5-3, 5-4 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-1: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
M01	66.405	23.343	5.004	2.102	1.376	0.652	0.627	0.388	0.104
M02	69.659	21.577	4.250	1.828	1.369	0.535	0.479	0.304	
M03	64.116	24.589	5.164	2.416	1.328	0.928	0.831	0.479	0.150
M04	69.786	22.153	3.344	2.072	1.385	0.485	0.460	0.316	

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il RI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano l'89.61% della variazione totale, (ii) 9 componenti dal metodo M02 di cui le prime 3 spiegano l'89.75%

della variazione totale, (iii) 10 componenti dal metodo M03 di cui le prime 3 spiegano l'89.46% della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il 90.40% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-2. Nelle Figure 5-5, 5-6, 5-7, 5-8 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-2: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9	10
M01	54.445	29.846	5.324	3.918	2.436	1.972	1.223	0.645	0.192	
M02	55.320	29.101	5.333	3.766	2.647	1.959	1.097	0.596	0.182	
M03	54.472	28.356	6.633	3.777	2.412	1.957	1.242	0.807	0.246	0.097
M04	56.348	28.409	5.643	3.780	2.078	1.691	1.152	0.689	0.210	

3.1.2 Variazioni stagionali derivate dalle serie temporali MCARI

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il MCARI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 91.80% della variazione totale, (ii) 9 componenti dal metodo M02 di cui le prime 3 spiegano il 96.01% della variazione totale, (iii) 10 componenti dal metodo M03 di cui le prime 3 spiegano l'89.81% della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il 94.50% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-3. Nelle Figure 5-9, 5-10, 5-11, 5-12 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-3: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9	10
M01	57.548	24.627	9.629	4.205	1.370	1.115	0.825	0.497	0.183	
M02	75.506	12.145	8.356	1.520	1.088	0.751	0.314	0.198	0.123	
M03	57.693	22.564	9.551	5.003	2.612	1.021	0.756	0.477	0.214	0.109
M04	73.263	14.624	6.611	2.121	1.895	0.726	0.374	0.248	0.139	

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il MCARI ha permesso di estrarre: (i) 10 componenti dal metodo M01 di cui le prime 3 spiegano il 91.28% della variazione totale, (ii) 9 componenti dal metodo M02 di cui le prime 3 spiegano il

il 94.40% della variazione totale, (iii) 11 componenti dal metodo M03 di cui le prime 3 spiegano l' 89.42% della variazione totale, (iiii) 10 componenti dal metodo M04 di cui le prime 3 spiegano il 95.42% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-4. Nelle Figure 5-13, 5-14, 5-15, 5-16 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-4: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9	10	11
M01	52.328	25.433	13.519	3.214	1.661	1.579	0.990	0.804	0.327	0.144	
M02	66.041	19.002	9.354	2.410	1.278	1.032	0.506	0.243	0.132		
M03	51.977	24.071	13.375	3.379	2.039	1.655	1.468	0.959	0.534	0.365	0.178
M04	63.805	25.113	6.505	1.549	1.291	0.682	0.511	0.278	0.163	0.103	

3.1.3 Variazioni stagionali derivate dalle serie temporali GNDVI

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il GNDVI ha permesso di estrarre: (i) 8 componenti dal metodo M01 di cui le prime 3 spiegano il 99.17% della variazione totale, (ii) 7 componenti dal metodo M02 di cui le prime 3 spiegano il 99.37% della variazione totale, (iii) 8 componenti dal metodo M03 di cui le prime 3 spiegano il 98.93% della variazione totale, (iiii) 7 componenti dal metodo M04 di cui le prime 3 spiegano il 99.25% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-5. Nelle Figure 5-17, 5-18, 5-19, 5-20 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-5: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8
M01	96.665	1.958	0.547	0.423	0.156	0.110	0.099	0.042
M02	97.091	1.848	0.433	0.362	0.122	0.094	0.051	
M03	96.100	2.047	0.791	0.489	0.223	0.165	0.116	0.070
M04	96.685	1.932	0.634	0.405	0.169	0.115	0.060	

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il GNDVI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 95.01% della variazione totale, (ii) 9 componenti dal metodo M02 di cui le prime 3 spiegano il 95.13% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 95.09%

della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il 95.51% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-6. Nelle Figure 5-21, 5-22, 5-23, 5-24 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-6: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
<i>M01</i>	67.742	20.639	6.631	1.578	1.311	0.945	0.725	0.320	0.109
<i>M02</i>	68.695	19.996	6.437	1.726	1.140	0.859	0.725	0.316	0.106
<i>M03</i>	68.565	20.167	6.359	1.597	1.266	0.914	0.757	0.263	0.112
<i>M04</i>	69.804	19.621	6.083	1.695	0.934	0.826	0.695	0.243	0.099

3.1.4 Variazioni stagionali derivate dalle serie temporali NDVI

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il NDVI ha permesso di estrarre: (i) 7 componenti dal metodo M01 di cui le prime 3 spiegano il 99.23% della variazione totale, (ii) 6 componenti dal metodo M02 di cui le prime 3 spiegano il 99.36% della variazione totale, (iii) 8 componenti dal metodo M03 di cui le prime 3 spiegano il 98.83% della variazione totale, (iiii) 7 componenti dal metodo M04 di cui le prime 3 spiegano il 99.18% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-7. Nelle Figure 5-25, 5-26, 5-27, 5-28 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-7: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8
<i>M01</i>	96.105	2.605	0.524	0.408	0.231	0.077	0.051	
<i>M02</i>	96.291	2.607	0.460	0.397	0.192	0.053		
<i>M03</i>	95.540	2.705	0.582	0.435	0.422	0.122	0.100	0.093
<i>M04</i>	95.977	2.734	0.471	0.414	0.281	0.067	0.055	

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il NDVI ha permesso di estrarre: (i) 8 componenti dal metodo M01 di cui le prime 3 spiegano il 94.61% della variazione totale, (ii) 8 componenti dal metodo M02 di cui le prime 3 spiegano il 94.81% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 94.60%

della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il 94.96% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-8. Nelle Figure 5-29, 5-30, 5-31, 5-32 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-8: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
M01	63.211	26.128	5.277	2.029	1.383	0.882	0.741	0.349	
M02	63.758	26.438	4.618	2.169	1.167	0.814	0.689	0.348	
M03	63.748	26.241	4.616	2.123	1.298	0.920	0.625	0.315	0.114
M04	64.547	26.621	3.795	2.198	1.088	0.815	0.538	0.303	0.094

3.1.5 Variazioni stagionali derivate dalle serie temporali NDRE

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il NDRE ha permesso di estrarre: (i) 8 componenti dal metodo M01 di cui le prime 3 spiegano il 98.49% della variazione totale, (ii) 8 componenti dal metodo M02 di cui le prime 3 spiegano il 98.58% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 98.07% della variazione totale, (iiii) 8 componenti dal metodo M04 di cui le prime 3 spiegano il 98.45% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-9. Nelle Figure 5-33, 5-34, 5-35, 5-36 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-9: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
M01	92.838	4.689	0.959	0.607	0.422	0.227	0.158	0.098	
M02	93.297	4.406	0.876	0.593	0.399	0.202	0.135	0.092	
M03	91.285	5.497	1.285	0.647	0.459	0.419	0.208	0.136	0.064
M04	92.687	4.635	1.127	0.618	0.415	0.276	0.150	0.092	

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il NDRE ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 93.60% della variazione totale, (ii) 9 componenti dal metodo M02 di cui le prime 3 spiegano il 93.94% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 93.87% della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il

94.29% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-10. Nelle Figure 5-37, 5-38, 5-39, 5-40 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-10: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
<i>M01</i>	59.587	27.562	6.449	1.939	1.703	1.241	0.984	0.417	0.118
<i>M02</i>	59.269	28.185	6.486	1.839	1.635	1.210	0.845	0.408	0.121
<i>M03</i>	59.668	28.514	5.692	1.948	1.758	1.058	0.827	0.408	0.128
<i>M04</i>	59.590	29.211	5.490	1.811	1.655	0.992	0.739	0.395	0.118

3.1.6 Variazioni stagionali derivate dalle serie temporali MNDWI

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il MNDWI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 96.77% della variazione totale, (ii) 8 componenti dal metodo M02 di cui le prime 3 spiegano il 97.58% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 95.78% della variazione totale, (iiii) 8 componenti dal metodo M04 di cui le prime 3 spiegano il 97.11% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-11. Nelle Figure 5-41, 5-42, 5-43, 5-44 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-11: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
<i>M01</i>	84.757	9.702	2.312	1.385	0.661	0.561	0.443	0.127	0.051
<i>M02</i>	86.338	9.081	2.161	1.097	0.514	0.457	0.260	0.091	
<i>M03</i>	81.593	11.606	2.587	1.567	1.009	0.800	0.525	0.202	0.111
<i>M04</i>	85.229	9.729	2.153	1.315	0.685	0.513	0.264	0.113	

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il MNDWI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 91.66% della variazione totale, (ii) 9 componenti dal metodo M02 di cui le prime 3 spiegano il 91.69% della variazione totale, (iii) 10 componenti dal metodo M03 di cui le prime 3 spiegano il 91.69% della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il 91.92% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in

Tabella 3-12. Nelle Figure 5-45, 5-46, 5-47, 5-48 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-12: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9	10
M01	72.237	15.985	3.439	2.534	2.143	1.928	1.073	0.485	0.177	
M02	72.038	16.398	3.251	2.632	2.223	1.868	1.044	0.378	0.167	
M03	72.379	15.876	3.438	2.553	1.964	1.821	1.029	0.620	0.239	0.080
M04	72.320	16.443	3.157	2.634	2.035	1.758	1.011	0.449	0.192	

3.1.7 Variazioni stagionali derivate dalle serie temporali NDMI

Per l'area di studio Monte Conero, la FPCA delle serie temporali costruite con il NDMI ha permesso di estrarre: (i) 8 componenti dal metodo M01 di cui le prime 3 spiegano il 97.91% della variazione totale, (ii) 8 componenti dal metodo M02 di cui le prime 3 spiegano il 98.04% della variazione totale, (iii) 8 componenti dal metodo M03 di cui le prime 3 spiegano il 97.70% della variazione totale, (iiii) 8 componenti dal metodo M04 di cui le prime 3 spiegano il 97.85% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in Tabella 3-13. Nelle Figure 5-49, 5-50, 5-51, 5-52 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-13: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8
M01	87.503	8.990	1.414	1.087	0.530	0.204	0.156	0.116
M02	87.612	9.027	1.399	0.997	0.530	0.181	0.142	0.113
M03	87.024	9.204	1.477	1.201	0.537	0.228	0.176	0.153
M04	87.307	9.146	1.395	1.163	0.514	0.200	0.149	0.124

Per l'area di studio Gola di Frasassi, la FPCA delle serie temporali costruite con il NDMI ha permesso di estrarre: (i) 9 componenti dal metodo M01 di cui le prime 3 spiegano il 94.07% della variazione totale, (ii) 8 componenti dal metodo M02 di cui le prime 3 spiegano il 94.22% della variazione totale, (iii) 9 componenti dal metodo M03 di cui le prime 3 spiegano il 94.22% della variazione totale, (iiii) 9 componenti dal metodo M04 di cui le prime 3 spiegano il 94.41% della variazione totale. Per ciascuna componente, gli *eigenvalues* sono riportati in

Tabella 3-14. Nelle Figure 5-53, 5-54, 5-55, 5-56 in Appendice vengono visualizzati i pattern spazio-temporali associati alle prime 3 componenti nei 4 metodi testati.

Tabella 3-14: Eigenvalues delle componenti principali estratte con la FPCA per ogni metodo. La variazione spiegata dalle componenti diminuisce progressivamente dalla prima all'ultima.

	1	2	3	4	5	6	7	8	9
<i>M01</i>	64.778	24.083	5.212	2.376	1.488	0.885	0.749	0.344	0.086
<i>M02</i>	65.014	24.034	5.174	2.362	1.509	0.875	0.705	0.327	
<i>M03</i>	64.763	24.316	5.139	2.364	1.416	0.825	0.693	0.365	0.121
<i>M04</i>	64.992	24.412	5.004	2.348	1.399	0.782	0.601	0.353	0.109

3.2 Mappatura della vegetazione

Nella parte seguente, vengono riportati i risultati dalla classificazione per le 4 tipologie di predittori utilizzate: serie temporali (TS), *scores* (SCT), *scores* ridotti al 99% di varianza spiegata (SCR) e serie temporali ricostruite (TSR). In totale, sono generati 256 modelli, 32 dei quali utilizzano simultaneamente i predittori derivati dagli indici, mentre i restanti 224 separatamente. Per ciascuna area di studio, i risultati vengono organizzati in tabelle a doppia entrata (metodo – predittore), in cui si restituiscono i valori percentuale di *overall accuracy* dei modelli. Data la mole di output, le mappe e le matrici di confusione associate vengono visualizzate in Appendice esclusivamente per i migliori modelli prodotti.

3.2.1 Mappatura dei predittori derivati dal RI

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dall'indice RI ha generato: (i) l'80% di accuratezza massima dal metodo M03 con le TS, (ii) l'81% di accuratezza massima dai metodi M01 e M03 con gli SCT, (iii) l'80% di accuratezza massima dai metodi M01 e M03 con gli SCR, (iiii) il 40% di accuratezza massima dal metodo M01 con le TSR. L'accuratezza massima globale è pari all'81% per i metodi M01 e M03 con l'uso degli SCT. L'accuratezza minima globale è pari al 35% per il metodo M04 con l'uso delle TSR. I modelli costruiti con i metodi M01 e M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratezze dei modelli viene riportato in Tabella 3-15.

Tabella 3-15: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	78	81	80	40
M02	75	77	77	37
M03	80	81	80	39
M04	75	76	76	35

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice RI ha generato: (i) il 73% di accuratezza massima dal metodo M03 con le TS, (ii) il 76% di accuratezza massima dal metodo M03 con gli SCT, (iii) il 77% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 38% di accuratezza massima dal metodo M04 con le TSR. L'accuratezza massima globale è pari al 77% per il metodo M03 con l'uso degli SCR. L'accuratezza minima globale è pari al 35% per i metodi M01 e M02 con l'uso delle TSR. I modelli costruiti con il metodo M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratezze dei modelli viene riportato in Tabella 3-16.

Tabella 3-16: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	72	73	73	35
M02	71	72	73	35
M03	73	76	77	36
M04	72	75	74	38

3.2.2 Mappatura dei predittori derivati dal MCARI

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dall'indice MCARI ha generato: (i) il 77% di accuratezza massima dal metodo M03 con le TS, (ii) l'80% di accuratezza massima dal metodo M03 con gli SCT, (iii) il 79% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 51% di accuratezza massima dal metodo M04 con le TSR. L'accuratezza massima globale è pari all'80% per il metodo M03 con l'uso degli SCT. L'accuratezza minima globale è pari al 40% per i metodi M01 e M03 con l'uso delle TSR. I modelli costruiti con il metodo M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratezze dei modelli viene riportato in Tabella 3-17.

Tabella 3-17: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	76	77	76	40
M02	71	71	72	46
M03	77	80	79	40
M04	74	76	75	51

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice MCARI ha generato: (i) il 64% di accuratezza massima dal metodo M03 con le TS, (ii) il 67% di accuratezza massima dal metodo M03 con gli SCT, (iii) il 67% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 36% di accuratezza massima dai metodi M02 e M04 con le TSR. L'accuratezza massima globale è pari al 67% per il metodo M03 con l'uso degli SCT e degli SCR. L'accuratezza minima globale è pari al 33% per il metodo M03 con l'uso delle TSR. I modelli costruiti con il metodo M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratezze dei modelli viene riportato in Tabella 3-18.

Tabella 3-18: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	62	66	65	35
M02	55	65	62	36
M03	64	67	67	33
M04	60	65	64	36

3.2.3 Mappatura dei predittori derivati del GNDVI

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dall'indice GNDVI ha generato: (i) il 78% di accuratezza massima dal metodo M04 con le TS, (ii) l'81% di accuratezza massima dal metodo M03 con gli SCT, (iii) l'82% di accuratezza massima dal metodo M04 con gli SCR, (iiii) il 58% di accuratezza massima dal metodo M04 con le TSR. L'accuratezza massima globale è pari all'82% per il metodo M04 con l'uso degli SCR. L'accuratezza minima globale è pari al 49% per il metodo M03 con l'uso delle TSR. I modelli costruiti con i metodi M03 e M04 e i predittori SCT e SCR si dimostrano i più performanti.

Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratezze dei modelli viene riportato in Tabella 3-19.

Tabella 3-19: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	75	79	78	52
M02	77	79	76	57
M03	76	81	81	49
M04	78	80	82	58

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice GNDVI ha generato: (i) il 67% di accuratezza massima dal metodo M01 con le TS, (ii) il 72% di accuratezza massima dal metodo M04 con gli SCT, (iii) il 70% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 40% di accuratezza massima dai metodi M01 e M04 con le TSR. L'accuratezza massima globale è pari al 72% per il metodo M04 con l'uso degli SCT. L'accuratezza minima globale è pari al 38% per il metodo M03 con l'uso delle TSR. I modelli costruiti con i metodi M03 e M04 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratezze dei modelli viene riportato in Tabella 3-20.

Tabella 3-20: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	67	69	68	40
M02	66	69	69	39
M03	66	71	70	38
M04	66	72	69	40

3.2.4 Mappatura dei predittori derivati dal NDVI

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice NDVI ha generato: (i) l'82% di accuratezza massima dal metodo M01 con le TS, (ii) il 79% di accuratezza massima dai metodi M01 e M03 con gli SCT, (iii) il 79% di accuratezza massima dai metodi M01 e M03 con gli SCR, (iiii) il 58% di accuratezza massima dal metodo M04 con le TSR. L'accuratezza massima globale è pari all'82% per il metodo M01 con l'uso delle TS. L'accuratezza minima globale è pari al 54% per i metodi M02 e M03 con l'uso delle TS. L'accuratezza minima globale è pari al 54% per i metodi M02 e M03 con l'uso delle TS.

TSR. I modelli costruiti con i metodi M01 e M03 e i predittori TS, SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. L'insieme delle accuratèzze dei modelli viene riportato in Tabella 3-21.

Tabella 3-21: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	82	79	79	55
M02	79	77	77	54
M03	81	79	79	54
M04	78	78	78	58

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice NDVI ha generato: (i) il 70% di accuratezza massima dai metodi M01 e M03 con le TS, (ii) il 72% di accuratezza massima dal metodo M01 con gli SCT, (iii) il 70% di accuratezza massima dal metodo M01 con gli SCR, (iiii) il 41% di accuratezza massima dal metodo M02 con le TSR. L'accuratezza massima globale è pari all'72% per il metodo M01 con l'uso degli SCT. L'accuratezza minima globale è pari al 38% per il metodo M01 con l'uso delle TSR. I modelli costruiti con i metodi M01 e M03 e i predittori TS, SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-22.

Tabella 3-22: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	70	72	70	38
M02	69	71	69	41
M03	70	71	69	40
M04	69	70	67	40

3.2.5 Mappatura dei predittori derivati dal NDRE

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dall'indice NDRE ha generato: (i) il 77% di accuratezza massima dal metodo M03 con le TS, (ii) l'81% di accuratezza massima dal metodo M03 con gli SCT, (iii) l'83% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 54% di accuratezza massima dal metodo M01 con le TSR. L'accuratezza massima globale è pari all'83% per il metodo M03 con l'uso degli SCR.

L'accuratezza minima globale è pari al 49% per il metodo M04 con l'uso delle TSR. I modelli costruiti con i metodi M01 e M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-23.

Tabella 3-23: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	76	80	80	54
M02	76	79	79	52
M03	77	81	83	52
M04	76	79	78	49

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice NDRE ha generato: (i) il 67% di accuratezza massima dal metodo M03 con le TS, (ii) il 68% di accuratezza massima dai metodi M03 e M04 con gli SCT, (iii) il 68% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 35% di accuratezza massima dal metodo M02 con le TSR. L'accuratezza massima globale è pari al 68% per i metodi M03 e M04 con l'uso degli SCT e degli SCR. L'accuratezza minima globale è pari al 32% per il metodo M01 con l'uso delle TSR. I modelli costruiti con il metodo M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-24.

Tabella 3-24: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	65	66	65	32
M02	63	65	63	35
M03	67	68	68	32
M04	66	68	64	33

3.2.6 Mappatura dei predittori derivati dal MNDWI

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dall'indice MNDWI ha generato: (i) il 69% di accuratezza massima dal metodo M03 con le TS, (ii) il 72% di accuratezza massima dal metodo M03 con gli SCT, (iii) il 72% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 46% di accuratezza massima dal metodo M02 con le

TSR. L'accuratezza massima globale è pari al 72% per il metodo M03 con l'uso degli SCT e degli SCR. L'accuratezza minima globale è pari al 41% per il metodo M04 con l'uso delle TSR. I modelli costruiti con il metodo M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-25.

Tabella 3-25: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	65	69	67	43
M02	63	68	68	46
M03	69	72	72	45
M04	65	69	69	41

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice MNDWI ha generato: (i) il 67% di accuratezza massima dal metodo M04 con le TS, (ii) il 67% di accuratezza massima dal metodo M04 con gli SCT, (iii) il 66% di accuratezza massima dal metodo M04 con gli SCR, (iiii) il 30% di accuratezza massima dal metodo M01 con le TSR. L'accuratezza massima globale è pari al 67% per il metodo M04 con l'uso delle TS e degli SCT. L'accuratezza minima globale è pari al 23% per il metodo M02 con l'uso delle TSR. I modelli costruiti con il metodo M04 e i predittori TS, SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-26.

Tabella 3-26: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi.
Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	62	63	63	30
M02	64	63	63	23
M03	65	64	63	28
M04	67	67	66	25

3.2.7 Mappatura dei predittori derivati dal NDMI

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dall'indice NDMI ha generato: (i) il 77% di accuratezza massima dai metodi M03 e M04 con le TS, (ii) il 78% di accuratezza massima dal metodo M01 con gli SCT, (iii) il 78% di accuratezza

massima dal metodo M01 con gli SCR, (iiii) il 53% di accuratezza massima dal metodo M02 con le TSR. L'accuratezza massima globale è pari al 78% per il metodo M01 con l'uso degli SCT e degli SCR. L'accuratezza minima globale è pari al 44% per il metodo M03 con l'uso delle TSR. I modelli costruiti con i metodi M01, M03 e M04 e i predittori TS, SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-27.

Tabella 3-27: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	76	78	78	45
M02	75	76	76	53
M03	77	77	77	44
M04	77	77	77	47

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dall'indice NDMI ha generato: (i) il 67% di accuratezza massima dal metodo M03 con le TS, (ii) il 70% di accuratezza massima dal metodo M03 con gli SCT, (iii) il 70% di accuratezza massima dal metodo M03 con gli SCR, (iiii) il 34% di accuratezza massima dal metodo M02 con le TSR. L'accuratezza massima globale è pari al 70% per il metodo M03 con l'uso degli SCT e degli SCR. L'accuratezza minima globale è pari al 31% per i metodi M03 e M04 con l'uso delle TSR. I modelli costruiti con il metodo M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-28.

Tabella 3-28: Tabella a doppia entrata metodo-predittore per l'area di studio Gola di Frasassi. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	65	66	64	32
M02	62	65	65	34
M03	67	70	70	31
M04	65	69	67	31

3.2.8 Mappatura dei predittori derivati dal totale degli indici

Con questo procedimento è possibile utilizzare simultaneamente tutti gli indici per le diverse categorie e valutare quale tra le combinazioni possibili riesca a produrre la mappa con

l'accuratezza più elevata. In Appendice vengono riportate le mappe e le matrici di confusione associate ai migliori modelli prodotti in termini di accuratezza.

Per l'area di studio Monte Conero, la classificazione dei predittori derivati dal totale degli indici ha generato: (i) l'84% di accuratezza massima dai metodi M01 e M03 con le TS, (ii) l'87% di accuratezza massima dal metodo M03 con gli SCT, (iii) l'87% di accuratezza massima dal metodo M01 con gli SCR, (iiii) il 78% di accuratezza massima dai metodi M01, M02 e M03 con le TSR. L'accuratezza massima globale è pari all'87% per i metodi M01 e M03 con l'uso degli SCT e degli SCR. L'accuratezza minima globale è pari al 77% per il metodo M04 con l'uso delle TSR. I modelli costruiti con i metodi M01 e M03 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-29.

Tabella 3-29: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
M01	84	86	87	78
M02	82	84	84	78
M03	84	87	86	78
M04	82	84	85	77

Le 5 più importanti variabili di predizione per il modello costruito dal metodo M01 con gli SCR derivano rispettivamente dagli indici RI, NDRE, MNDWI e NDVI. La matrice di confusione e il risultato della mappatura vengono riportati in Appendice, Tabella 5-1 e Figura 5-57. Le 5 più importanti variabili di predizione per il modello costruito dal metodo M03 con gli SCT derivano rispettivamente dagli indici RI, NDRE e MCARI. La matrice di confusione e il risultato della mappatura vengono riportati in Appendice, Tabella 5-2 e Figura 5-57.

Per l'area di studio Gola di Frasassi, la classificazione dei predittori derivati dal totale degli indici ha generato: (i) il 76% di accuratezza massima dal metodo M04 con le TS, (ii) il 79% di accuratezza massima da tutti i metodi con gli SCT, (iii) l'80% di accuratezza massima dai metodi M03 e M04 con gli SCR, (iiii) il 69% di accuratezza massima dal metodo M04 con le TSR. L'accuratezza massima globale è pari all'80% per i metodi M03 e M04 con l'uso degli SCR. L'accuratezza minima globale è pari al 67% per il metodo M02 con l'uso delle TSR. I modelli costruiti con i metodi M03 e M04 e i predittori SCT e SCR si dimostrano i più performanti. Per ogni metodo, i modelli meno performanti sono costruiti con le TSR. I valori di accuratezza di ciascun modello sono riportati in Tabella 3-32.

Tabella 3-30: Tabella a doppia entrata metodo-predittore per l'area di studio Monte Conero. Riassume i valori percentuale di accuratezza associati ai diversi modelli di classificazione prodotti.

	TS	SCT	SCR	TSR
<i>M01</i>	0.75	0.79	0.78	0.68
<i>M02</i>	0.75	0.79	0.79	0.67
<i>M03</i>	0.75	0.79	0.80	0.68
<i>M04</i>	0.76	0.79	0.80	0.69

Le 5 più importanti variabili di predizione per il modello costruito dal metodo M03 con gli SCR derivano rispettivamente dagli indici GNDVI, RI, NDVI, MCARI e NDRE. La matrice di confusione e il risultato della mappatura vengono riportati in Appendice, Tabella 5-3 e Figura 5-59. Le 5 più importanti variabili di predizione per il modello costruito dal metodo M03 con gli SCR derivano rispettivamente dagli indici GNDVI, NDVI, RI e MCARI. La matrice di confusione e il risultato della mappatura vengono riportati in Appendice, Tabella 5-4 e Figura 5-60.

3.3 Confronto tra i metodi

Dai risultati ottenuti in entrambe le aree di studio emerge che i modelli prodotti dal metodo M03 (funzione *tsclean* e GAM sui dati giornalieri) si dimostrano i più prestanti in termini di accuratezza. Nello specifico, per l'area di studio Monte Conero, i modelli costruiti con il metodo M03 rientrano sempre tra i migliori, raggiungendo accuratezze superiori o uguali all'80% per gli indici RI, MCARI, GNDVI, NDVI, NDRE. A seguire, M01 ottiene buoni valori di accuratezza nei modelli costruiti con gli indici RI, NDVI e NDRE. I metodi M02 e M04 performano solo per il GNDVI.

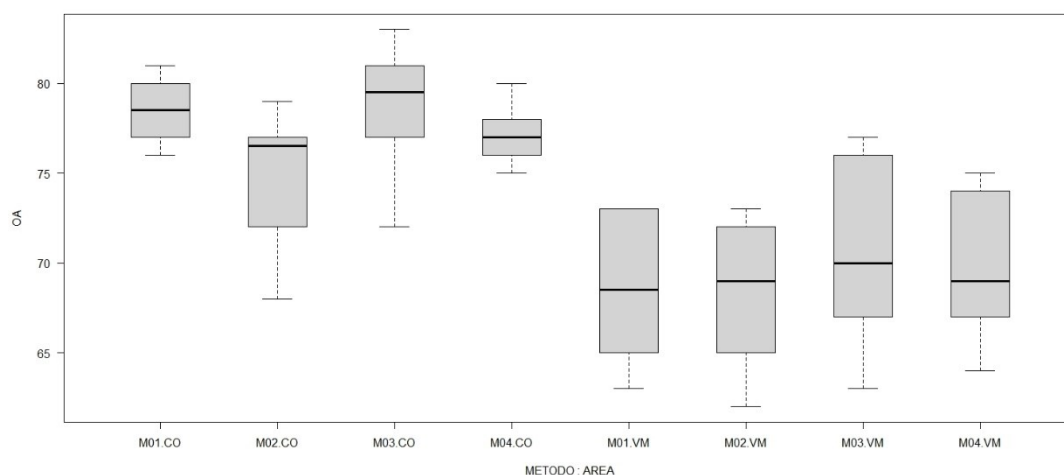


Figura 3-1: Grafico boxplot relativo ai valori di overall accuracy (OA) per metodo (M01, M02, M03, M04) e area di studio (CO, VA). M03 si dimostra il metodo più performante.

Nell'area di studio Gola di Frasassi, sebbene le accuratèzze raggiunte dai modelli siano generalmente inferiori (Figura 4-3), il metodo M03 rientra tra i più prestanti in 6 indici su 7 (RI, MCARI, GNDVI, NDVI, NDRE e NDMI). M01 e M04 seguono ottenendo buone accuratèzze per i modelli costruiti rispettivamente con gli indici NDVI e GNDVI. Tra i modelli costruiti considerando il totale degli indici, con il metodo M03 si raggiungono il massimo delle accuratèzze, pari all'87% nell'area di studio Monte Conero e all'80% nell'area di studio Gola di Frasassi. È necessario sottolineare che tali risultati sono raggiunti anche dai modelli prodotti con i metodi M01 e M04, rispettivamente per la prima e la seconda area di studio.

In Figura 4-4 vengono riportati i *boxplot* relativi al confronto delle accuratèzze tra gli indici e i metodi di pulizia proposti (funzioni *tsclean* e *boxplot*). Per gli indici GNDVI, MCARI, NDMI, NDRE, NDVI e RI la modalità di individuazione dei valori anomali che massimizza le accuratèzze dei modelli è la funzione *tsclean* (pacchetto *forecast*). In particolare, i modelli prodotti dall'indice GNDVI mostrano una differenza poco marcata tra i due metodi proposti. Nel caso del MNDWI si osserva un comportamento a favore della funzione *boxplot* (pacchetto *graphics*), che tuttavia riveste un ruolo marginale nel contesto generale delle accuratèzze dei modelli.

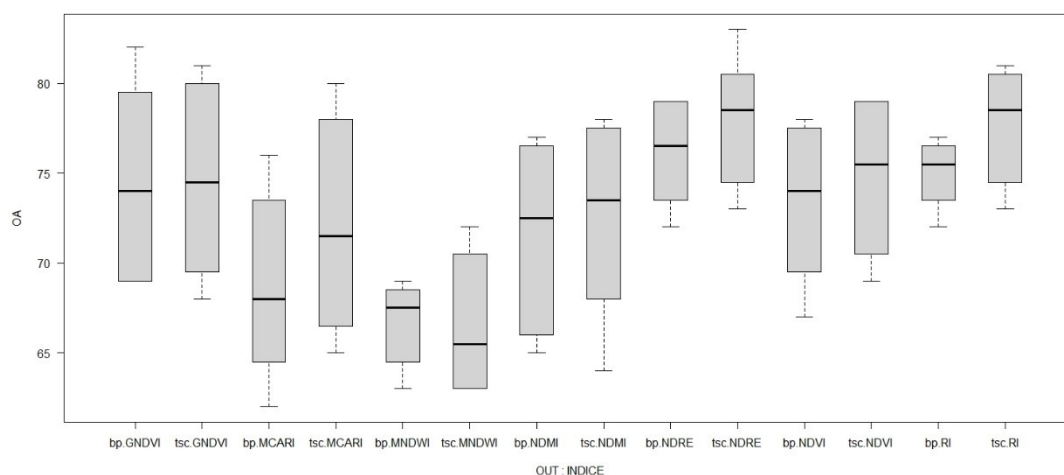


Figura 3-2: Grafico boxplot relativo ai valori di overall accuracy (OA) per indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI) e metodo di pulizia (bp, tsc). La funzione tsclean si dimostra il metodo più performante per gli indici GNDVI, MCARI, NDMI, NDRE, NDVI e RI.

In Figura 4-5 vengono riportati i *boxplot* relativi al confronto delle accuratèzze tra gli indici e i metodi proposti per l'applicazione della GAM (GAM sui dati aggregati e GAM sui dati giornalieri). Per gli indici GNDVI, MCARI, MNDWI, NDMI, NDRE e RI l'applicazione della GAM sui dati giornalieri massimizza le accuratèzze dei modelli prodotti. Solo nel caso dell'indice NDVI si osserva un comportamento pressoché analogo nei confronti di entrambi i metodi.

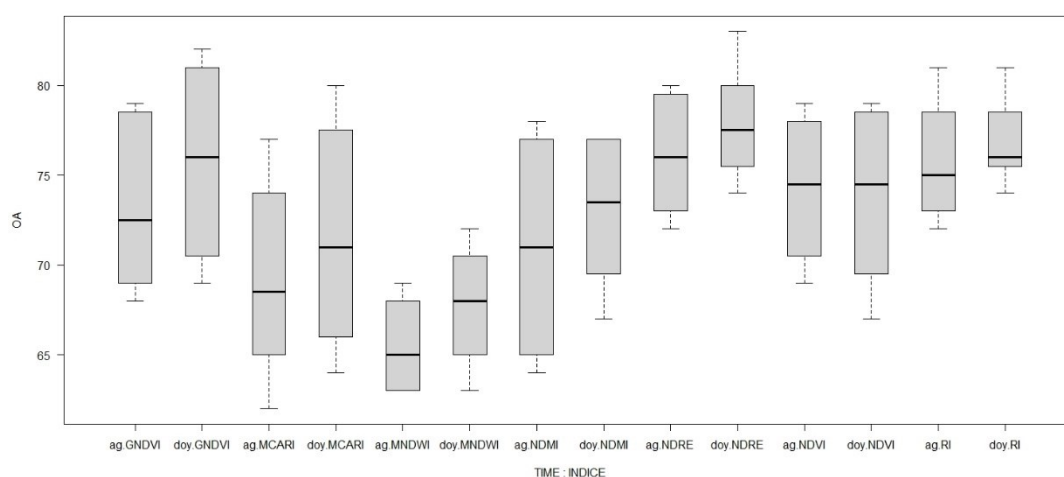


Figura 3-3: Grafico boxplot relativo ai valori di overall accuracy (OA) per indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI) e metodo di applicazione della GAM e delle aggregazioni

(ag, doy). La modalità di applicazione della GAM sui dati giornalieri precedente alle aggregazioni si dimostra il metodo più performante per gli indici GNDVI, MCARI, MNDWI, NDMI, NDRE e RI.

3.4 Confronto tra i predittori

I risultati ottenuti dalla classificazione dei predittori mostrano una sostanziale differenza tra le accuratezze raggiunte dai modelli TS, SCT e SCR e TSR (Figura 4-1). Per ogni metodo e indice, in entrambe le aree di studio, le TSR si dimostrano incapaci di produrre dei modelli che superino un'accuratezza massima del 58% (GNDVI Monte Conero – M04). Solo nei modelli che utilizzano il totale degli indici vengono raggiunte delle accuratezze discrete. La classificazione delle TS rappresenta una soluzione efficace in contesti poco diversificati con l'indice NDVI (NDVI Monte Conero – M01 e M03). Tuttavia, la variabilità intrinseca del dato impedisce ai modelli TS di raggiungere le accuratezze massime. I modelli migliori sono stati prodotti dagli SCT, che riescono a sintetizzare adeguatamente l'informazione contenuta all'interno delle TS. Gli SCR performano altrettanto bene con gli indici NDMI, MCARI e NDRE, si osserva una diminuzione delle accuratezze per GNDVI, MNDWI e NDVI. RI e MCARI producono accuratezze maggiori per l'area di studio Gola di Frasassi. Questo comportamento è dovuto all'informazione contenuta nelle componenti più piccole che, sebbene ridotta, si dimostra influente nella costruzione dei modelli di classificazione.

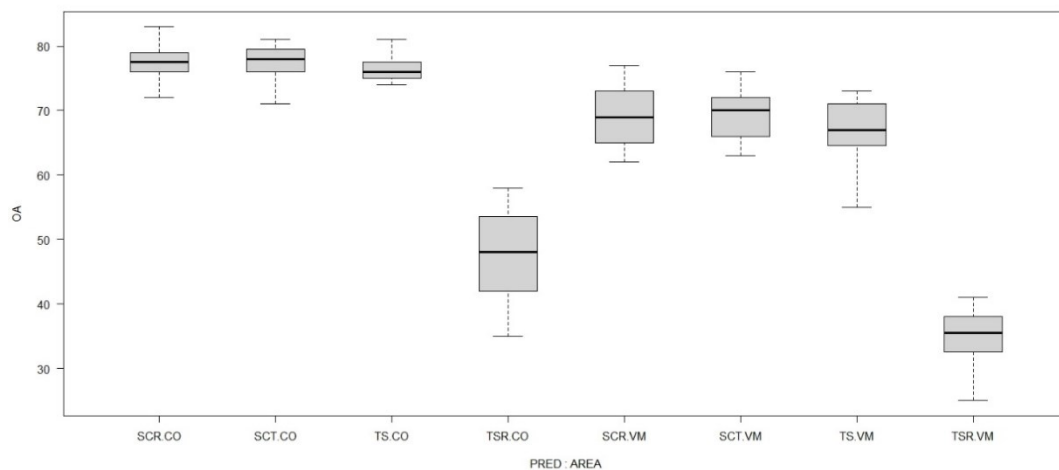


Figura 3-4: Grafico boxplot relativo ai valori di overall accuracy (OA) per predittori (SCR, SCT, TS, TSR) e area di studio (CO, VA). SCT e SCR si dimostrano i predittori più performanti.

Tuttavia, in alcuni casi gli SCR restituiscono dei modelli più accurati, come ad esempio per il Monte Conero indice NDRE metodo M03 (83%). Ipoteticamente, ciò può essere dovuto alla

semplicità vegetazionale che caratterizza l'area del Monte Conero, che conduce quindi il classificatore a necessitare di meno componenti per descrivere il contesto.

Nel grafico riportato in Figura 4-2 viene visualizzate le performance di ciascun indice, per entrambe le aree di studio. Dal più prestante troviamo rispettivamente NDRE, RI, GNDVI e NDVI, NDMI, MCARI e MNDWI.

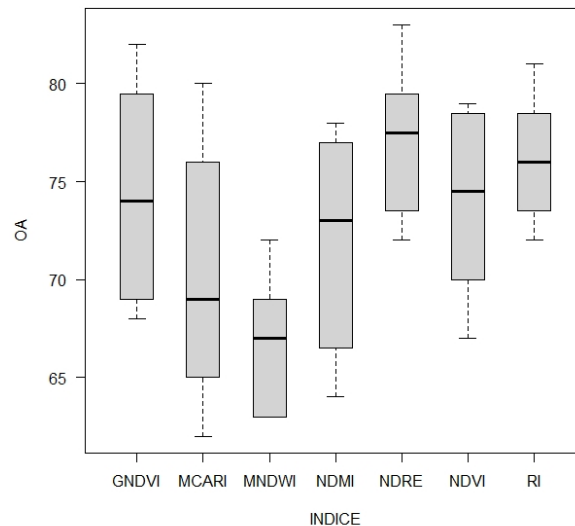


Figura 3-5: Grafico boxplot relativo ai valori di overall accuracy (OA) per indice (RI, MCARI, GNDVI, NDVI, NDRE, MNDWI, NDMI)

Per entrambe le aree di studio, la classificazione effettuata con il totale degli indici ha prodotto le accuratezze migliori. Anche in questo caso i valori più elevati sono stati raggiunti dai modelli costruiti con gli SCT e gli SCR (87% Monte Conero e 80% Gola di Frasassi).

Capitolo 4

DISCUSSIONI E CONCLUSIONI

La classificazione supervisionata dei predittori derivati dalle serie temporali di 186 immagini Sentinel-2 ha permesso di mappare le principali associazioni vegetali presenti in due aree di studio dell'Italia Centrale. In questo lavoro sono stati confrontati 4 differenti metodi di processamento delle serie temporali, che combinano 2 modalità di individuazione dei valori anomali (funzione *tsclean* pacchetto *forecast* e funzione *boxplot* pacchetto *graphics*) e 2 modalità di applicazione delle tecniche di *smoothing* (GAM sui dati aggregati e GAM sui dati giornalieri). I predittori sono derivati da 7 indici di vegetazione, usati per l'addestramento di 256 modelli, 224 individuali e 32 con il totale. I risultati sono valutati in funzione delle accuratezze (*overall accuracy*) raggiunte dai modelli di classificazione. Le serie temporali sono state elaborate con la FPCA per l'estrazione delle componenti principali, descritte dagli *scores*, dagli *eigenvalues* e dalle *eigenfunctions*. Gli *scores* vengono usati nel loro insieme e in una frazione che spiega il 99% della varianza del dataset. Da questi, attraverso le *eigenfunction*, le serie temporali sono state ricostruite. Le serie temporali (originali e ricostruite) e gli *scores* (totali e ridotti) sono stati usati come variabili di predizione per la classificazione delle fitocenosi vegetali.

Dai risultati ottenuti, il metodo che determina una maggiore performance dei modelli prevede l'utilizzo della funzione *tsclean* e l'applicazione della GAM sui dati giornalieri. La funzione *tsclean* permette un'adeguata pulizia del dataset dai valori anomali con tempi di computazione contenuti e si dimostra un importante strumento di processamento delle serie temporali costruite dagli indici di vegetazione. La funzione *boxplot* ha il vantaggio di consentire la sola rimozione dei valori anomali e non la loro sostituzione. Inoltre, sebbene i tempi di computazione si dilatino leggermente, dal punto di vista della programmazione il codice si presta più facilmente ad essere modificato. In contesti più eterogenei come la Gola di Frasassi, la funzione *boxplot* può risultare per alcuni indici più performante. L'applicazione della GAM sui dati giornalieri permette di effettuare le fasi di aggregazione in un secondo momento, dando al codice una maggiore adattabilità in funzione degli obiettivi. La possibilità di manipolare il dataset a partire dai giorni dell'anno rappresenta un vantaggio nella logica del

codice e, di fatto, un efficientamento. Per tutti gli indici, le accuratèzze dei modelli prodotti dal metodo in questione sono analoghe o superiori alle stesse restituite dai modelli che prevedono l'applicazione della GAM sui dati aggregati. I tempi di computazione si dilatano poichè è maggiore l'intervallo di valori entro il quale interpolare i dati (da 1 a 365). Tuttavia, si rende necessario un solo step di aggregazione successivo che in parte compensa quanto aggiunto ed ha la possibilità di modificare il tempo di aggregazione in modo rapido ed efficace.

Sebbene gli algoritmi implementati nel metodo GAM abbiano dimostrato la capacità di interpolare adeguatamente i dati aggregati e giornalieri, l'utilizzo di altre tecniche di *smoothing* è auspicabile per migliorare la costruzione delle serie temporali. Tra queste si suggerisce il metodo LOESS (*locally estimated scatterplot smoothing*), un modello di regressione polinomiale locale ampiamente diffuso per l'analisi statistica. Anche per la fase di individuazione dei valori anomali è possibile pensare a ulteriori soluzioni. Un esempio è la funzione *tsoutliers* (pacchetto *forecast*), che individua la posizione dei valori anomali all'interno del vettore contenente la serie temporale e, a differenza della funzione *tsclean*, non ne effettua la sostituzione automatica. Sono da valutare i tempi di computazione e le difficoltà di applicazione, tuttavia sembra essere una promettente alternativa. Lavori futuri potranno implementare il processo di costruzione delle serie temporali, testare modalità diverse di *smoothing* e individuazione dei valori anomali, nonché altri indici e algoritmi di classificazione.

Per le due aree di studio la differenza nei valori di accuratezza dei modelli prodotti delle classificazioni è evidente e sostanziale. Questi risultati sono da imputare al differente grado di complessità delle fitocenosi vegetali e della copertura del suolo in generale. Nell'area di studio Monte Conero le classi ricercate identificate sono 4, 2 delle quali si riferiscono a boschi della stessa specie (*Quercus ilex*). Nell'area di studio Gola di Frasassi si ritrovano 8 classi inserite in un contesto geomorfologico e topografico assai più complesso. Ciò pone l'accento sulla necessità di testare le metodologie descritte in contesti diversi e diversificati, così da valutarne l'affidabilità per più ambienti. L'uso di immagini a differente risoluzione spaziale può incidere sui risultati finali della classificazione. Come visto per Quattrini et al., (2020) alcune categorie si prestano meglio ad essere descritte con una risoluzione maggiore (20 metri). In questo lavoro si sono utilizzate immagini Sentinel-2 a 10 metri di risoluzione; pertanto, sarebbe interessante valutare a parità di condizioni quanto questa incida nella zona della Gola di Frasassi. Per l'area del Monte Conero i risultati hanno confermato le precedenti classificazioni effettuate (Pesaresi et al., 2020), sebbene queste abbiano utilizzato immagini Landsat a 30 metri di risoluzione.

Le accuratezze massime pervenute sono pari all'87% per l'area di studio Monte Conero e all'80% per l'area di studio Gola di Frasassi. Gli *scores* si sono confermati i predittori che meglio riescono a descrivere l'informazione contenuta nelle serie temporali riuscendo al contempo a ridurre la dimensione del dataset e diminuire i tempi di computazione. Dalle evidenze raggiunte in questo lavoro è possibile affermare che il codice sia stato implementato adeguatamente, efficientandone alcune porzioni e garantendo allo stesso tempo dei risultati conformi ai precedenti studi. In Quattrini et al., (2020), per l'area di studio Gola di Frasassi, la classificazione degli *scores* derivati dalle serie temporali Sentinel-2 hanno raggiunto l'85% di accuratezza. Tuttavia, nel suddetto lavoro sono state prese in considerazione più classi rispetto a quelle individuate nell'ambito di questa tesi, nonché ulteriori variabili topografiche e litologiche; inoltre, la risoluzione spaziale delle immagini è di 20 metri anziché 10. Per l'area di studio Monte Conero l'accuratezza massima ottenuta dalla classificazione è in linea con quanto riportato in Pesaresi *et al.*, (2020), sebbene venga usato esclusivamente il NDVI calcolato su immagini Landsat (30 metri di risoluzione spaziale).

La generazione di mappe della vegetazione è oggi eseguita principalmente mediante fotointerpretazione, che è un metodo che richiede tempo e denaro. Questo lavoro si inserisce all'interno di una filiera, nel contesto di una partnership tra le più interessanti all'interno dell'ateneo UNIVPM tra il Dipartimento di Ingegneria Informatica e il Dipartimento di Scienze Agrarie, Alimentari e Ambientali, che tenta di presentare un nuovo quadro metodologico basato sulla FPCA per la mappatura delle associazioni vegetali. I risultati ottenuti mostrano che è possibile identificare le principali variazioni stagionali intra-annuali per i 4 metodi proposti utilizzando immagini Sentinel-2. Il metodo che prevede l'uso della funzione *tsclean* e della GAM applicata ai dati giornalieri si dimostra efficiente e migliorativo del procedimento. Le implicazioni da questo punto di vista sono molteplici, ma il vantaggio principale riguarda la possibilità di manipolare il codice con facilità, abbattendo i tempi di computazione. Gli *scores* hanno un'enorme potenzialità di sviluppare analisi integrate tra i tradizionali dati fitosociologici di campo e i dati telerilevati. Questo tipo di approccio è efficiente e accurato e offre la possibilità di ripetere le analisi periodicamente per produrre classificazioni sempre aggiornate.

Capitolo 5

APPENDICE

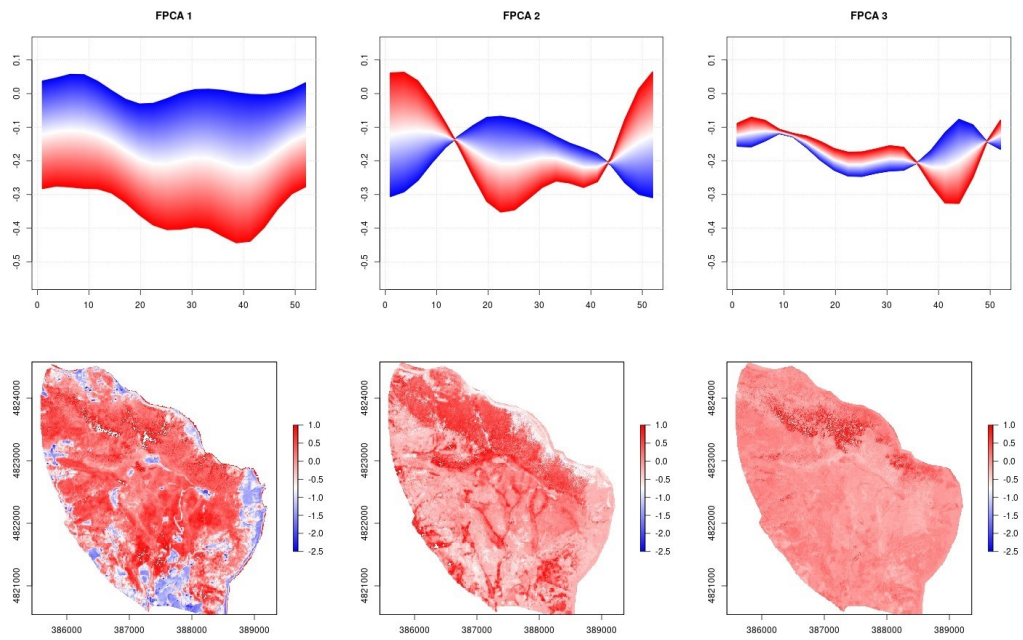


Figura 5-1: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice RI.

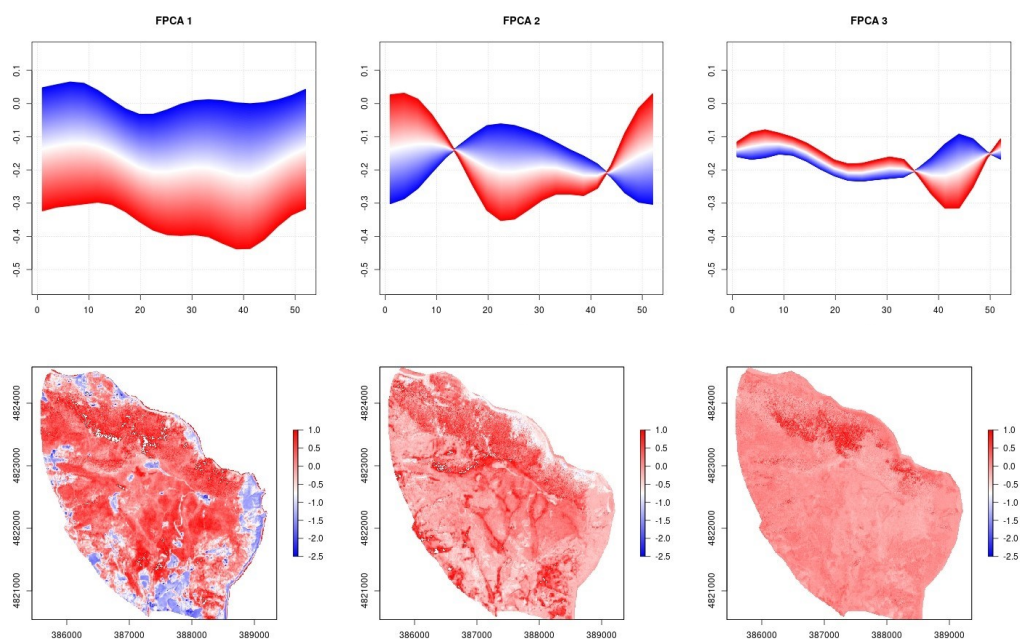


Figura 5-2: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice RI.

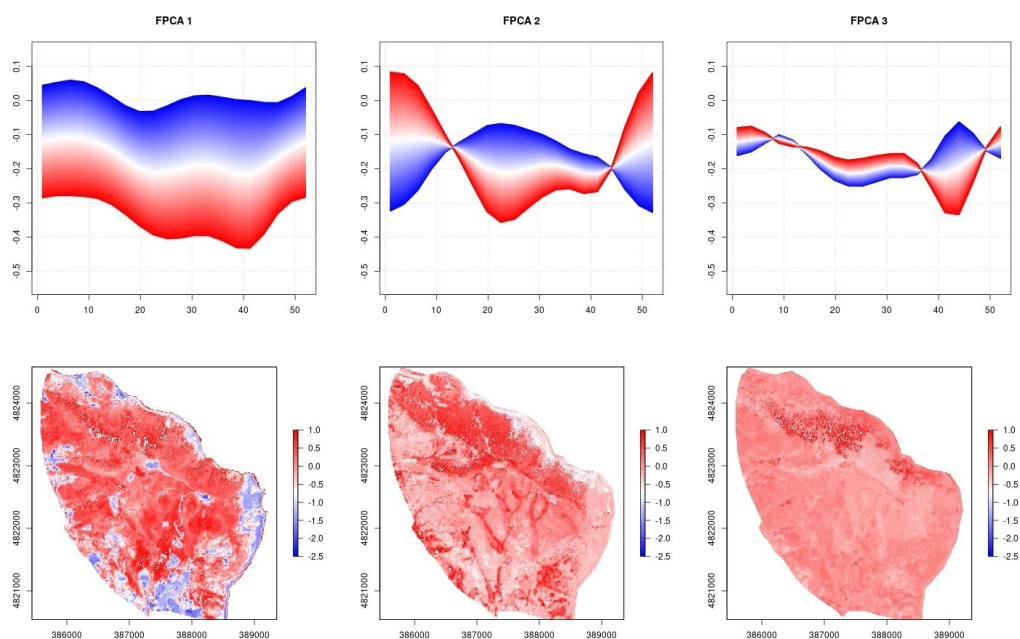


Figura 5-3: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice RI.

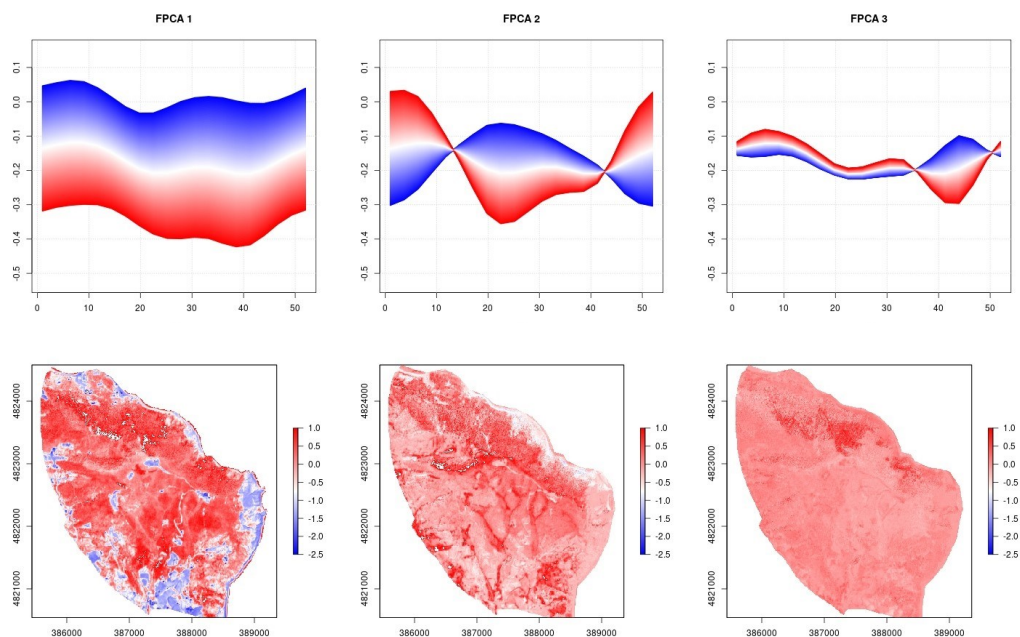


Figura 5-4: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice RI.

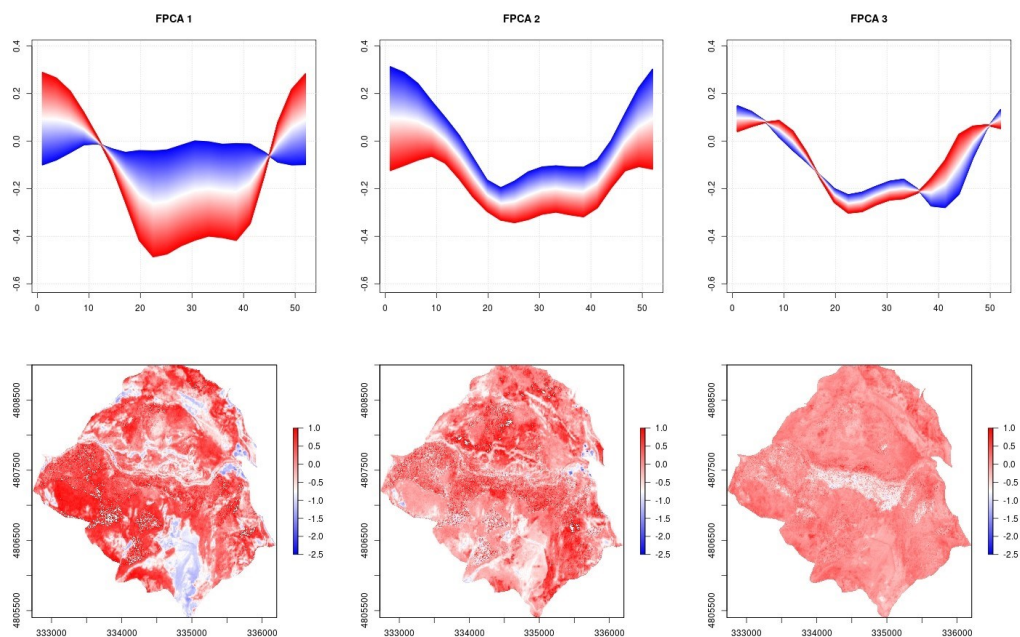


Figura 5-5: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice RI.

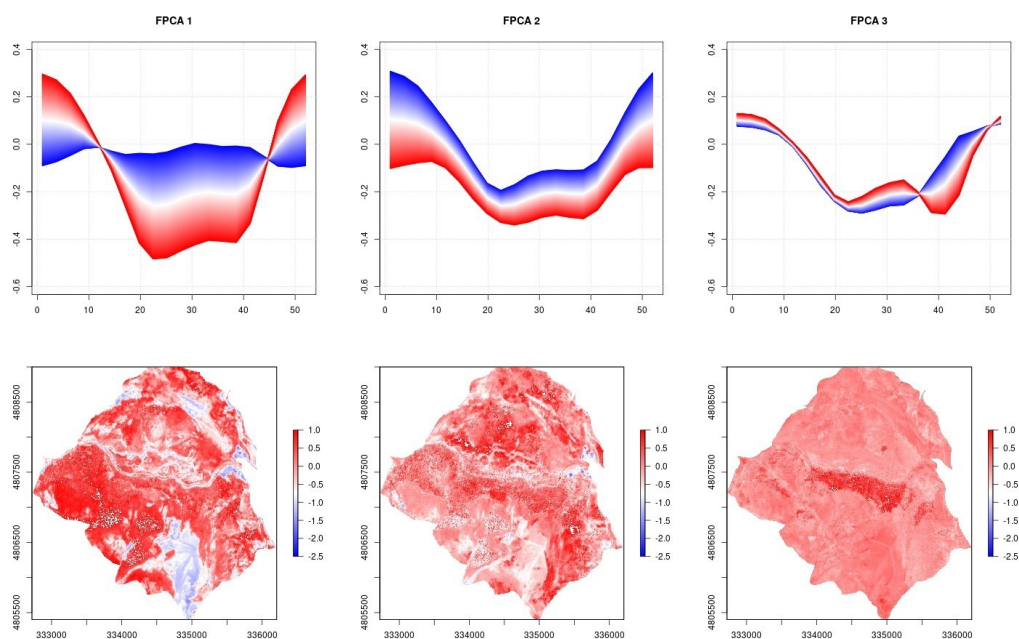


Figura 5-6: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice RI.

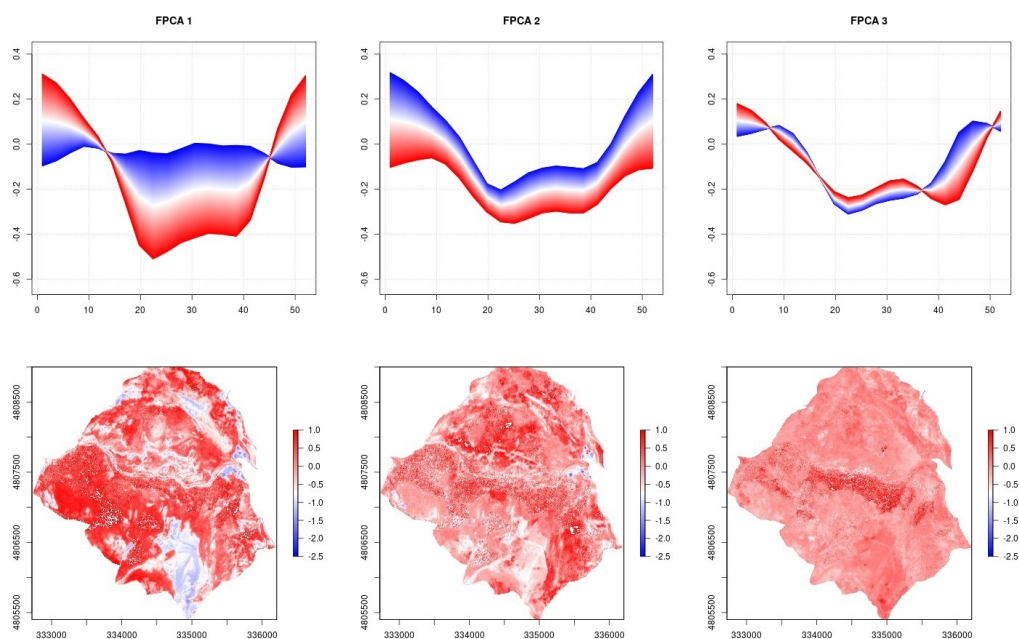


Figura 5-7: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice RI.

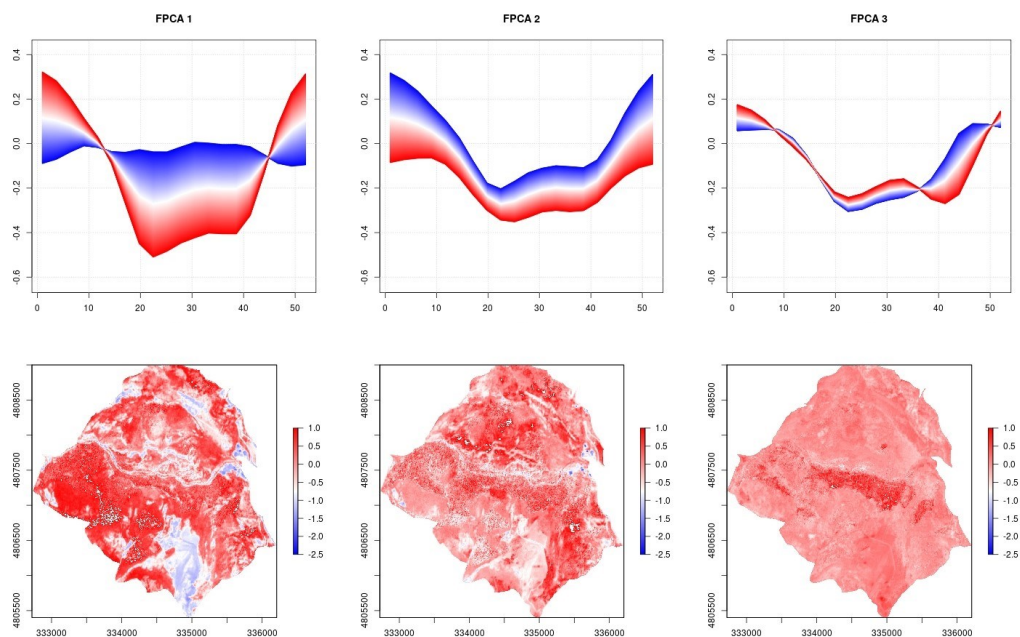


Figura 5-8: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice RI.

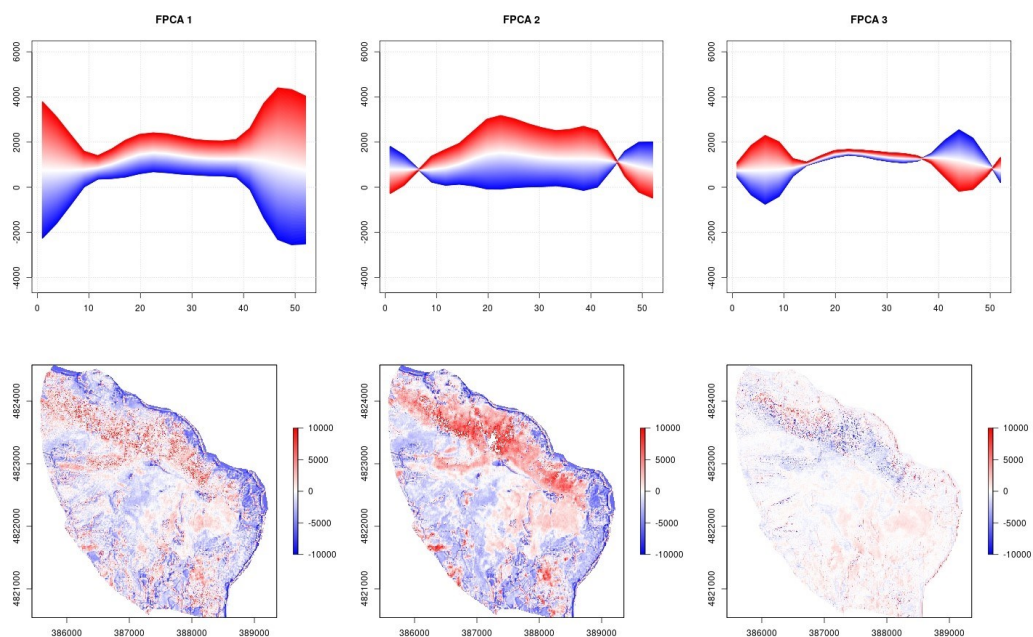


Figura 5-9: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice MCARI.

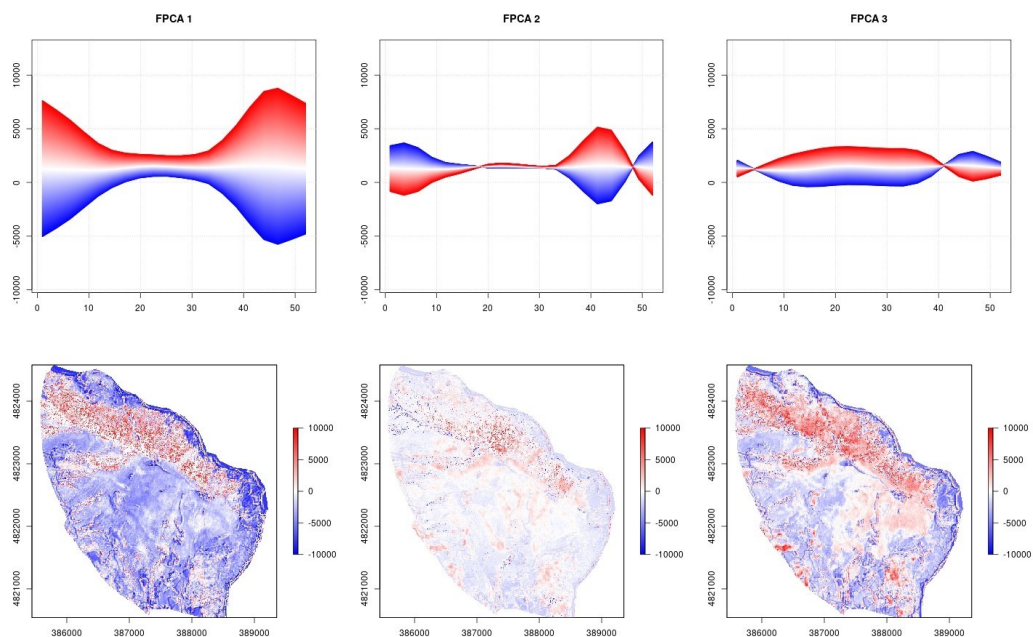


Figura 5-10: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice MCARI.

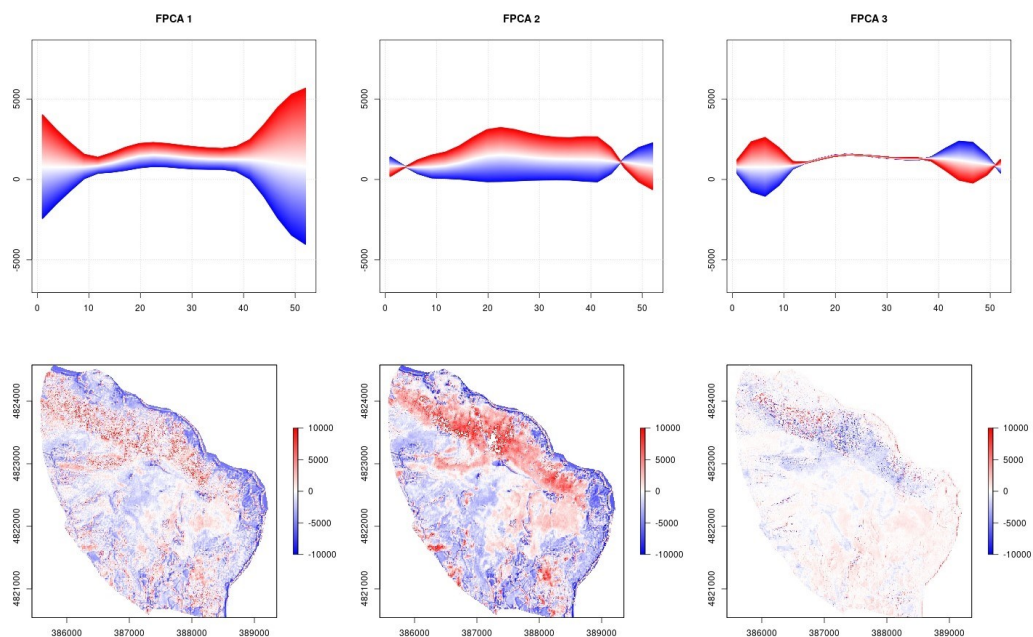


Figura 5-11: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice MCARI.

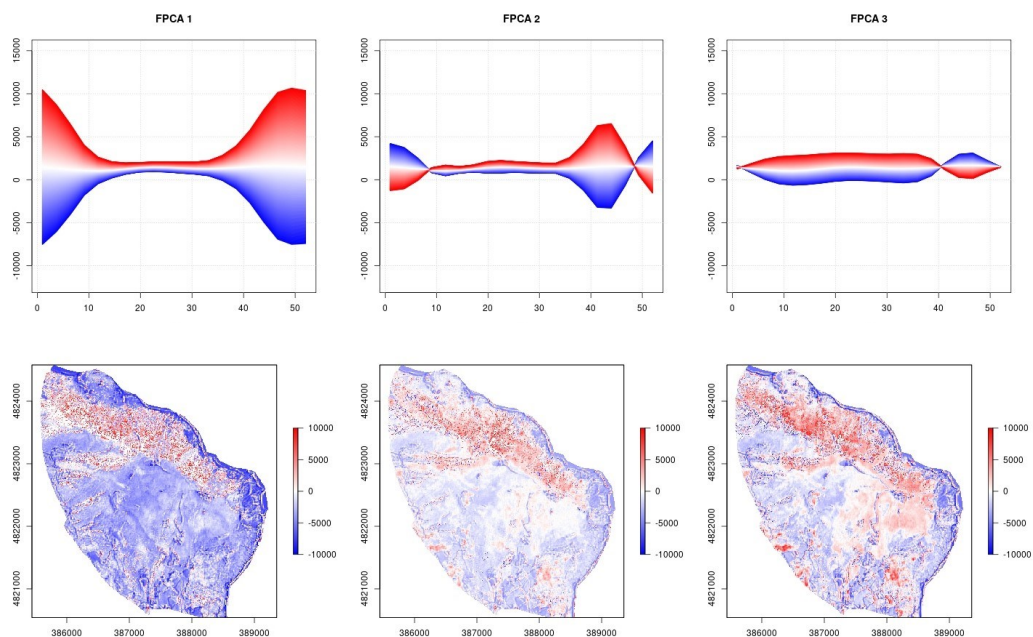


Figura 5-12: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice MCARI.

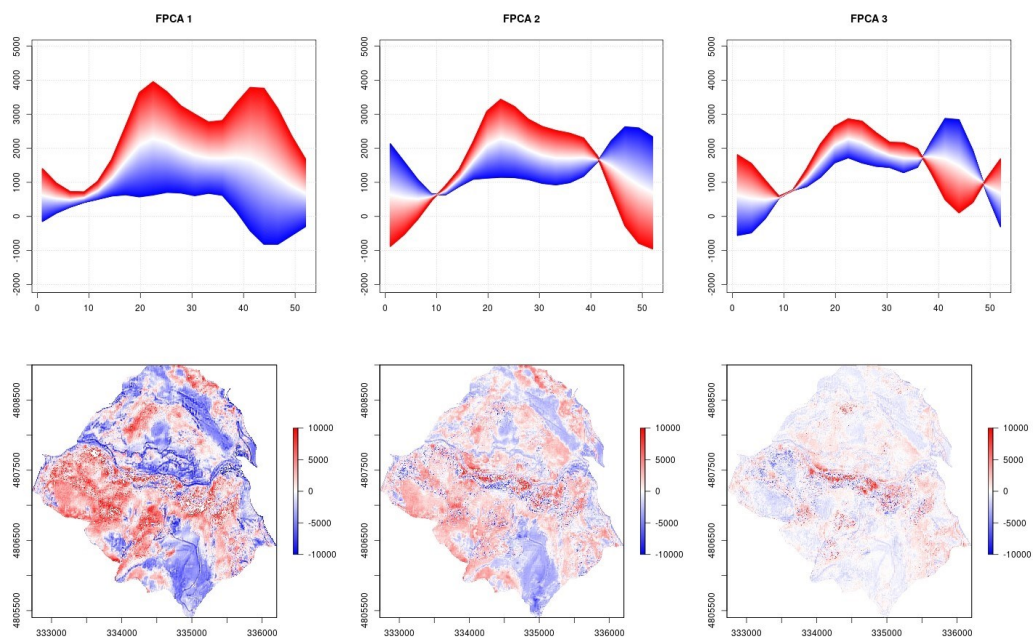


Figura 5-13: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice MCARI.

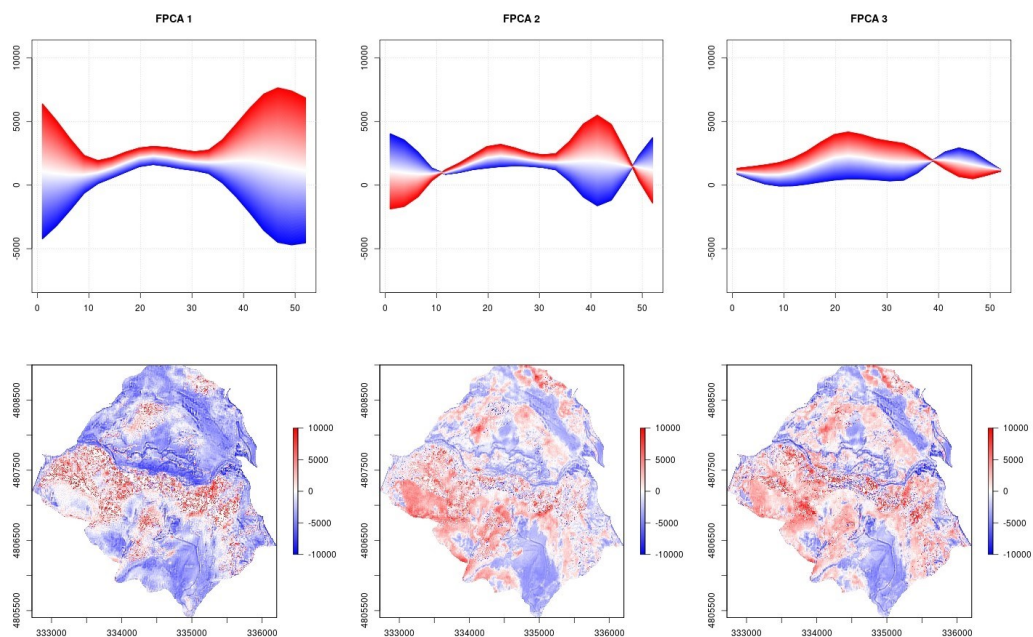


Figura 5-14: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice MCARI.

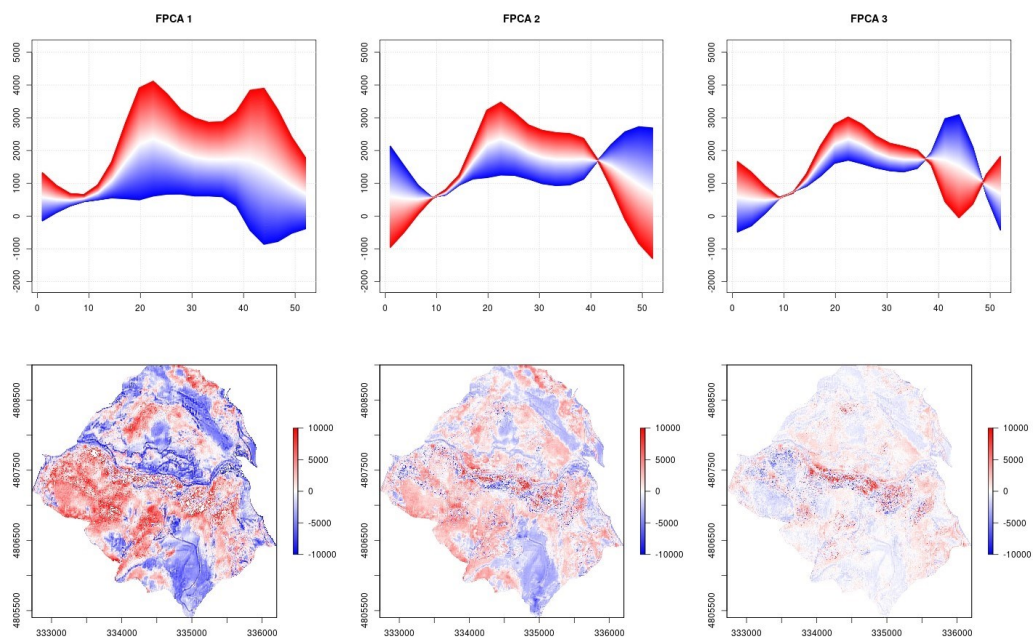


Figura 5-15: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice MCARI.

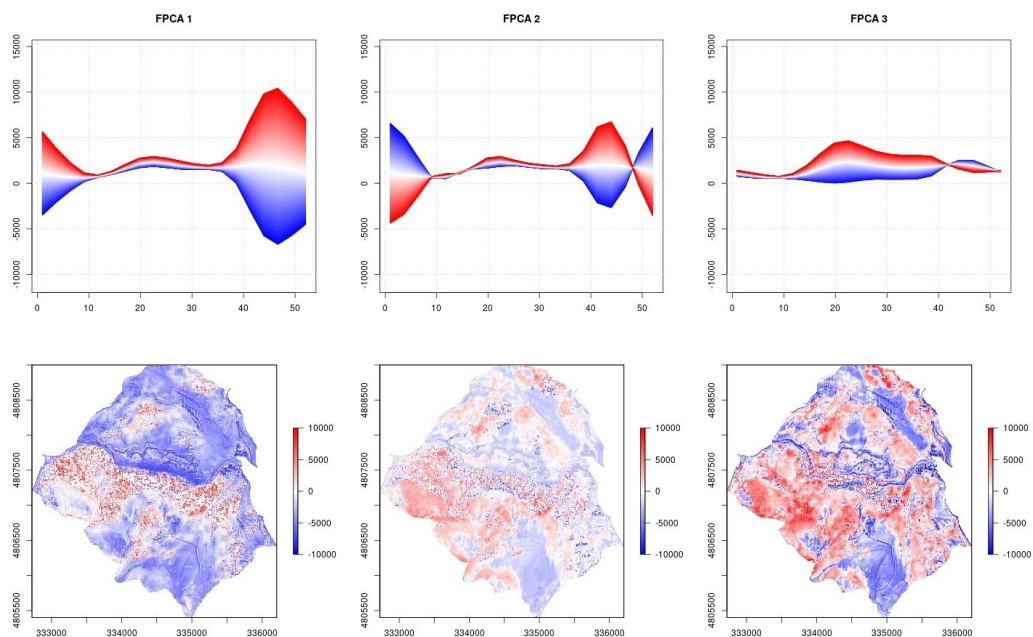


Figura 5-16: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice MCARI.

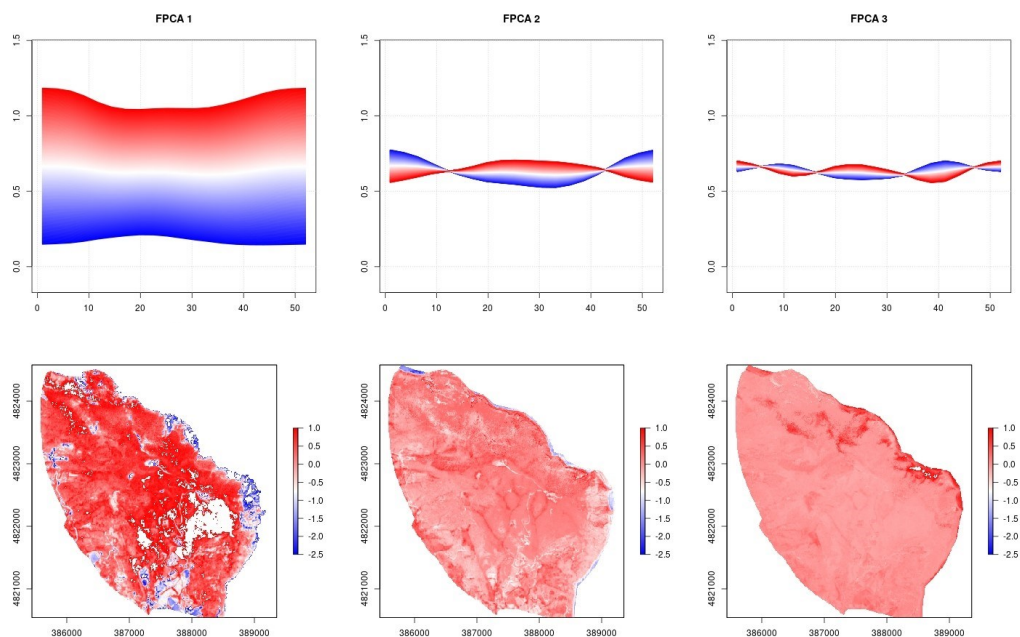


Figura 5-17: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice GNDVI.

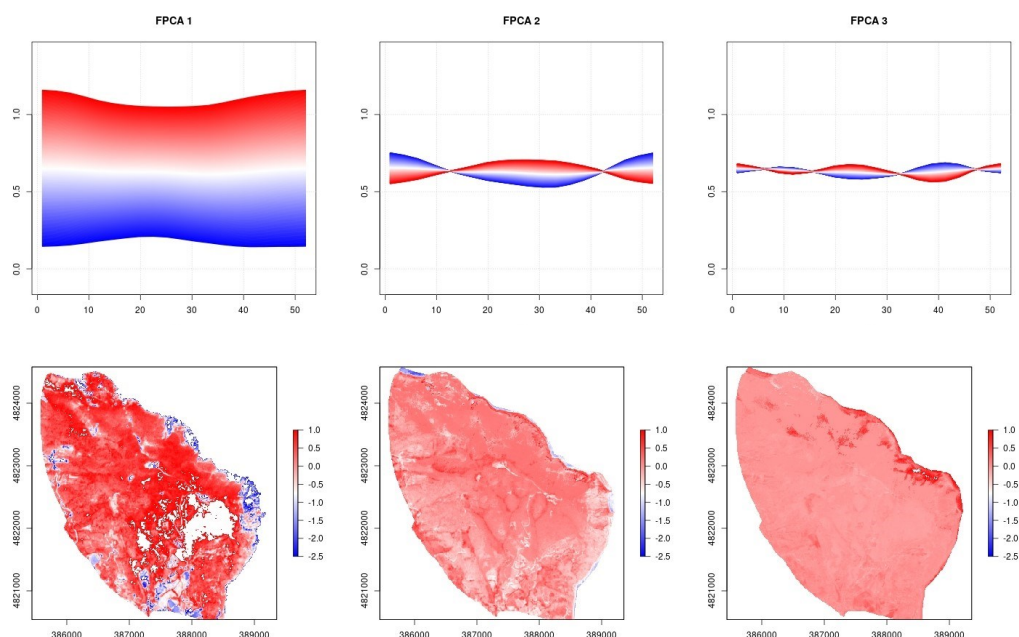


Figura 5-18: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice GNDVI.

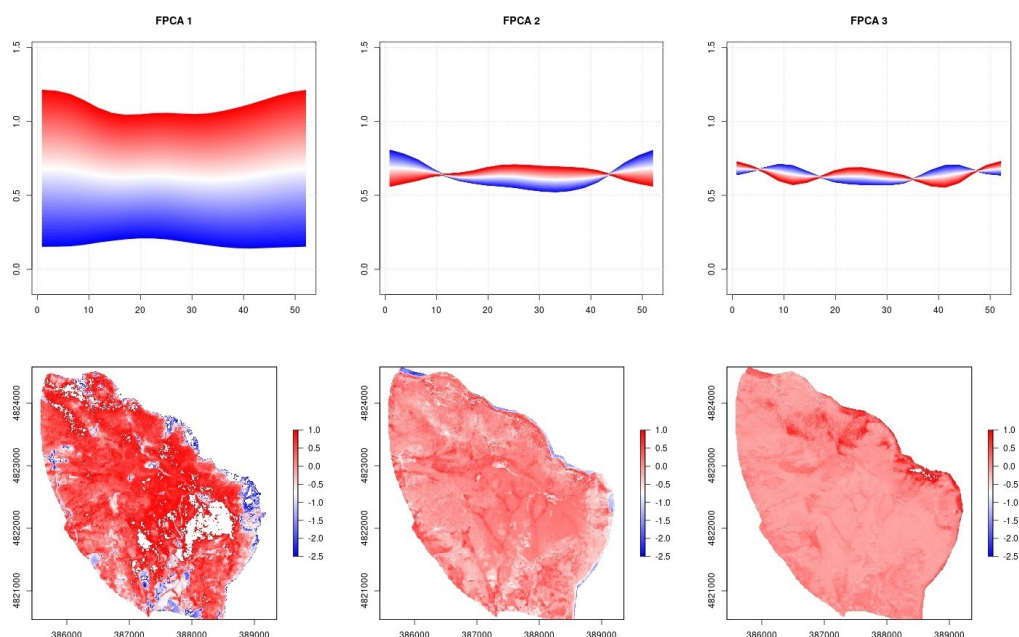


Figura 5-19: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice GNDVI.

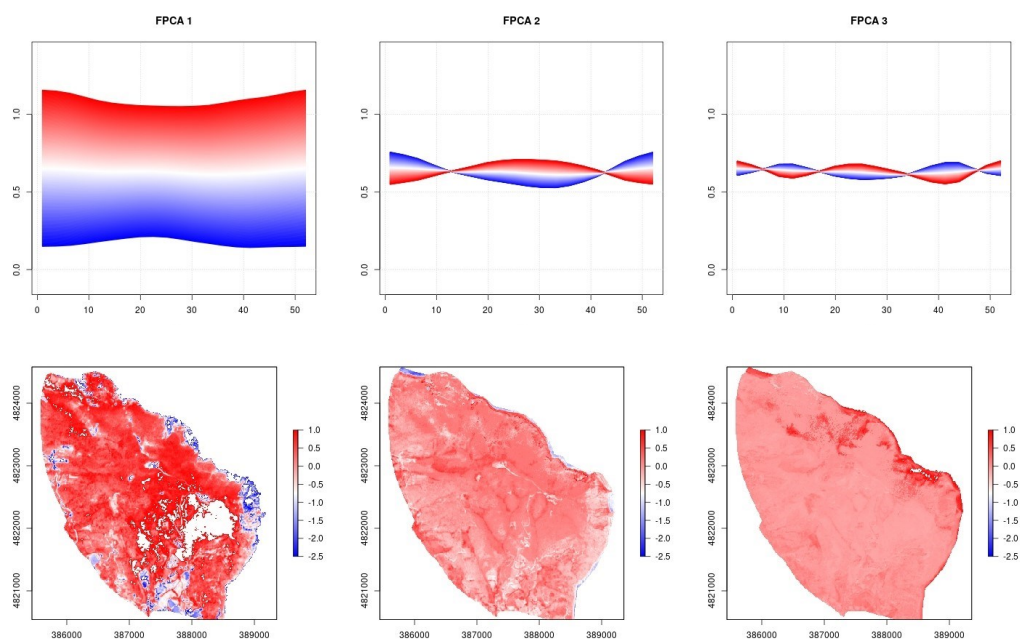


Figura 5-20: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice GNDVI.

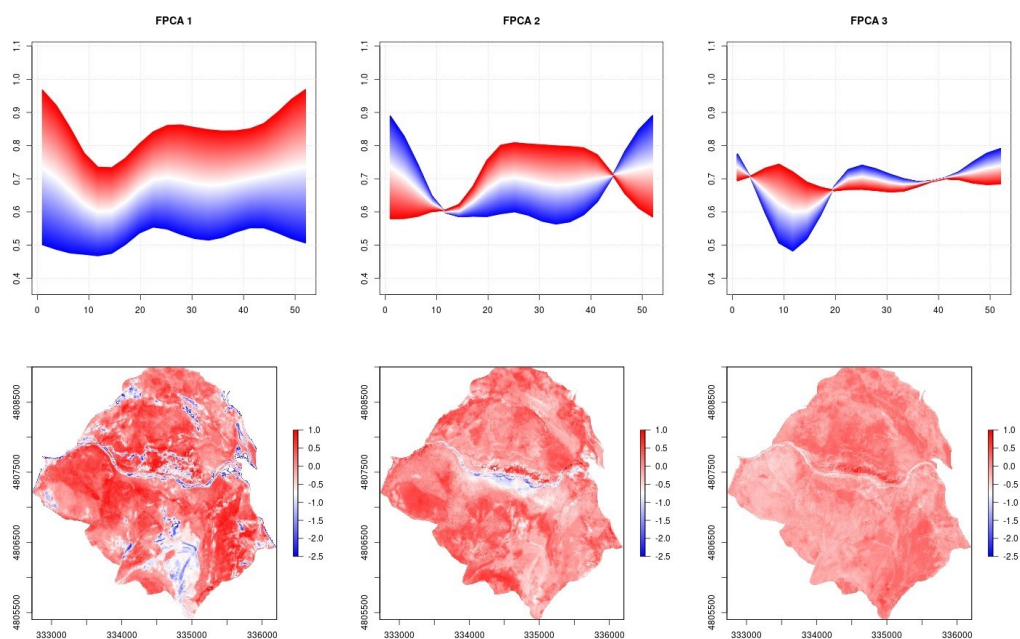


Figura 5-21: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice GNDVI.

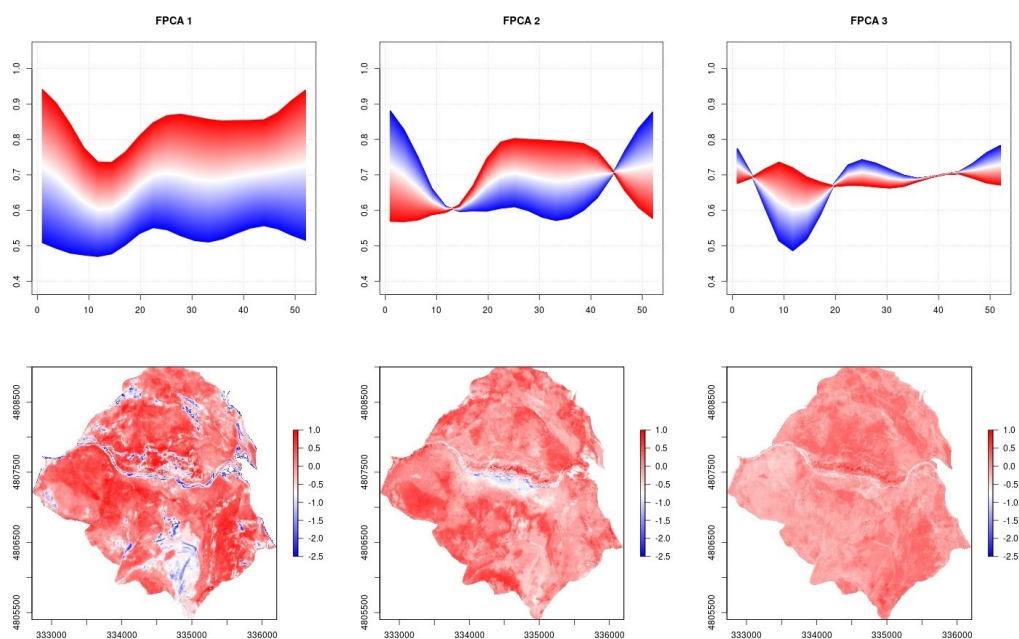


Figura 5-22: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice GNDVI.

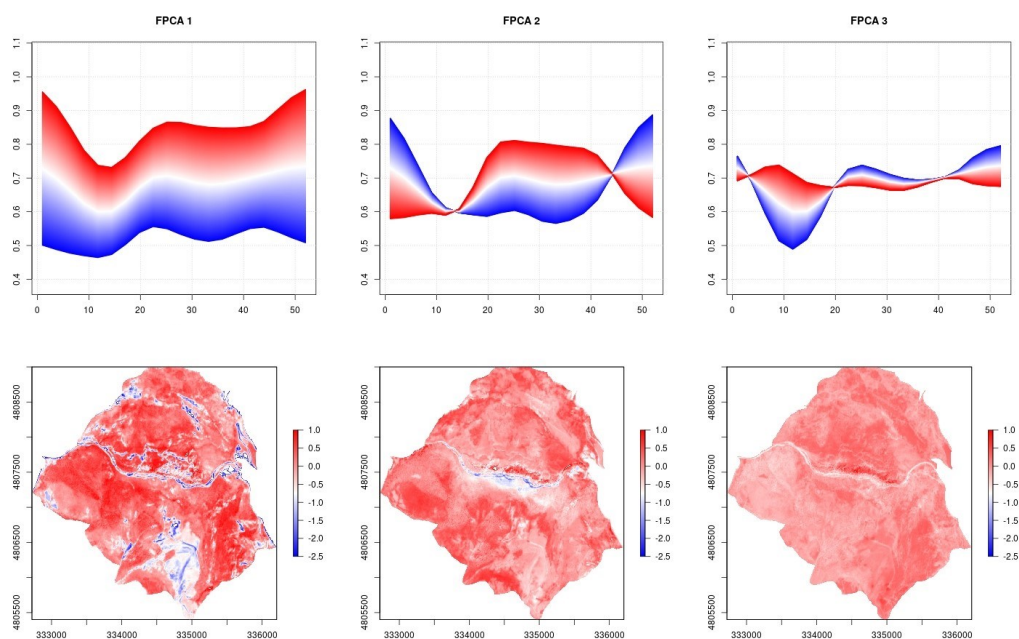


Figura 5-23: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice GNDVI.

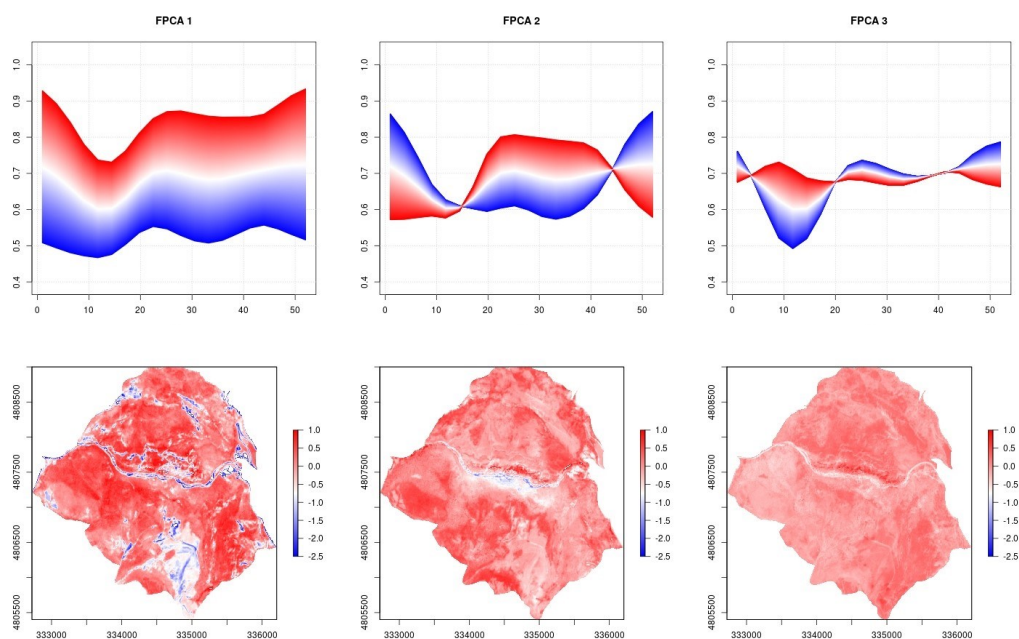


Figura 5-24: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice GNDVI.

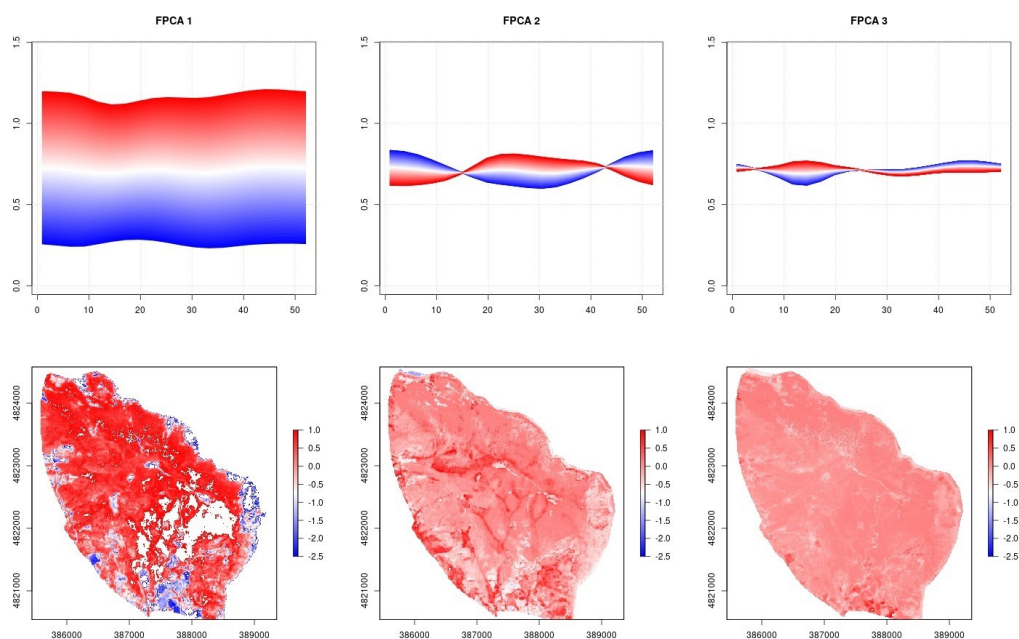


Figura 5-25: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice NDVI.

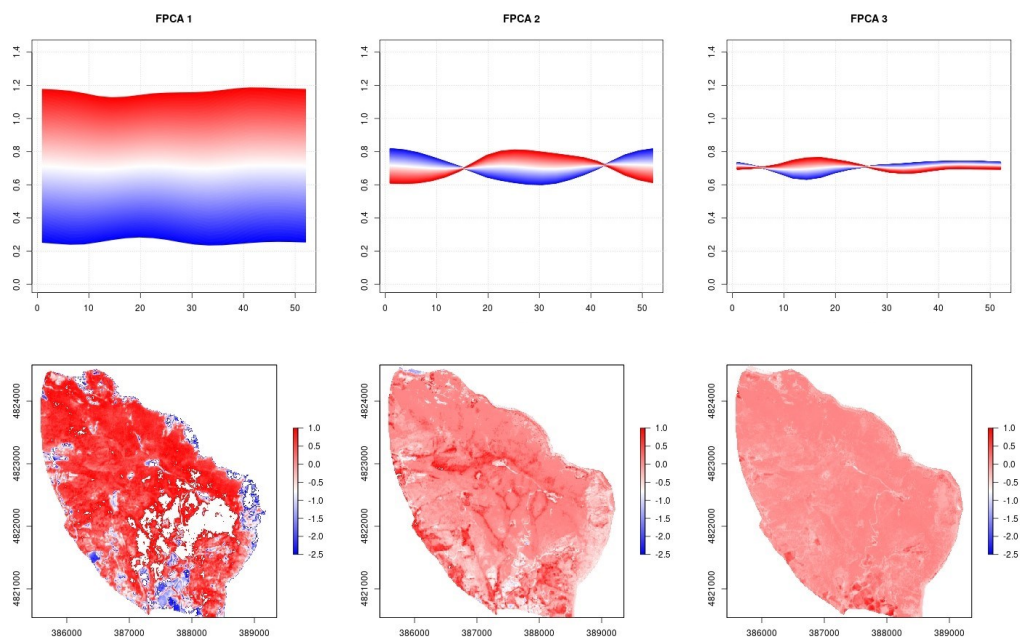


Figura 5-26: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice NDVI.

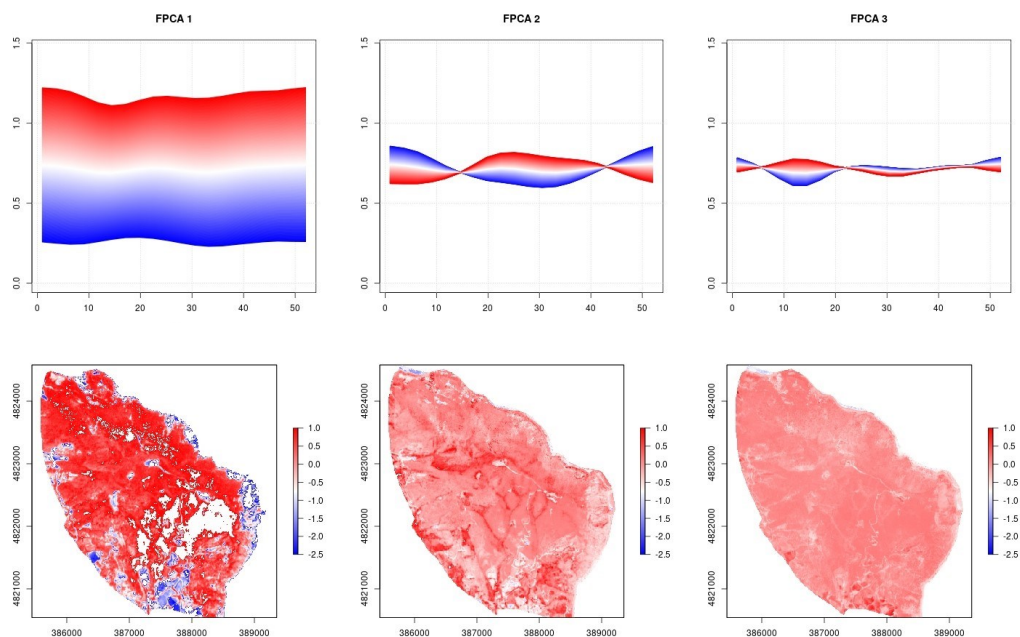


Figura 5-27: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice NDVI.

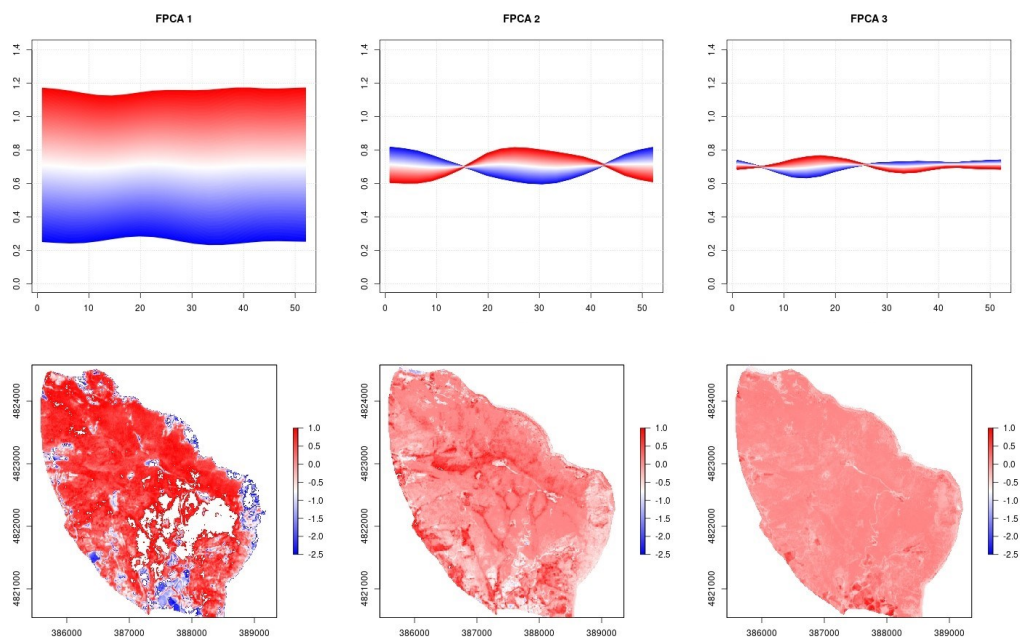


Figura 5-28: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice NDVI.

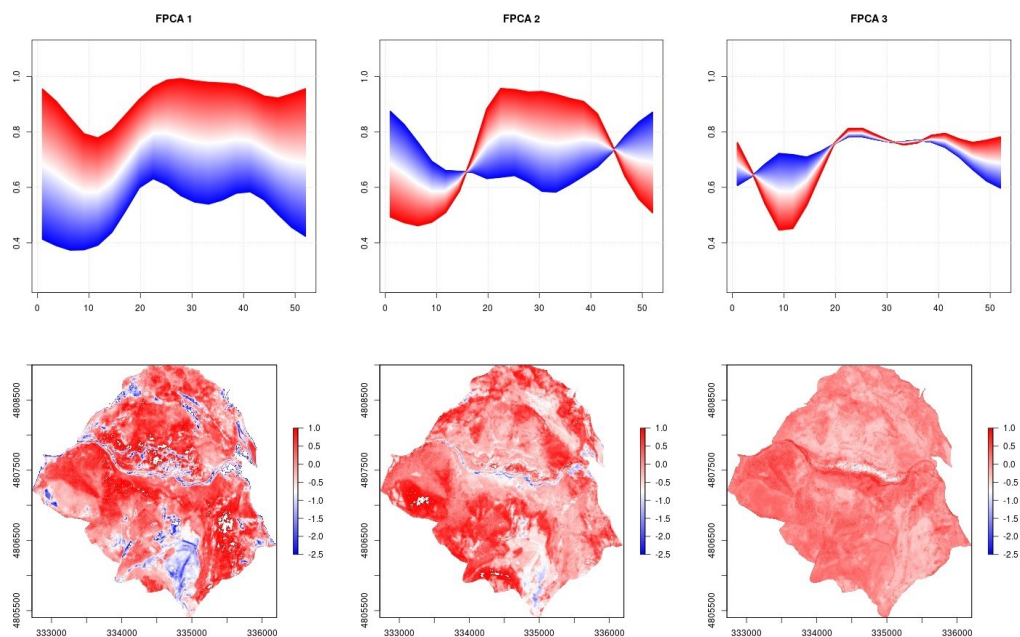


Figura 5-29: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice NDVI.

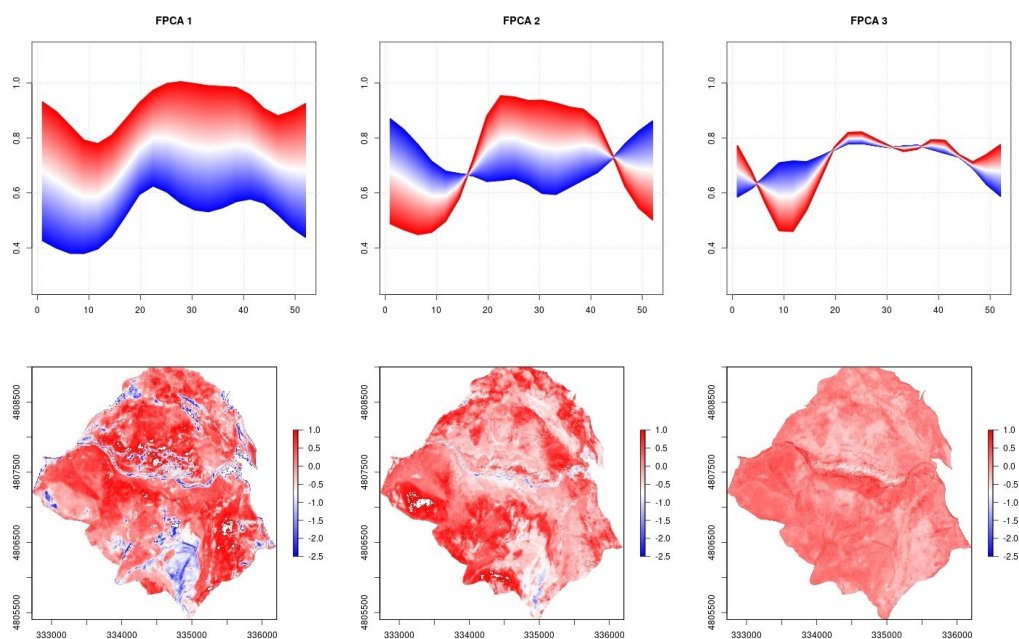


Figura 5-30: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice NDVI.

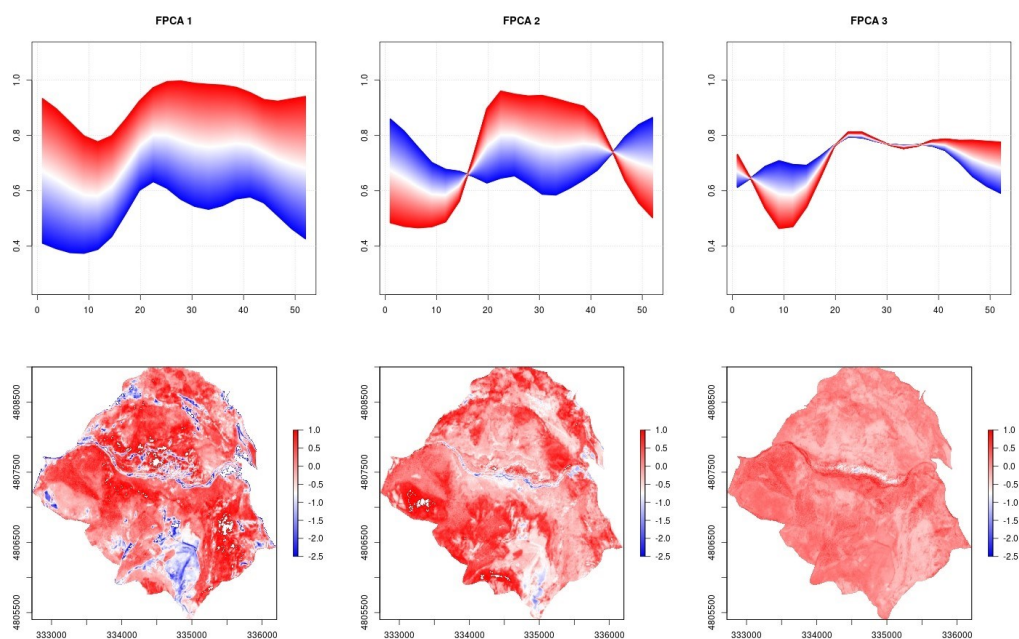


Figura 5-31: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice NDVI.

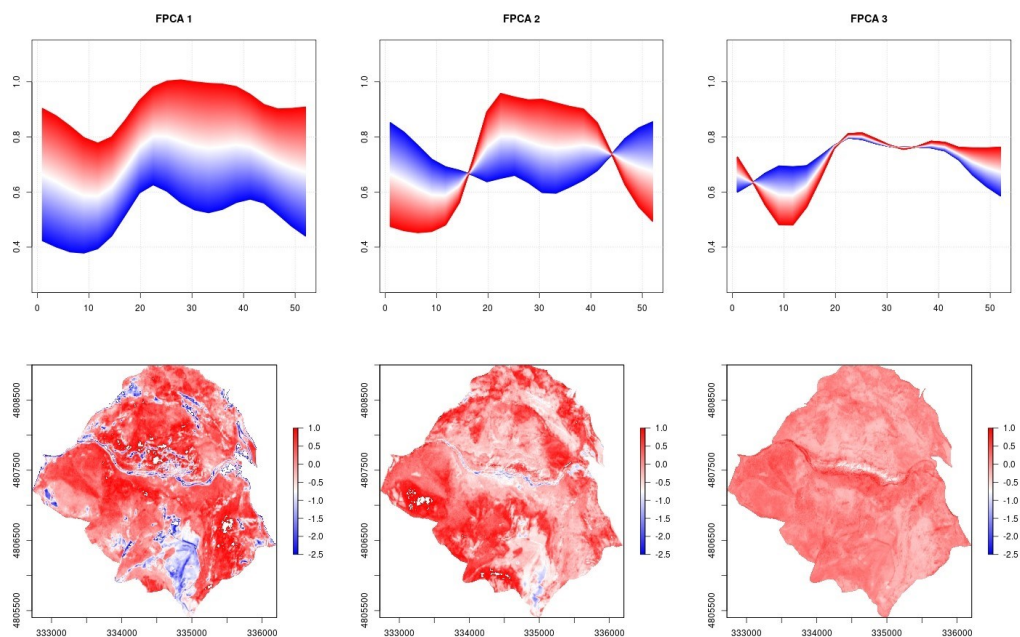


Figura 5-32: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice NDVI.

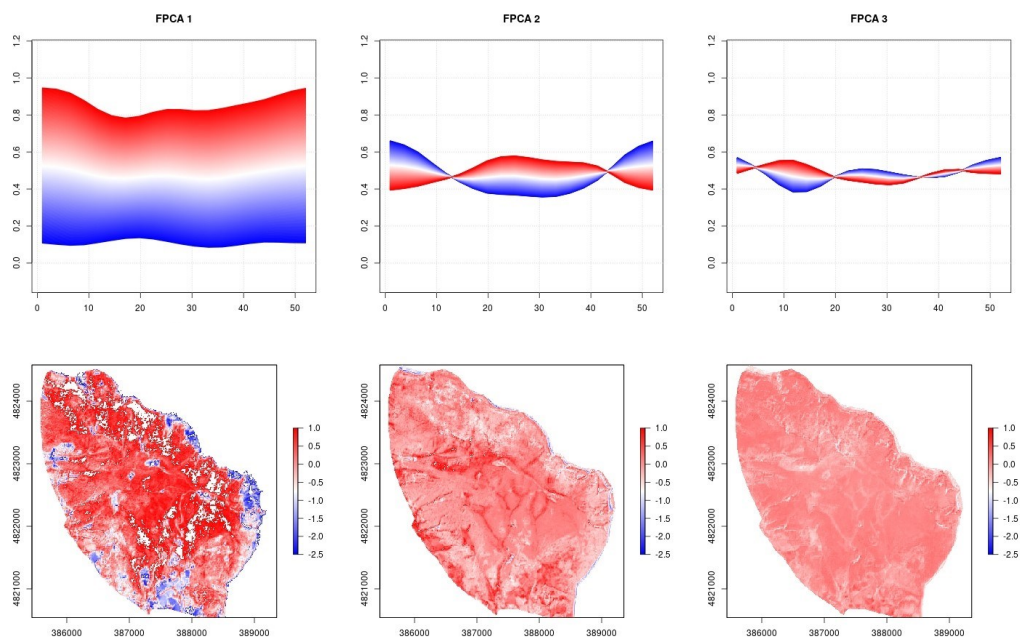


Figura 5-33: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice NDRE.

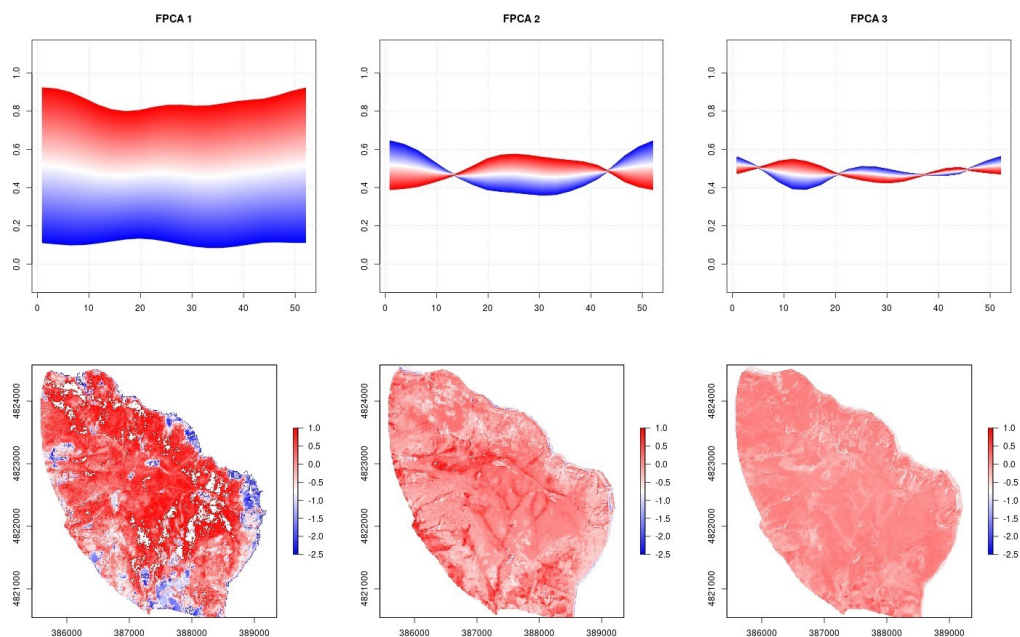


Figura 5-34: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice NDRE.

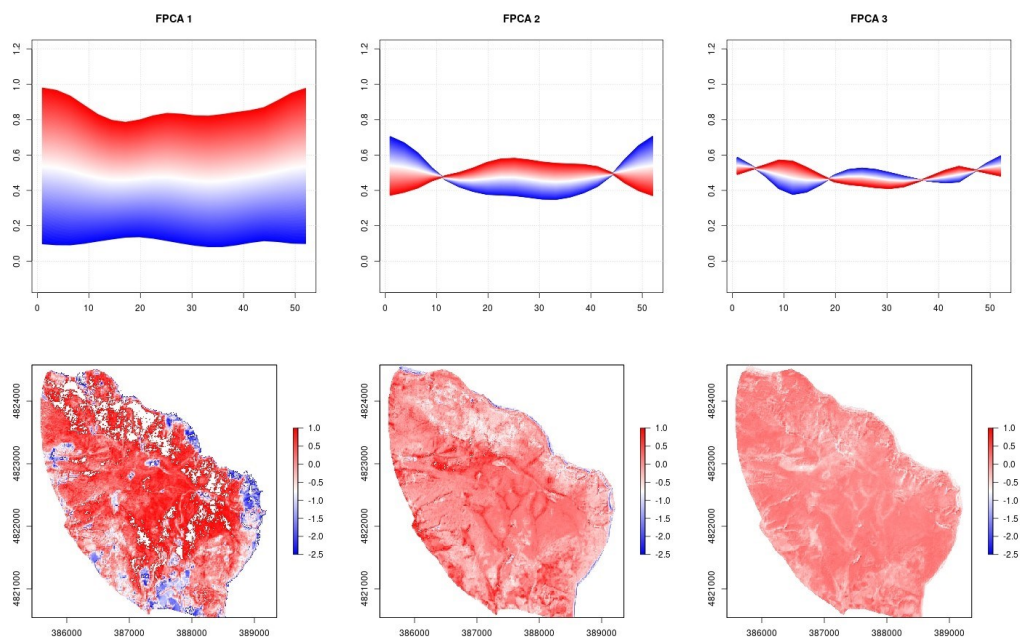


Figura 5-35: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice NDRE.

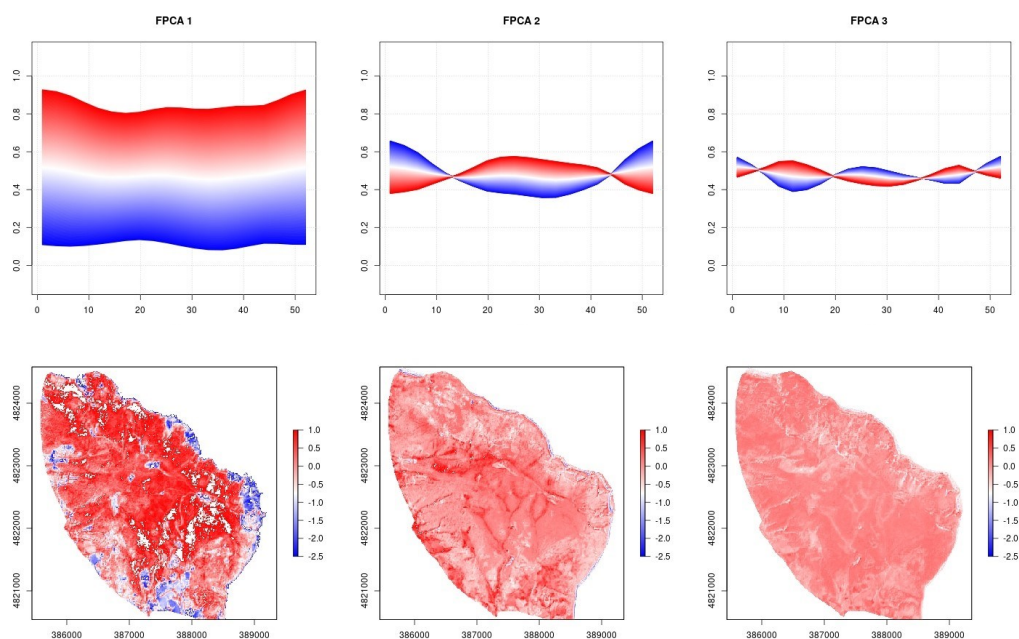


Figura 5-36: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice NDRE.

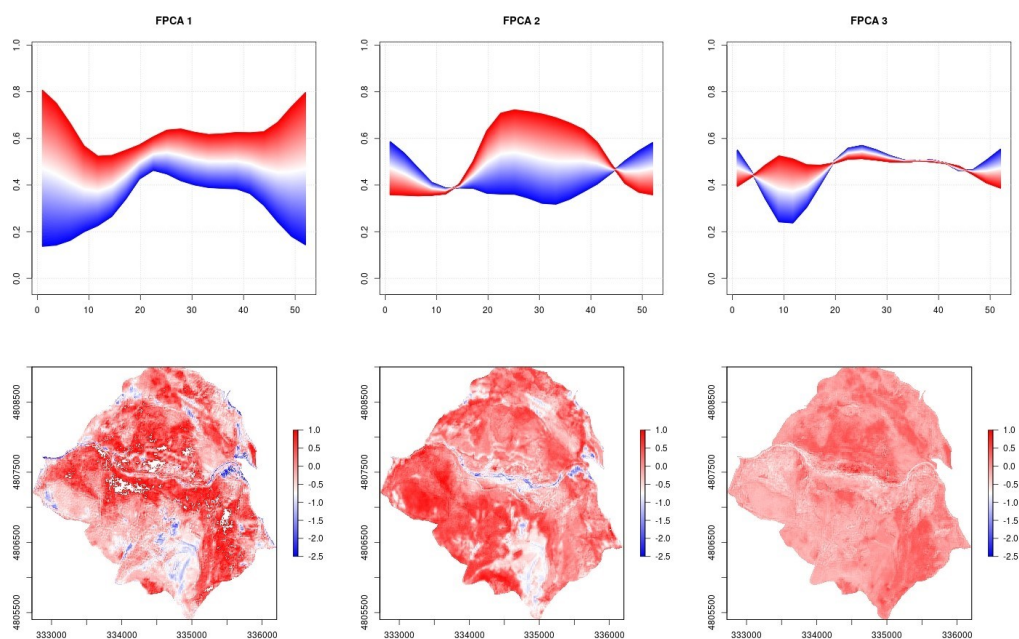


Figura 5-37: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice NDRE.

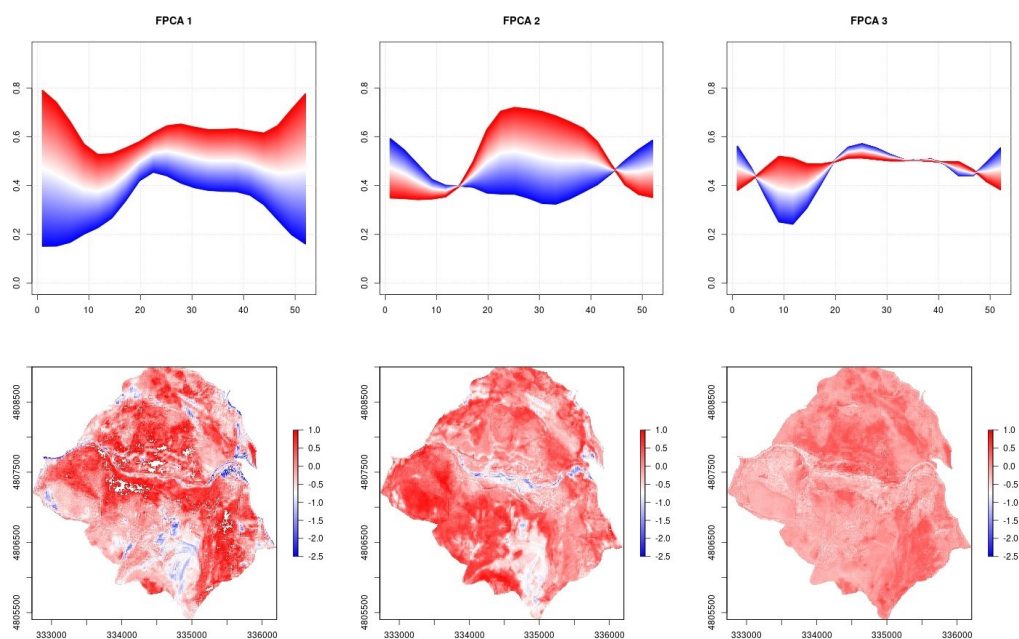


Figura 5-38: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice NDRE.

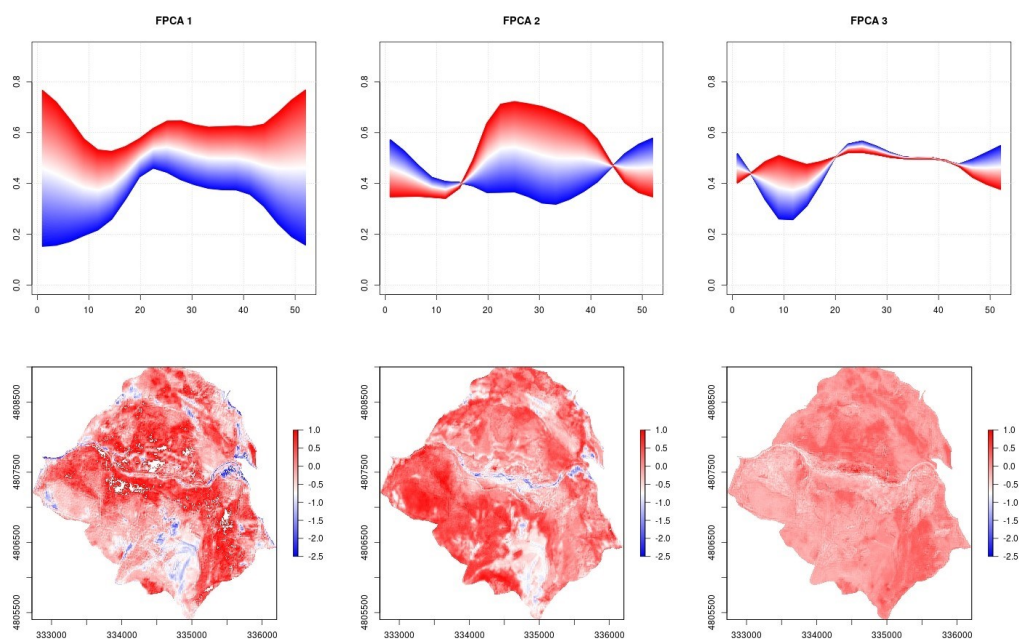


Figura 5-39: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice NDRE.

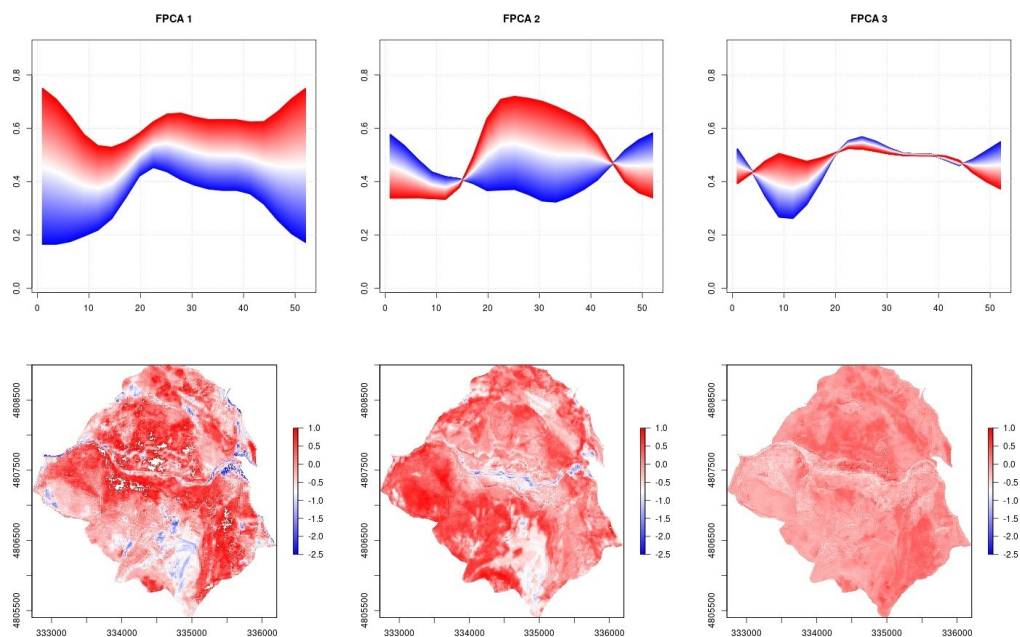


Figura 5-40: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice NDRE.

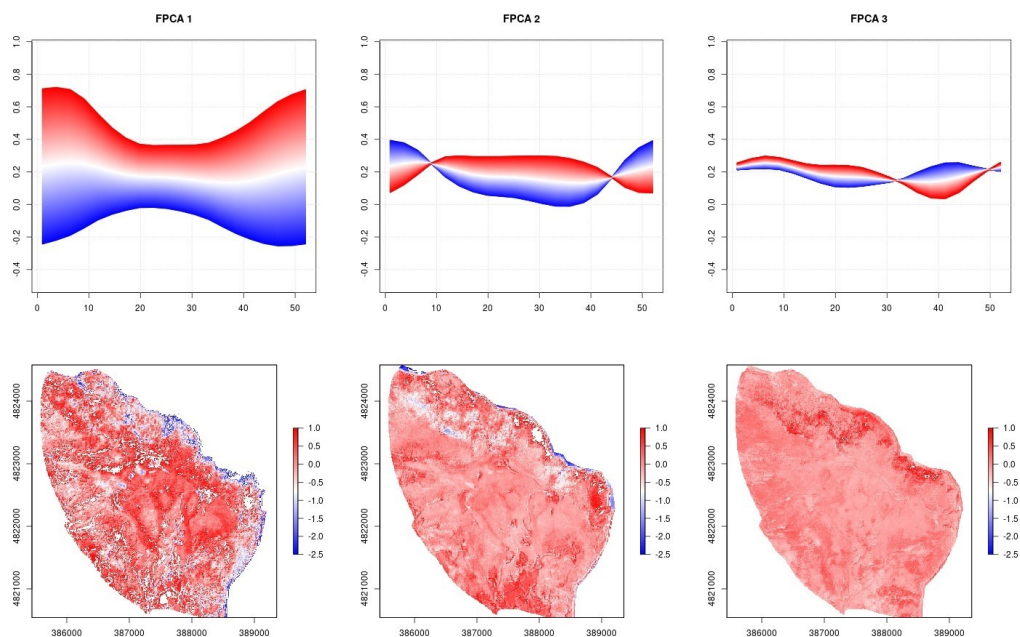


Figura 5-41: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice MNDWI.

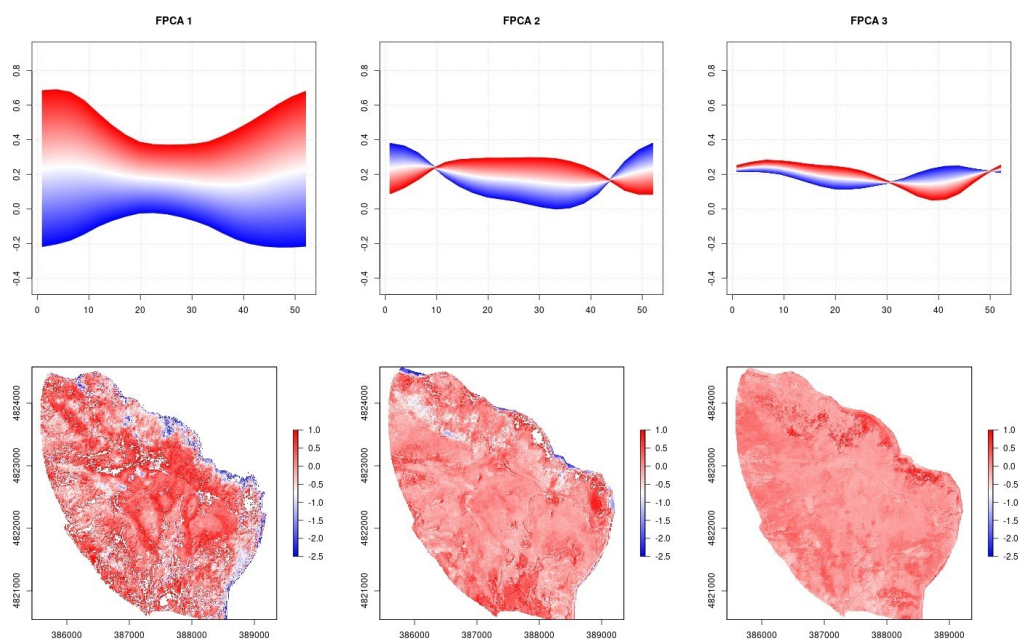


Figura 5-42: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice MNDWI.

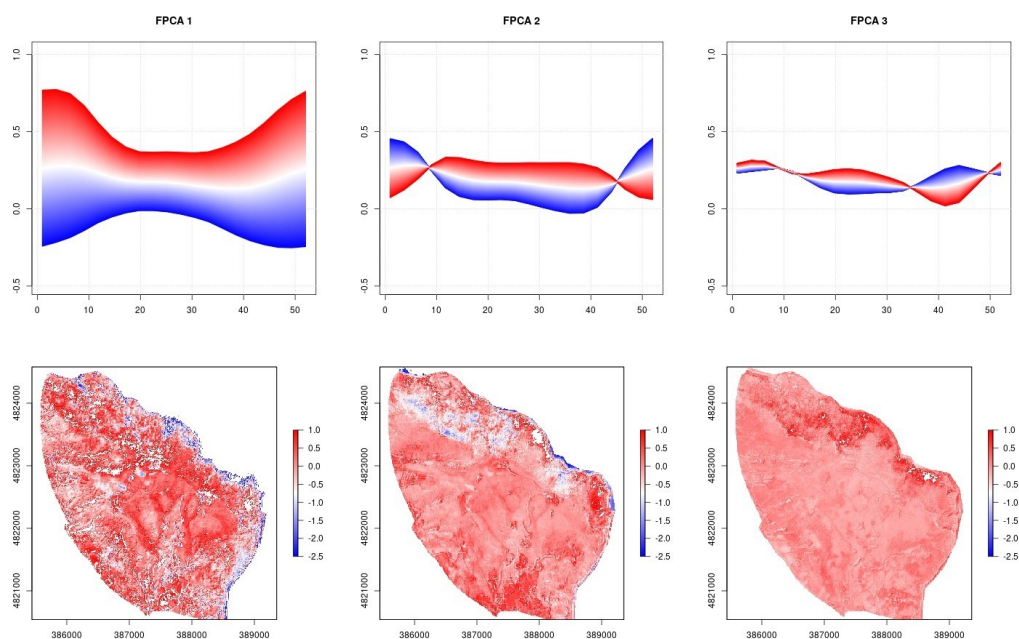


Figura 5-43: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice MNDWI.

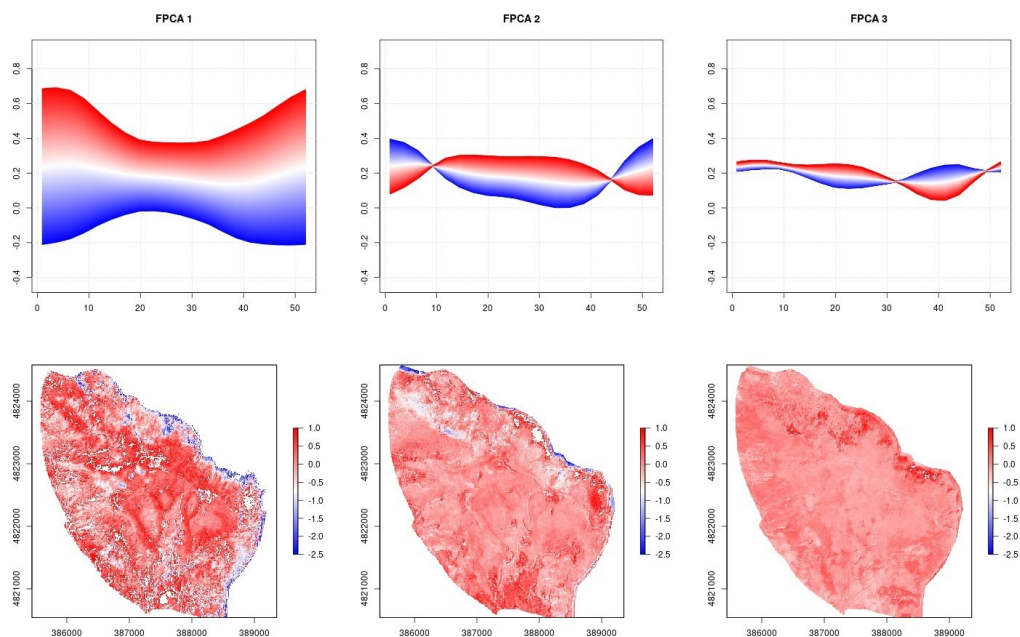


Figura 5-44: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice MNDWI.

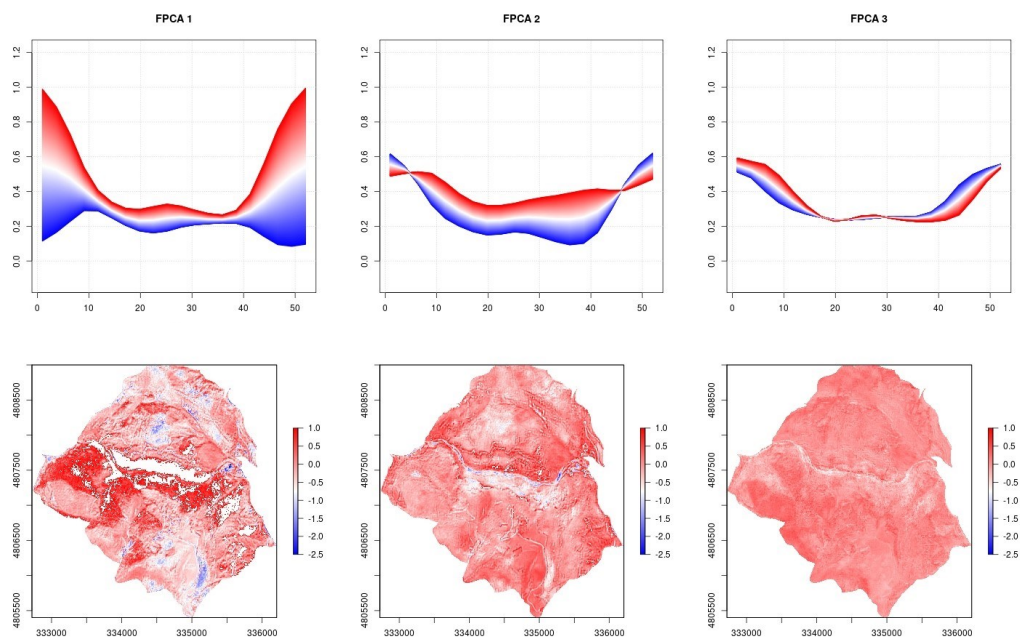


Figura 5-45: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice MNDWI.

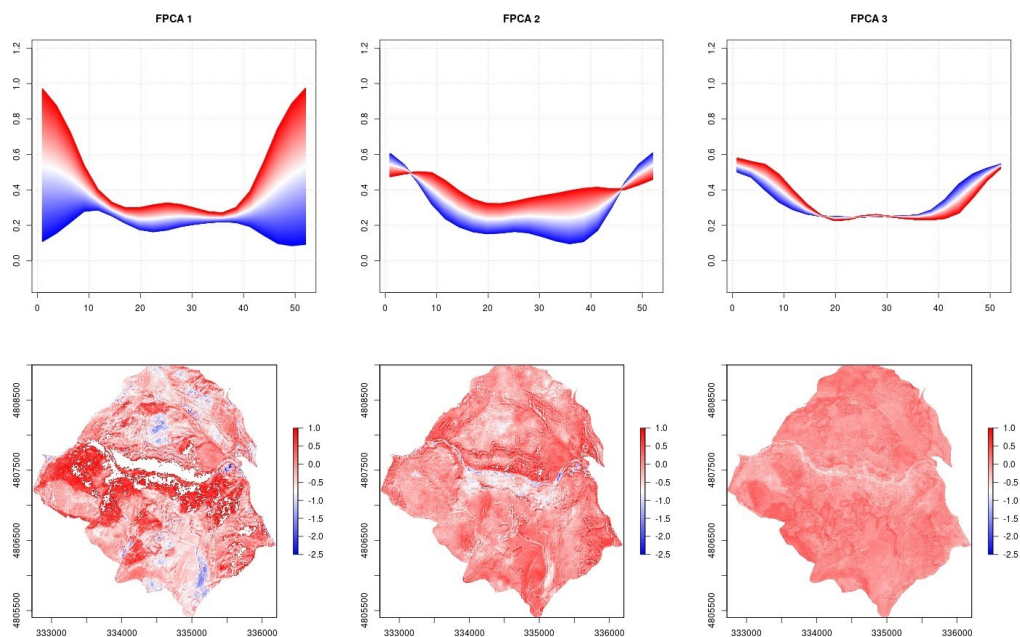


Figura 5-46: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice MNDWI.

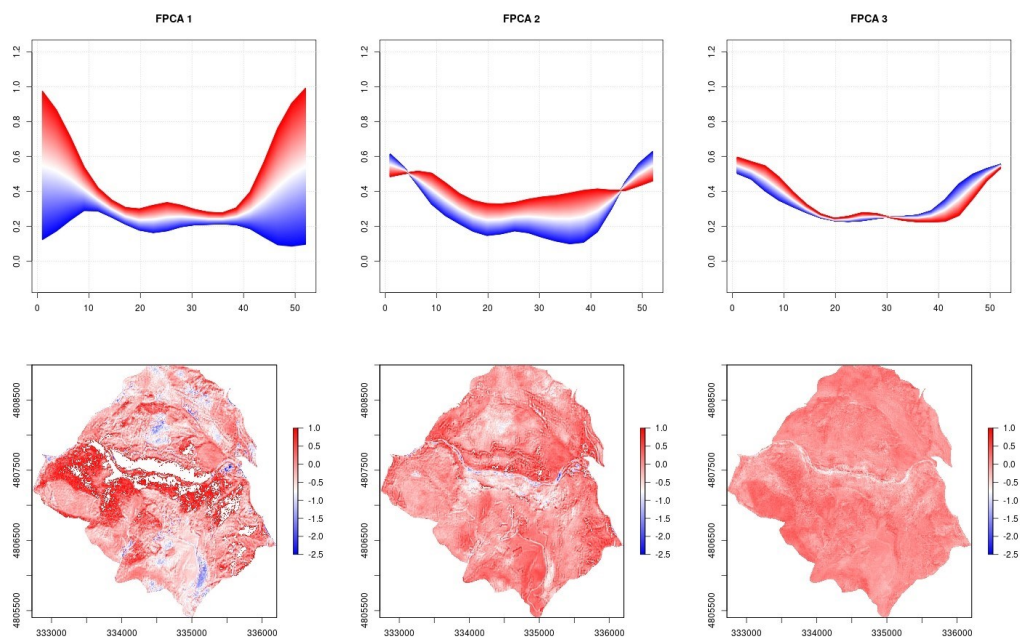


Figura 5-47: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice MNDWI.

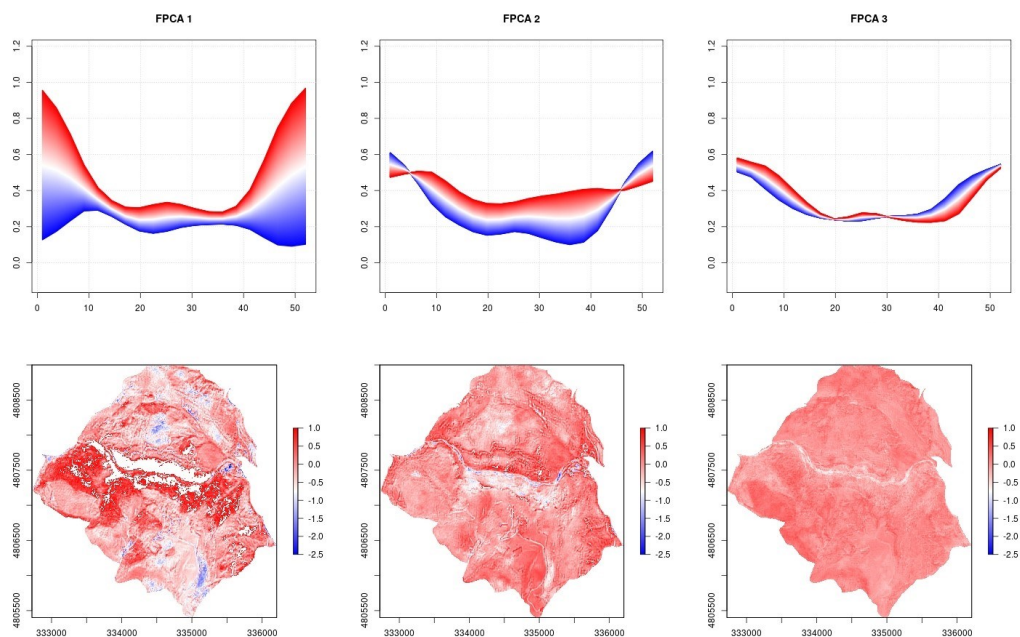


Figura 5-48: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice MNDWI.

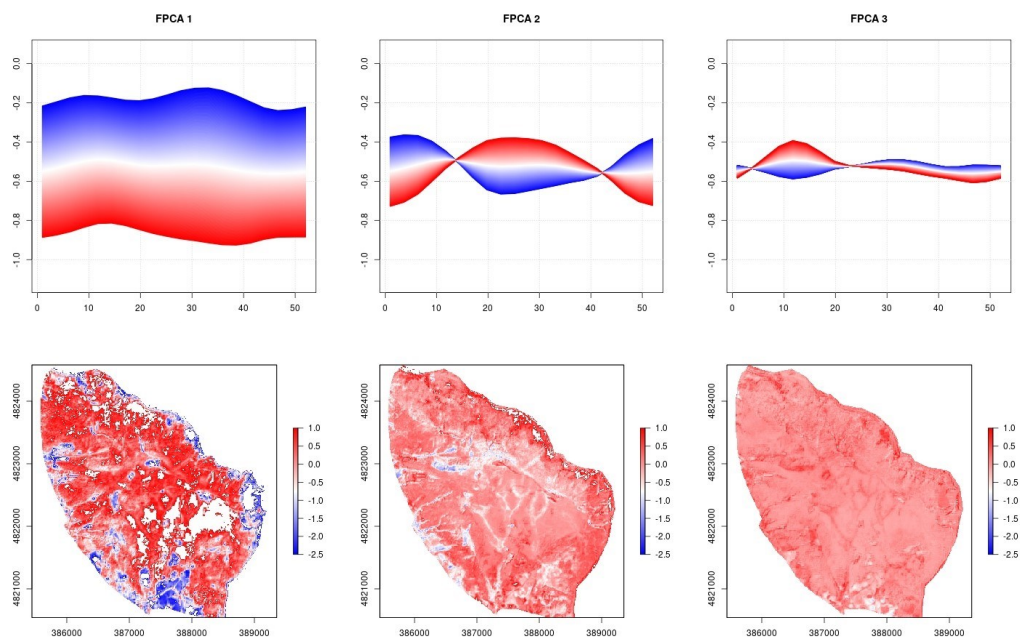


Figura 5-49: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M01, indice NDMI.

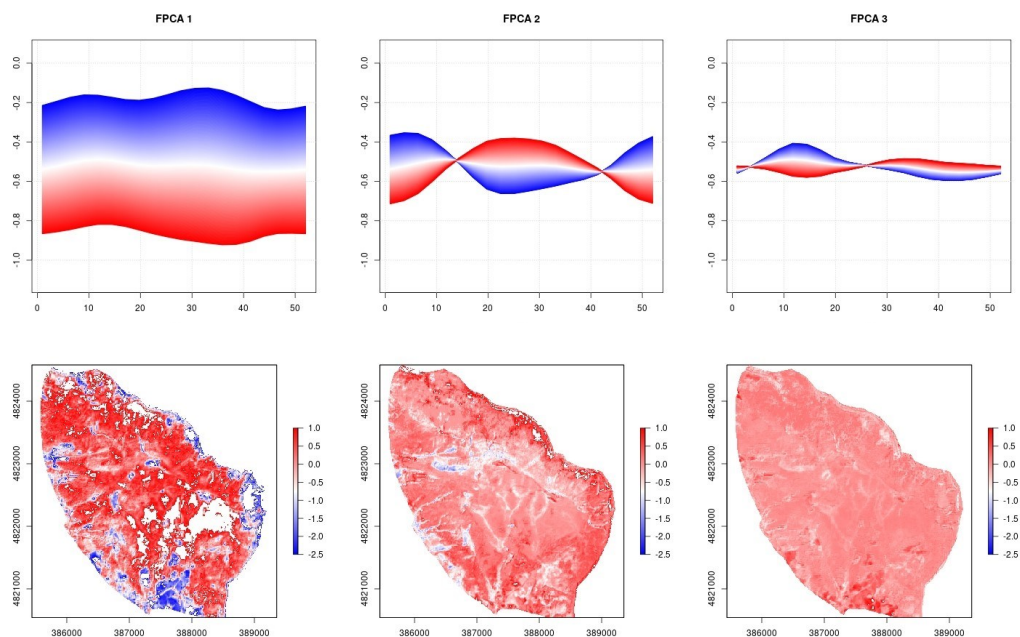


Figura 5-50: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M02, indice NDMI.

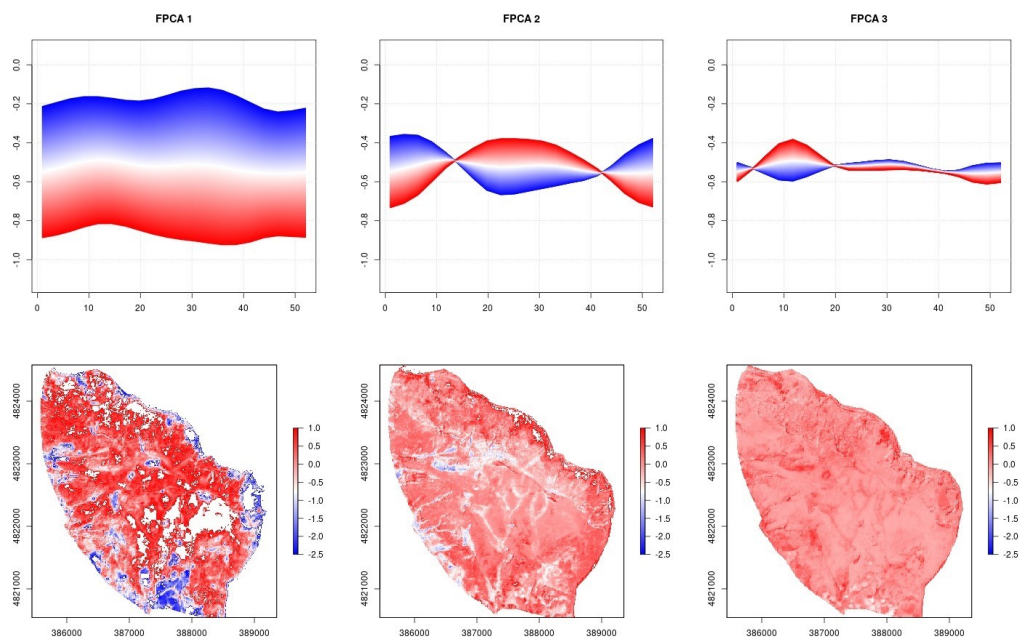


Figura 5-51: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M03, indice NDMI.

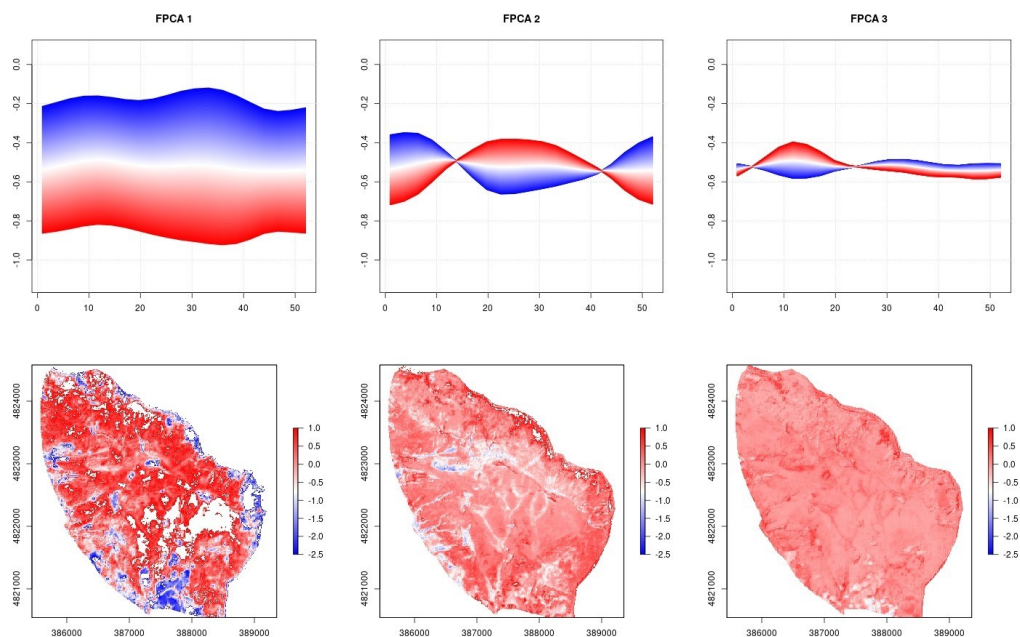


Figura 5-52: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Monte Conero, metodo M04, indice NDMI.

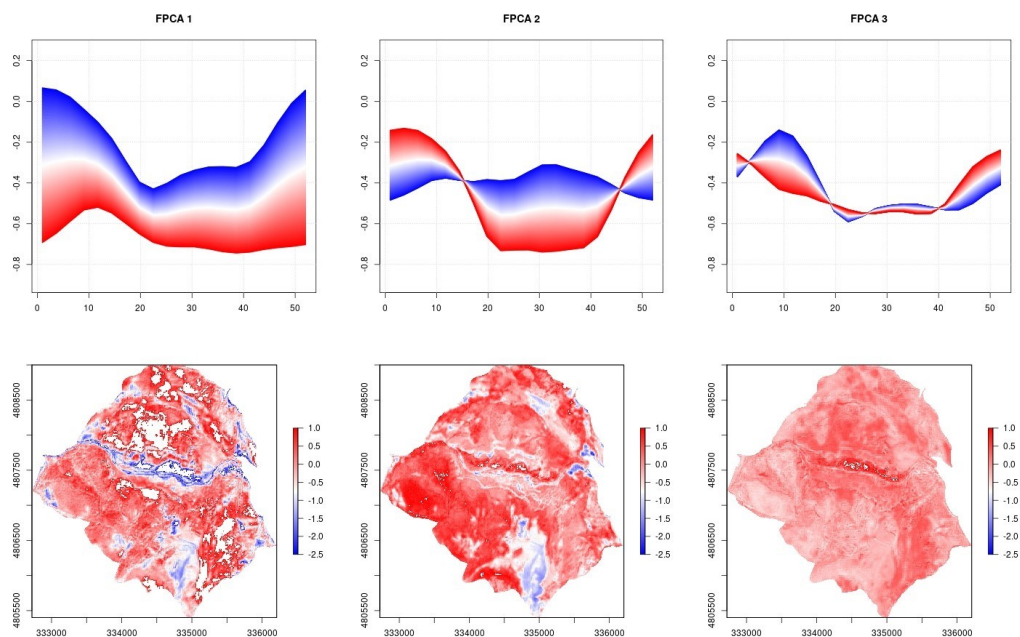


Figura 5-53: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M01, indice NDMI.

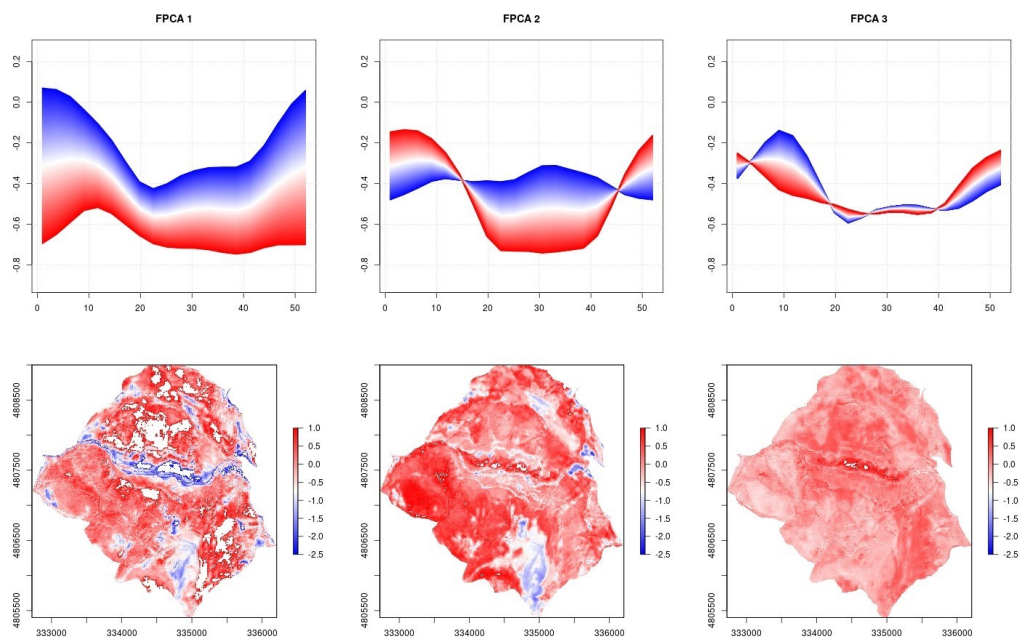


Figura 5-54: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M02, indice NDMI.

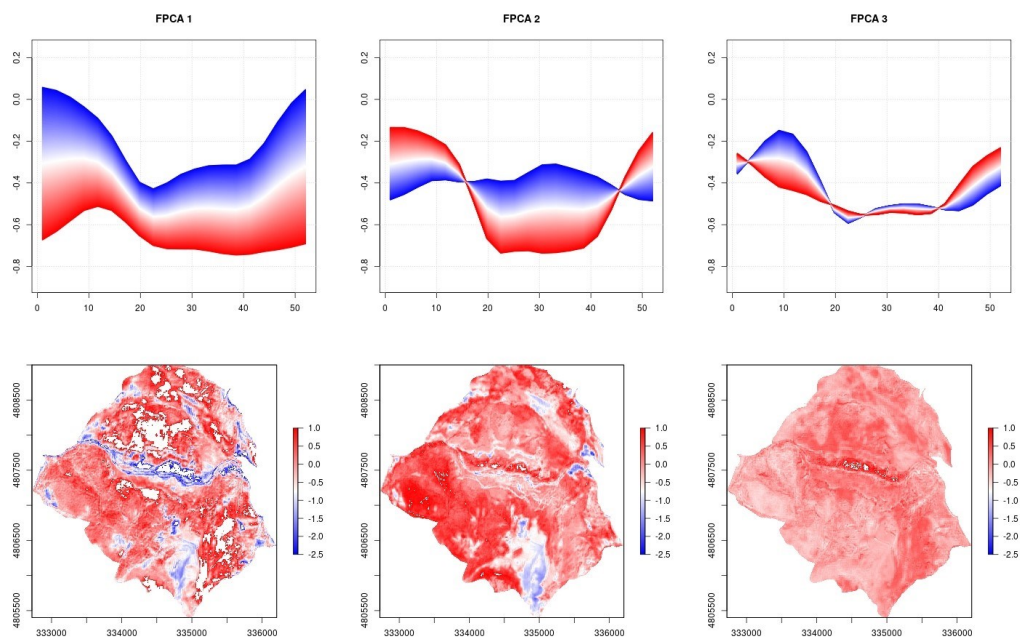


Figura 5-55: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M03, indice NDMI.

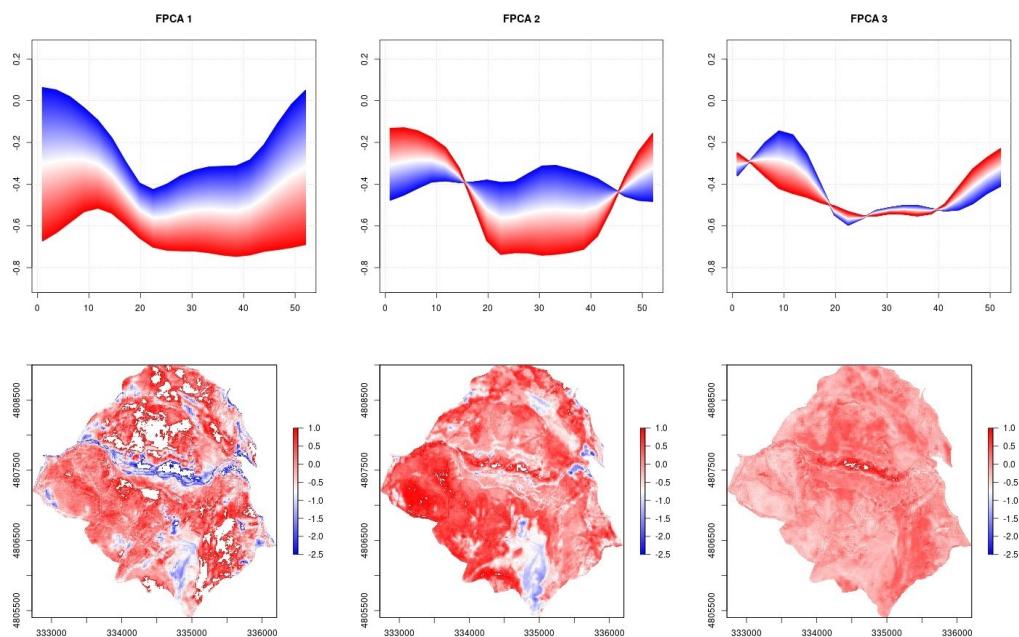


Figura 5-56: Pattern spazio-temporali descritti dalle prime 3 componenti principali. Area di studio Gola di Frasassi, metodo M04, indice NDMI.

Tabella 5-1: Matrice di confusione prodotta dal modello M01-SCR, area di studio Monte Conero, con i valori percentuale di producer accuracy e user accuracy.

	<i>Lt</i>	<i>Lm</i>	<i>O</i>	<i>R</i>	<i>producer</i>
<i>Lt</i>	57	8	6	0	80%
<i>Lm</i>	7	27	0	3	73%
<i>O</i>	1	0	12	0	92%
<i>R</i>	3	2	1	48	89%
<i>user</i>	84%	73%	63%	94%	

Tabella 5-2: Matrice di confusione prodotta dal modello M03-SCT, area di studio Monte Conero, con i valori percentuale di producer accuracy e user accuracy.

	<i>Lt</i>	<i>Lm</i>	<i>O</i>	<i>R</i>	<i>producer</i>
<i>Lt</i>	59	7	4	1	83%
<i>Lm</i>	8	28	0	1	76%
<i>O</i>	1	0	12	0	92%
<i>R</i>	3	3	1	47	87%
<i>user</i>	83%	74%	71%	96%	

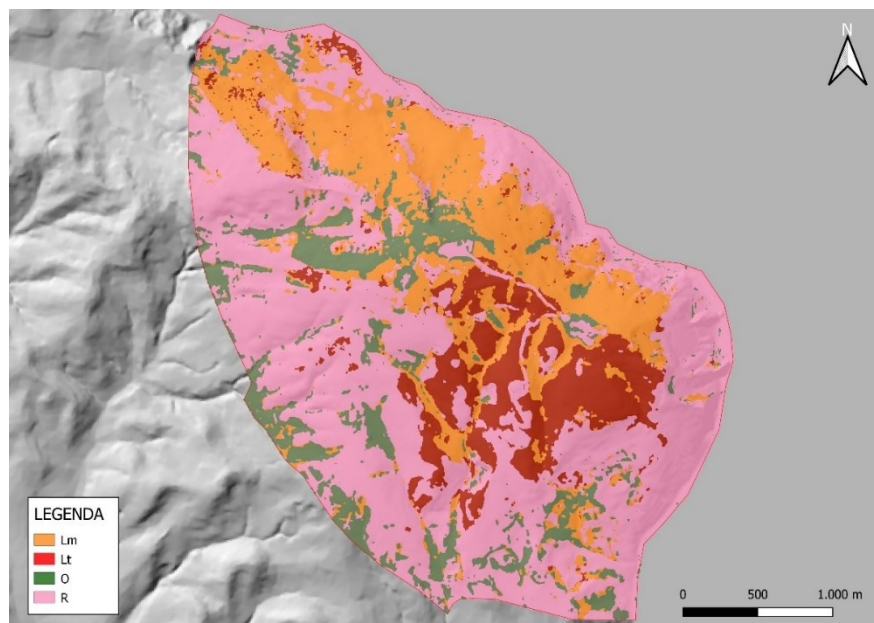


Figura 5-57: Mappa della vegetazione prodotta dal modello M01-SCR per l'area di studio Monte Conero. Categorie mappate: Lm – boschi mesofili a prevalenza di *Quercus ilex* (*Cephaanthero longifoliae* – *Quercetum ilicis*); Lt – boschi termofili a prevalenza di *Quercus ilex* (*Cyclamino hederifolii* – *Quercetum ilicis*); O – boschi a prevalenza di *Ostrya carpinifolia* (*Asparago acutifolii* – *Ostryetum carpinifoliae*); R – rimboschimenti.

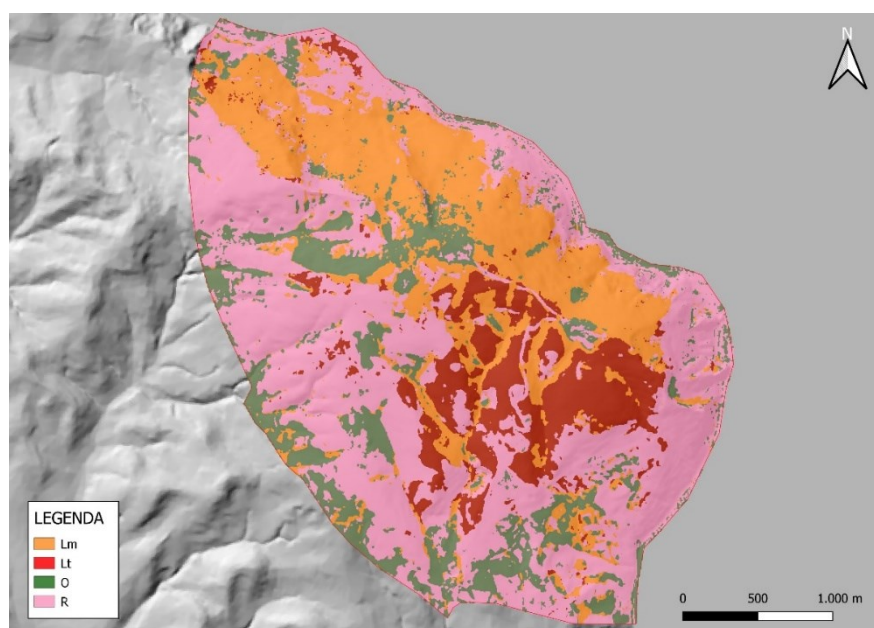


Figura 5-58: Mappa della vegetazione prodotta dal modello M03-SCT per l'area di studio Monte Conero. Categorie mappate: Lm – boschi mesofili a prevalenza di *Quercus ilex* (*Cephaanthero longifoliae* – *Quercetum ilicis*); Lt – boschi termofili a prevalenza di *Quercus ilex* (*Cyclamino hederifolii* – *Quercetum ilicis*); O – boschi a prevalenza di *Ostrya carpinifolia* (*Asparago acutifolii* – *Ostryetum carpinifoliae*); R – rimboschimenti.

Tabella 5-3: Matrice di confusione prodotta dal modello M03-SCR, area di studio Gola di Frasassi, con i valori percentuale di producer accuracy e user accuracy.

	<i>G</i>	<i>J</i>	<i>L</i>	<i>M</i>	<i>O</i>	<i>P</i>	<i>Q</i>	<i>R</i>	<i>producer</i>
<i>G</i>	13	1	0	0	0	0	2	0	81%
<i>J</i>	1	13	0	0	0	1	0	0	87%
<i>L</i>	0	0	30	0	1	0	0	3	88%
<i>M</i>	0	1	1	43	0	1	0	0	93%
<i>O</i>	0	0	4	0	36	0	15	1	64%
<i>P</i>	1	0	0	2	0	13	0	0	81%
<i>Q</i>	1	0	0	1	4	0	22	0	79%
<i>R</i>	1	0	2	0	1	0	0	27	87%
<i>user</i>	76%	87%	81%	93%	86%	87%	56%	87%	

Tabella 5-4: Matrice di confusione prodotta dal modello M04-SCR, area di studio Gola di Frasassi, con i valori percentuale di producer accuracy e user accuracy.

	<i>G</i>	<i>J</i>	<i>L</i>	<i>M</i>	<i>O</i>	<i>P</i>	<i>Q</i>	<i>R</i>	<i>producer</i>
<i>G</i>	12	1	0	0	0	0	3	0	75%
<i>J</i>	1	13	0	0	0	1	0	0	87%
<i>L</i>	0	0	30	0	1	0	0	3	88%
<i>M</i>	0	1	0	44	0	1	0	0	96%
<i>O</i>	2	0	5	0	35	0	13	1	63%
<i>P</i>	1	0	0	1	0	14	0	0	88%
<i>Q</i>	1	0	0	0	4	0	23	0	82%
<i>R</i>	0	1	3	0	1	0	0	26	84%
<i>user</i>	71%	81%	79%	98%	85%	88%	59%	87%	

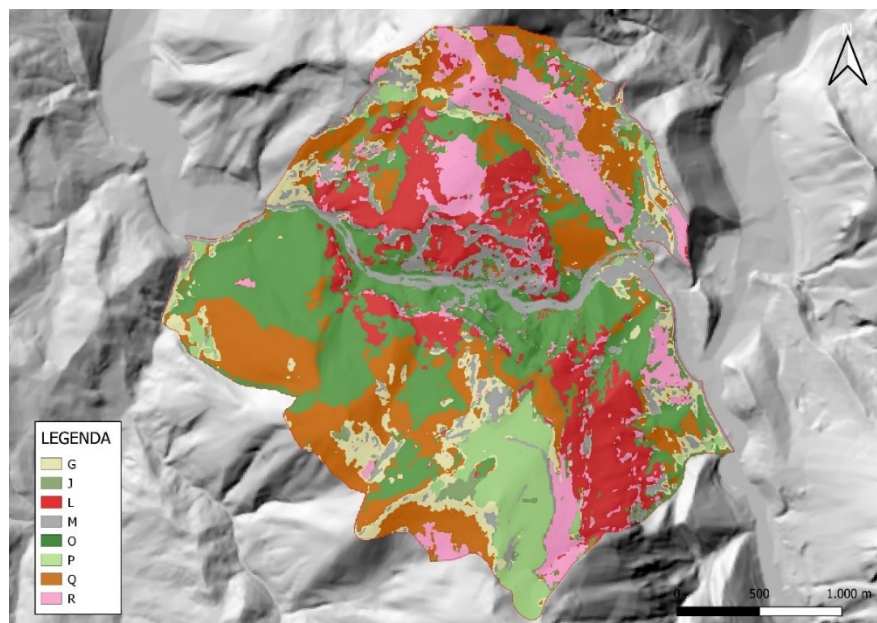


Figura 5-59: Mappa della vegetazione prodotta dal modello M03-SCR per l'area di studio Gola di Frasassi. Categorie mappate: G – ginestre (Spartio juncei – Cytisetum sessilifolii variante a Spartium junceum); J – gineprei (Spartio juncei – Cytisetum sessilifolii variante a Juniperus oxycedrus e Juniperus communis); L – boschi a prevalenza di Quercus ilex (Cephalanthero longifoliae – Quercetum ilicis); M – mosaico a gariga e macchia; O – boschi a prevalenza di Ostrya carpinifolia (Scutellario columnae – Ostryetum carpinifoliae); P – praterie (Asperulo purpureae – Brometum erecti); Q – boschi a prevalenza di Quercus pubescens (Cytiso sessilifolii – Quercetum pubescentis); R – rimboschimenti.

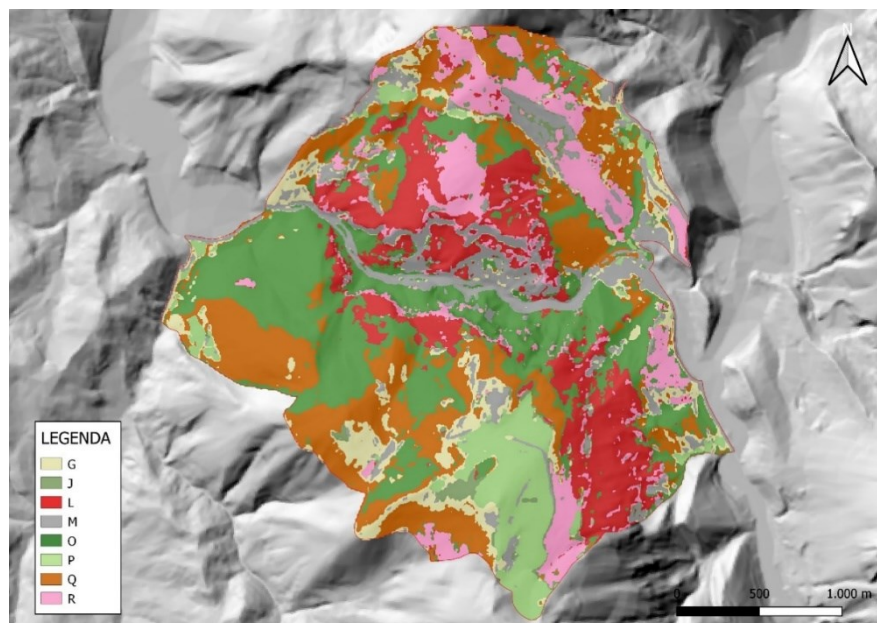


Figura 5-60: Mappa della vegetazione prodotta dal modello M04-SCR per l'area di studio Gola di Frasassi. Categorie mappate: G – ginestre (Spartium juncei – Cytisetum sessilifolii variante a Spartium junceum); J – ginepreti (Spartium juncei – Cytisetum sessilifolii variante a Juniperus oxycedrus e Juniperus communis); L – boschi a prevalenza di Quercus ilex (Cephalanthero longifoliae – Quercetum ilicis); M – mosaico a gariga e macchia; O – boschi a prevalenza di Ostrya carpinifolia (Scutellario columnae – Ostryetum carpinifoliae); P – praterie (Asperulo purpureae – Brometum erecti); Q – boschi a prevalenza di Quercus pubescens (Cytiso sessilifolii – Quercetum pubescentis); R – rimboschimenti.

BIBLIOGRAFIA

- Adelchi, A. and SCarpa, B. (2004) *Data Analysis and Data Mining*. Oxford: Oxford University Press.
- Badeck, F.W. *et al.* (2004) "Responses of spring phenology to climate change," *New Phytologist*, 162(2), pp. 295–309. doi:10.1111/j.1469-8137.2004.01059.x.
- Bannari, A. *et al.* (1995) "A review of vegetation indices," *Remote Sensing Reviews*, 13(1–2), pp. 95–120. doi:10.1080/02757259509532298.
- Benson, J.F. *et al.* (2019) "Extinction vortex dynamics of top predators isolated by urbanization," *Ecological Applications*, 29(3). doi:10.1002/eap.1868.
- Biondi, E., Casavecchia, S., Angelini, E., *et al.* (2009) *Carta fitosociologica della vegetazione SIC Monte Conero IT5320007*. Ancona.
- Biondi, E., Casavecchia, S., Galassi, S., *et al.* (2009) *Carta fitosociologica della vegetazione ZPS Gola della Rossa e di Frasassi IT5320017*. Ancona.
- Biondi, E. *et al.* (2012) "Diagnosis and syntaxonomic interpretation of Annex I Habitats (Dir. 92/43/EEC) in Italy at the alliance level," *Plant Sociology*, 49(1), pp. 5–37. doi:10.7338/pls2012491/01.
- Biondi, E., Bagella, S. and Casavecchia, S. (2001) "Analisi geobotaniche integrate per l'elaborazione del Piano di gestione naturalistica del Parco naturale regionale del Conero," *Informatore botanico italiano*, 33(1). Available at: <http://eprints.uniss.it/4514/>.
- Blázquez-García, A. *et al.* (2020) "A review on outlier/anomaly detection in time series data." Available at: <http://arxiv.org/abs/2002.04236>.
- vanden Borre, J. *et al.* (2011) "Integrating remote sensing in Natura 2000 habitat monitoring: Prospects on the way forward," *Journal for Nature Conservation*, 19(2), pp. 116–125. doi:10.1016/j.jnc.2010.07.003.
- Caldararu, S., Purves, D.W. and Palmer, P.I. (2014) "Phenology as a strategy for carbon optimality: A global model," *Biogeosciences*, 11(3), pp. 763–778. doi:10.5194/bg-11-763-2014.
- Caparros-Santiago, J.A., Rodriguez-Galiano, V. and Dash, J. (2021) "Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review," *ISPRS*

Journal of Photogrammetry and Remote Sensing, 171, pp. 330–347. doi:10.1016/j.isprsjprs.2020.11.019.

Cleveland, R.B. *et al.* (1990) “STL: A seasonal-trend decomposition procedure based on loess,” *Journal of Official Statistics*, 6, pp. 3–33.

Commissione Europea (2013) *Manuale di Interpretazione degli Habitat dell'Unione Europea*. Bruxelles.

Congalton, R.G. (2001) “Accuracy assessment and validation of remotely sensed and other spatial information,” *International Journal of Wildland Fire*, 10(3–4), pp. 321–328. doi:10.1071/wf01031.

Copernicus Europe’s eyes on Earth (2021) *Informazioni su Copernicus*.

Corbane, C. *et al.* (2015) “Remote sensing for mapping natural habitats and their conservation status - New opportunities and challenges,” *International Journal of Applied Earth Observation and Geoinformation*, 37, pp. 7–16. doi:10.1016/j.jag.2014.11.005.

Dash, J. and Ogutu, B.O. (2016) “Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems,” *Progress in Physical Geography*, 40(2), pp. 322–351. doi:10.1177/0309133316639403.

Dowle, M. (2021) “Package ‘data.table’ - Extension of ‘data.frame’.”

Estiarte, M. and Peñuelas, J. (2015) “Alteration of the phenology of leaf senescence and fall in winter deciduous species by climate change: Effects on nutrient proficiency,” *Global Change Biology*, 21(3), pp. 1005–1017. doi:10.1111/gcb.12804.

Evans, D. (2006) “The Habitats of the European Union Habitats Directive,” *Source: Biology and Environment: Proceedings of the Royal Irish Academy*, 106(3), pp. 167–173. Available at: http://www.jstor.orgURL:http://www.jstor.org/stable/20728591http://www.jstor.org/stable/20728591?seq=1&cid=pdf-reference#references_tab_contents.

Feret, J.B., Corbane, C. and Alleaume, S. (2015) “Detecting the Phenology and Discriminating Mediterranean Natural Habitats with Multispectral Sensors-An Analysis Based on Multiseasonal Field Spectra,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(5), pp. 2294–2305. doi:10.1109/JSTARS.2015.2431320.

Forconi, R., Mancini, A. and Pesaresi, S. (2019) *Classificazione di time-series da immagini Sentinel-2 mediante Functional Principal Component Analysis nell'ambito della Land Surface Phenology*.

- Forkel, M. *et al.* (2016) “Enhanced seasonal CO₂ exchange caused by amplified plant productivity in northern ecosystems,” *Science*, 351(6274), pp. 696–699. doi:10.1126/science.aac4971.
- Franklin, J. *et al.* (2015) “Regional variation in Caribbean dry forest tree species composition,” *Plant Ecology*, 216(6), pp. 873–886. doi:10.1007/s11258-015-0474-8.
- Friedman, J.H. (1984) “A variable span smoother,” *Journal of American Statistical Association* [Preprint].
- Garrity, S.R. *et al.* (2011) “A comparison of multiple phenology data sources for estimating seasonal transitions in deciduous forest carbon exchange,” *Agricultural and Forest Meteorology*, 151(12), pp. 1741–1752. doi:10.1016/j.agrformet.2011.07.008.
- Gigante, D. *et al.* (2016) “A methodological protocol for Annex i Habitats monitoring: The contribution of vegetation science,” *Plant Sociology*, 53(2), pp. 77–87. doi:10.7338/pls2016532/06.
- Gitelson, A.A. and Merzlyak, M.N. (1998) “Remote Sensing Of Chlorophyll Concentration In Higher Plant Leaves,” *Adv. Space Res.*, 22(5).
- Hijmans, R.J. (2021) “Package ‘raster’ Type Package Title Geographic Data Analysis and Modeling.” Available at: <https://github.com/rspsatial/raster/issues/>.
- Huete, A. *et al.* (2002) “Overview of the radiometric and biophysical performance of the MODIS vegetation indices,” *Remote Sensing for Environment* [Preprint]. Available at: www.elsevier.com/locate/rse.
- Hurley, C.M. *et al.* (2014) “Background factors predicting accuracy and improvement in micro expression recognition,” *Motivation and Emotion*, 38(5), pp. 700–714. doi:10.1007/s11031-014-9410-9.
- Hyndman, R. *et al.* (2021) “Package ‘forecast’ Title Forecasting Functions for Time Series and Linear Models Description Methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling.” Available at: <https://orcid.org/0000-0002-3665-9021>.
- Ichter, J. *et al.* (2017) “State-of-the-art of vegetation mapping in Europe: results of a European survey and contribution to the French program CarHAB,” *Documents phytosociologiques Série 3*, 6, pp. 335–352. Available at: <http://spn>.
- Khun, M. (2021) “Package ‘caret’ Title Classification and Regression Training.”
- Leonel Jorge (2018) *Supervised Learning, Medium*.
- Leutner, B. *et al.* (2019) “Package ‘RStoolbox’ Title Tools for Remote Sensing Data Analysis.”

Lieth, H. (1973) “Phenology in Productivity Studies,” in *Analysis of Temperate Forest Ecosystems*. Springer, Berlin, Heidelberg, pp. 29–46. doi:https://doi.org/10.1007/978-3-642-85587-0_4.

Lieth, H. (1974) “Purposes of a Phenology Book,” in *Phenology and Seasonality Modeling*. Springer, Berlin, Heidelberg, pp. 3–19. doi:https://doi.org/10.1007/978-3-642-51863-8_1.

Marshak, A. *et al.* (2000) “Cloud - vegetation interaction: Use of normalized difference cloud index for estimation of cloud optical thickness,” *Geophysical Research Letters*, 27(12), pp. 1695–1698. doi:10.1029/1999GL010993.

Marzialetti, F. *et al.* (2019) “Capturing coastal dune natural vegetation types using a phenology-based mapping approach: The potential of Sentinel-2,” *Remote Sensing*, 11(12). doi:10.3390/rs11121506.

Mini, L. *et al.* (2015) *Piano di Gestione dei Siti Natura 2000 Area di progetto “Gola della Rossa e di Frasassi” – ZPS IT5320017, SIC ITT5320004, SIC IT5320003*.

Mizunuma, T. *et al.* (2013) “The relationship between carbon dioxide uptake and canopy colour from two camera systems in a deciduous forest in southern England,” *Functional Ecology*, 27(1), pp. 196–207. doi:10.1111/1365-2435.12026.

Nicol, F. (2016) *Functional Principal Component Analysis of Aircraft Trajectories*. Available at: <https://hal-enac.archives-ouvertes.fr/hal-01349113>.

Paul Murrell, M. (2020) “Title Data and Functions from the Book R Graphics, Third Edition.” Available at: <https://www.stat.auckland.ac.nz/~paul/RG3e/index.html>.

Pesaresi, S. *et al.* (2020) “Mapping mediterranean forest plant associations and habitats with functional principal component analysis using Landsat 8 NDVI time series,” *Remote Sensing*, 12(7). doi:10.3390/rs12071132.

Polgar, C., Gallinat, A. and Primack, R.B. (2014) “Drivers of leaf-out phenology and their implications for species invasions: Insights from Thoreau’s Concord,” *New Phytologist*, 202(1), pp. 106–115. doi:10.1111/nph.12647.

Quattrini, G. *et al.* (2020) *Mappatura della vegetazione e degli habitat tramite l’analisi della fenologia telerilevata: il caso studio della ZSC “IT5320003 - Gola di Frasassi.”*

Quattrini, G., Pesaresi, S. and Casavecchia, Si. (2016) *Classificazione e mappatura delle serie temporali NDVI di alcune comunità forestali del Monte Conero*.

Ranghetti, L. *et al.* (2020) “‘sen2r’: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data,” *Computers and Geosciences*, 139. doi:10.1016/j.cageo.2020.104473.

- Rapinel, S. *et al.* (2019) “Evaluation of Sentinel-2 time-series for mapping floodplain grassland plant communities,” *Remote Sensing of Environment*, 223, pp. 115–129. doi:10.1016/j.rse.2019.01.018.
- Richardson, A.D. *et al.* (2013) “Climate change, phenology, and phenological control of vegetation feedbacks to the climate system,” *Agricultural and Forest Meteorology*, 169, pp. 156–173. doi:10.1016/j.agrformet.2012.09.012.
- Sims, D.A. and Gamon, J.A. (2002) “Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages,” *Remote Sensing for Environment*, 81, pp. 337–354. Available at: www.elsevier.com/locate/rse.
- Soriga, A. *et al.* (2015) *Piano di gestione dei siti Natura 2000 inclusi all'interno del territorio del Parco del Conero – S.I.C. “Portonovo e falesia calcarea a mare” (IT5320006), S.I.C. “Monte Conero (IT5320007), Z.P.S. “Monte Conero” (IT5320015).*
- Sundaresan, A., Varshney, P.K. and Arora, M.K. (2007) “Robustness of Change Detection Algorithms in the Presence of Registration Errors,” *Photogrammetric Engineering & Remote Sensing* [Preprint].
- T Daughtry, C.S. *et al.* (1999) “Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance,” *Remote Sensing for Environment*, 74, pp. 229–239.
- The European Space Agency (2021) *Sentinel-2 MSI User Guide*.
- Tucker, C.J. (1979) “Red and Photographic Infrared Linear Combinations for Monitoring Vegetation,” *Remote Sensing of Environment*, 8, pp. 127–150.
- Verrelst, J. *et al.* (2019) “Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods,” *Surveys in Geophysics*, 40(3), pp. 589–629. doi:10.1007/s10712-018-9478-y.
- Viciani, D. *et al.* (2016) “Detailed Natura 2000 and CORINE Biotopes habitat maps of the island of Elba (Tuscan Archipelago, Italy),” *Journal of Maps*, 12(3), pp. 492–502. doi:10.1080/17445647.2015.1044040.
- Walther, G.-R. *et al.* (2002) “Ecological responses to recent climate change,” *Nature*, 416, pp. 389–395. doi:<https://doi.org/10.1038/416389a>.
- Wolkovich, E.M. *et al.* (2012) “Warming experiments underpredict plant phenological responses to climate change,” *Nature*, 485(7399), pp. 494–497. doi:10.1038/nature11014.
- Wood, S.N. (2017) “Generalized additive models: An introduction with R, second edition,” *Generalized Additive Models: An Introduction with R, Second Edition*, pp. 1–476. doi:10.1201/9781315370279.

Xue, J. and Su, B. (2017) “Significant remote sensing vegetation indices: A review of developments and applications,” *Journal of Sensors*. Hindawi Limited. doi:10.1155/2017/1353691.

Young, N.E. *et al.* (2017) “A survival guide to Landsat preprocessing,” *Ecology*, 98(4), pp. 920–932. doi:10.1002/ecy.1730.

Zlinszky, A. *et al.* (2015) “Mapping natura 2000 habitat conservation status in a pannonic salt steppe with airborne laser scanning,” *Remote Sensing*, 7(3), pp. 2991–3019. doi:10.3390/rs70302991.